Lecture Notes in Networks and Systems 960

Fausto Pedro García Márquez
Akhtar Jamil
Alaa Ali Hameed
Isaac Segovia Ramírez *Editors*

Emerging Trends and Applications in Artificial Intelligence

Selected papers from the International Conference on Emerging Trends and Applications in Artificial Intelligence (ICETAI)

Springer

Lecture Notes in Networks and Systems 960

The series “Lecture Notes in Networks and Systems” publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the worldwide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

For proposals from Asia please contact Aninda Bose (aninda.bose@springer.com).

Fausto Pedro García Márquez · Akhtar Jamil ·
Alaa Ali Hameed · Isaac Segovia Ramírez
Editors

Emerging Trends and Applications in Artificial Intelligence

Selected papers from the International Conference on Emerging Trends and Applications in Artificial Intelligence (ICETAI)

Springer

Editors

Fausto Pedro García Márquez
Ingenium Research Group
University of Castilla-La Mancha
Ciudad Real, Spain

Akhtar Jamil
National University of Computer
and Emerging Sciences
Islamabad, Pakistan

Alaa Ali Hameed
Department of Computer Engineering
Istinye University
Istanbul, Türkiye

Isaac Segovia Ramírez
Ingenium Research Group
University of Castilla-La Mancha (UCLM)
Ciudad Real, Spain

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-3-031-56727-8 ISBN 978-3-031-56728-5 (eBook)
https://doi.org/10.1007/978-3-031-56728-5

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Preface

Emerging Trends and Applications in Artificial Intelligence

Selected papers from the International Conference on Emerging Trends and Applications in Artificial Intelligence (ICETAI)

This book is a compilation of the selected papers presented at the International Conference on Emerging Trends and Applications in Artificial Intelligence (ICETAI) in 2023.

The conference has been organized by the Istanbul Medipol University, Turkey, on September 08–09, 2023. This event brought together leading experts, researchers, scholars, and professionals from around the world to share their latest findings and explore the newest advances in the field of artificial intelligence. As technology continues to shape our lives, the role of artificial intelligence has become increasingly significant. This conference provided a unique opportunity to gain insights into the latest developments and applications of artificial intelligence in the digital age. From cutting-edge research to real-world applications, the conference provided a comprehensive overview of the field and its impact on society.

This conference managed a large number of submissions of original, high-quality research papers, where only a few were accepted. Authors submitted their work in areas related to artificial intelligence and its applications, including, but not limited to, machine learning, deep learning, computer vision, natural language processing, robotics, and more. All submissions were reviewed by a panel of experts in the field, and the accepted papers are presented in this book. This is an excellent opportunity for researchers, scholars, and professionals to showcase their work and contribute to the advancement of the field. Submissions were made through the conference website following the submission guidelines.

Each paper was peer-reviewed by at least two reviewers and evaluated based on originality, technical depth, correctness, relevance to conference, contributions, and readability. The papers were accepted based on technical merit, interest, applicability, and how well they fit a coherent and balanced technical program.

The conference was carried out in hybrid mode.

The book highlights some of the latest research advances and cutting-edge analysis of real-world case studies on computational intelligence, data analytics, IoT, and applications from a wide range of international contexts. It also identified business applications and the latest findings and innovations in Operations Management and the Decision Sciences, e.g.:

Data Analysis and Visualization

- Exploratory Data Analysis
- Statistical and mathematical modeling
- Business Intelligence

- Big Data Analysis
- Data Mining
- Cloud Computing Architecture and Systems
- ETL and Big Data Warehousing
- Business Intelligence
- Data Visualization
- Statistical Analysis

Computer Vision

- Document Analysis
- Biometrics and Pattern Recognition
- Remote Sensing & GIS
- Medical Image Processing
- Image and Video Retrieval
- Motion Analysis
- Structure from Motion
- Object Detection and Recognition
- Image Restoration
- Speech and Audio Processing
- Signal Processing

Artificial Intelligence

- Machine Learning
- Pattern Recognition
- Deep Learning
- Human–Computer Interactions
- Medical Image Processing
- Image and Video Retrieval
- Audio Video Processing
- Text Analytics
- Natural Language Processing
- Information Retrieval
- Robotics Applications

Internet of Things

- 3D Printing
- Securing IoT infrastructure
- Future of IoT and Big Data
- Internet of Things
- Intelligent Systems for IoT
- Security, Privacy, and Trust
- Visual Analytics IoT
- Data Compression for IoT Devices
- IoT Services and Applications
- Education and Learning
- Social Networks Analysis

Communication Systems and Networks

- Antennas, Propagation and RF Design
- Transmission and Communication Theory
- Wireless/Radio Access Technologies
- Optical Networks and NGN
- 5G & 6G Cellular systems and SON
- Sensor Networks
- Multimedia and New Media
- High-Speed Communication.
- Computational Intelligence in Telecommunications

Software Engineering

- Requirements Engineering
- Security Aspects
- Agile Software Engineering
- Software Evolution & Reuse
- Reverse Engineering
- Software Dependability
- Data & AI Monetization and Products
- Data as a Service/Platform
- Biomedical Experiments and Simulations
- Decision Support Systems

Fausto Pedro García Márquez
Akhtar Jamil
Alaa Ali Hameed
Isaac Segovia Ramírez

Organization

General Chairs

Gökhan Silahtaroğlu	School of Business and Management Sciences, Istanbul Medipol University, Turkiye
Samee U. Khan	Department of Electrical & Computer Engineering, Mississippi State University, USA

Advisory Committee

Ahmed Abdelgawad	Central Michigan University, USA
Gabriella Casalino	Università degli Studi di Bari Aldo Moro Bari, Italy
Hasan Dincer	Istanbul Medipol University, Turkiye
Rene Vinicio Sanchez Loja	Universidad Politécnica Salesiana, Ecuador
Serhat Yuksel	Istanbul Medipol University, Turkiye

Program Chairs

Fausto Pedro Garcia Marquez	University of Castilla–La Mancha, Spain
Kivanç Kök	Istanbul Medipol University, Turkiye

Organizing Committee

Kemal Özdemir	Istanbul Medipol University, Turkiye
Bahadır Güntürk	Istanbul Medipol University, Turkiye
Elif Baykal	Istanbul Medipol University, Turkiye
Humera Azam	University of Karachi, Pakistan
Özge Doguc	Istanbul Medipol University, Turkiye
Kailas Hambarde	University of Beira Interior, Portugal
Kevser Şahinbaş	Istanbul Medipol University, Turkiye
Tuba Khalifa	Istanbul Medipol University, Turkiye
Esra Baytören	Istanbul Medipol University, Turkiye
David Yeregui Marcos	University of Leon, Spain

Technical Program Chairs

Alaa Ali Hameed	Istinye University, Turkiye
Akhtar Jamil	National University of Computer and Emerging Sciences, Pakistan

Publication And Publicity Chairs

Dilek Yomralıoğlu	Istanbul Medipol University, Turkiye
Nada Misk	Istanbul Medipol University, Turkiye

Registration Committee

Sümeyye Özdemir	Istanbul Medipol University, Turkiye
Müge Eke	Istanbul Medipol University, Turkiye

Technical Program Committee

Abd Ullah Khan	National University of Sciences and Technology, Pakistan
Ahsan Altaf	Senior RF Systems Expert, Ahsan Altaf, Sweden
Alfredo Peinado Gonzalo	University of Castilla-La Mancha, Spain
Ali Javed	University of Engineering and Technology Taxila, Pakistan
Ali Osman Serdar Çıtak	Istanbul Medipol University, Turkiye
Alimatu-Saadia Yussiff	University of Cape Coast
Ameur Bensefia	Higher Colleges of Technology, Abu Dhabi, UAE
Atınç Yılmaz	Istanbul Beykent University, Istanbul, Türkiye
Aymen M. Khodayer Al-Dulaimi	Al-Farahidi University, Iraq
Bharat Bhushan	Sharda University, India
Chawki Djeddi	University of Rouen, France
Enkeleda Lulaj	University Haxhi Zeka, Kosovo
Faezeh Soleimani	Ball State University, USA
Ferhat Özgür Çatak	University of Stavanger, Norway
Francoise Contreras	Colombia, Universidad del Rosario
Ghulam Abid	Kinnaird College For Women, Pakistan
Haroon Rashid	Universiti Kebangsaan Malaysia/Xpertopedia Academy, Malaysia

Hasan Ali Khattak	National University of Sciences and Technology, Pakistan
Ihsan Ali	University of Malaya, Malaysia
Isaac Segovia Ramírez	University of Castilla-La Mancha, Spain
Isidro Peña García-Pardo	University of Castilla-La Mancha, Spain
J. Satpathy	Management University of Africa, Kenya and Srinivas University, India
Kiran Sood	Chitkara University, India
Mahavir Arjun Devmane.	VPPCOE & VA, India
María del Valle Fernández Moreno	University of Castilla-La Mancha, Spain
Marina Karpitskaya	Yanka Kupala University, Belarus
Momina Mustehsan	Bahria University, Pakistan
Muhammad Bilal	University of Engineering and Technology, Lahore, Pakistan
Muhammad Ilyas	Altinbas University, Turkiye
Muhammad Zeshan Alam	Brandon University, Canada
Muhammed Davud	Istanbul Sabahattin Zaim University, Turkiye
Muhsin Jaber Jweeg C	Al-Farahidi University, Iraq
Muneshwar Rajesh Niranjan	Amrutvahini College of Engineering Sangamner, India
Murat Kuzlu	Old Dominion University, Norfolk, VA, USA
Mustafa Al-asadi	KTO Karatay University, Turkey
Mustafa Takaoğlu	TÜBİTAK-BİLGEM, Kocaeli, Turkey
Naresh Kumar	University of Nizwa, Oman
Natalia Markovskaya	Yanka Kupala University, Belarus
Öznur Gülen Ertosun	Istanbul Medipol University, Turkiye
Özlem İlday	Istanbul Medipol University, Turkiye
Pedro José Bernalte Sánchez	University of Castilla-La Mancha, Spain
Rajasekaran S.	University of Technology and Applied Sciences, India
Raed Khalid Ibraheem	Al-Farahidi University, Iraq
Rana Atabay	Istanbul Medipol University, Turkiye
Reda CHEFIRA	Private University of Marrakesh, Morocco
S. G. Gollagi	KLE College of Engineering and Technology, India
Salih Sarp	Virginia Commonwealth University, Richmond, VA, USA
Şebnem Özdemir	Istanbul İstinye University, Istanbul, Türkiye
Serkan Eti	Istanbul Medipol University, Istanbul, Türkiye
Shivaji Ramdas Lahane	R. H. Sapat College of Engineering Management Studies and Research Nashik, India
Sibel Senan	Department of Computer Engineering, Istanbul University, Turkiye

Sobhan Sarkar	University of Edinburgh, UK
Subhan Ullah	National University of Computer and Emerging Sciences, Pakistan
Syed Attique Shah	Birmingham City University, UK
Tareq Hafdi Al-Khayat	Al-Farahidi University, Iraq
Ufuk Baytar	Istanbul Topkapı University, Istanbul, Türkiye
Ümit Çelebi	Istanbul Medipol University, Turkiye
Varsha Bhole	A.C. Patil College of Engineering, India
Vinayakumar Ravi	Prince Mohammad Bin Fahd University, Saudi Arabia
Waleed Ead	Beni-Suef University, Egypt
Waleed Khalid Ibraheem	Al-Farahidi University, Iraq
Yurdagül Meral	Istanbul Medipol University, Turkiye
Zeynep Orman	Istanbul University, Turkiye

Contents

About the Editors

Fausto Pedro Garcia Marquez Fausto works at UCLM as Full Professor (Accredited as Full Professor from 2013), Spain, Honorary Senior Research Fellow at Birmingham University, UK, Lecturer at the Postgraduate European Institute, Research Fellow at INTI International University & Colleges, Malaysia, and he has been Senior Manager in Accenture (2013-2014). He obtained his European PhD with maximum distinction. He has been distinguished with the prices: Runner Prize (2023), Nominate Prize (2022), Gran Maestre (2022), Grand Prize (2021), Runner Prize (2020) and Advancement Prize (2018), Runner (2015), Advancement (2013) and Silver (2012) by the International Society of Management Science and Engineering Management (ICMSEM), First International Business Ideas Competition 2017 Award (2017), etc. He has published more than 248 papers (156 JCR: 74-Q1; 42-Q2; 32-Q3; 8-Q4), some recognized as: "Progress in Photovoltaics: Research and Applications" (Q1, IF. 8.49, one of the most downloaded in first 12 months of publications, 2023), "Applied Energy" (Q1, IF 9.746, as "Best Paper 2020"), "Renewable Energy" (Q1, IF 8.001, as "Best Paper 2014"); "ICMSEM" (as "excellent"), "Int. J. of Automation and Computing" and "IMechE Part F: J. of Rail and Rapid Transit" (most downloaded), etc. He is the author and editor of over 50 books (Elsevier, Springer, Pearson, Mc-GrawHill, Intech, IGI, Marcombo, AlfaOmega,…), >100 international chapters, and 6 patents. He is the Editor of 5 Int. Journals, Committee Member more than 70 Int. Conferences. He has been Principal Investigator in 4 European Projects, 8 National Projects, and more than 150 projects for universities, companies, etc. His main interests are: artificial intelligence, maintenance, management, renewable energy, transport, advanced analytics, and data science.

He is being: Expert in the European Union in AI4People (EISMD), and ESF.; Director of www.ingeniumgroup.eu.; Senior Member at IEEE, 2021-… ; Honored Honorary Member of the Research Council of Indian Institute of Finance, 2021-… ; Committee Chair of The International

Society for management science and Engineering Management (ISMSEM), 2020-…. His main interests are: artificial intelligence, maintenance, management, renewable energy, transport, advanced analytics, data science.

Dr. Akhtar Jamil is Associate Professor in the Department of Computer Science at the National University of Computer and Emerging Sciences, Islamabad, Pakistan. Before joining FAST, he served as Assistant Professor and Vice Head of the Computer Engineering Department at Istanbul Sabahattin Zaim University, Istanbul, Turkey. He also served as a Lecture at COMSATS University, Islamabad. He has also worked in the industry as a developer for several years. He received his Ph.D. in machine learning and remote sensing from Yildiz Technical University, Istanbul, Turkey. He has published more than 50 high-quality papers in well-known journals and top conferences. He received a fully funded Ph.D. scholarship from the Turkish government. He is the founding member of the ICMI, ICAETA and ICCIDA conferences. He serves as a reviewer for several journals and conferences. He focuses on applied research for solving real-world problems. His current research interests include statistical machine learning, deep learning, pattern recognition, data analytics, image classification, and remote sensing.

Alaa Ali Hameed received his Master's degree in computer engineering from Eastern Mediterranean University, North Cyprus, in 2012, and his Ph.D. degree from the Department of Computer Engineering at Selcuk University, Turkey, in 2017. He worked as an Assistant Professor in the Department of Computer Engineering, at Istanbul Aydin University, Turkey, from 2017 to 2019. He then moved to Istanbul Sabahattin Zaim University, Turkey, where he worked as an Assistant Professor in the Department of Computer Engineering from 2019–2022. Currently, he is Assistant Professor in the Department of Computer Engineering at Istinye University, Turkey. He has published more than 60 technical articles in top international journals and conferences in a short span of time. He has served as a Program Chair and a Technical Program Chair member for many international conferences; also he has served as a Guest Editor for many SCIE journals. His research interests include digital signal and image processing, adaptive filters, adaptive computing, data mining, machine, and deep learning, big data and data analytics, neural networks and self-learning systems, and artificial intelligence.

Isaac Segovia Ramirez Industrial Engineer at ETSII of the University of Castilla-La Mancha, Ciudad Real (2015) and Master of Industrial Engineering at ETSII of the University of Castilla-La Mancha, Ciudad Real (2019). Associate professor of electronic courses in 2019–2020 in ETSII of the University of Castilla-La Mancha, Ciudad Real. Current PhD student, he is in collaboration with several national and European projects with the Department of Business Administration of the University of Castilla La Mancha and the (June 2013-present) Ingenium group. Winner of the international contest “Entrepreneurship 5+5” in Tunisia. He has been awarded with the “Advancement Prize for Management Science and Engineering Management with the Nominated Prize (2018)” for the article “Remotely Piloted Aircraft System and Engineering Management: A Real Case Study”. His main research interests are related to maintenance management, UAVs, renewable energy, detection of elements in surface by infrared radiation, etc.

Simultaneous Optimization of Ride Comfort and Energy Harvesting Through a Regenerative, Active Suspension System Using Genetic Algorithm

Hassan Sayyaadi(✉) and Jamal Seddighi

Department of Mechanical Engineering, Sharif University of Technology, Tehran, Iran
sayyaadi@sharif.edu, seddighi.sayyedjamal@mech.sharif.edu

Abstract. Active suspension systems have long been recognized as an effective means of improving ride comfort and vehicle handling. However, high energy consumption and a lack of economic justification have hindered their commercial adoption in the industry. In order to address the challenges, this research proposed an innovative control structure that utilizes linear electromagnetic actuators capable of functioning in both motor and generator modes. To implement the proposed method, a suitable vehicle dynamic model available within the Adams software was selected. An analytical model corresponding to the software model was then extracted and verified to ensure its accuracy and reliability for use in GA optimization algorithms. Assuming only ride maneuvers, a feedback control structure based on meaningful terms in vehicle dynamics was developed. Then by using a GA algorithm, the ride comfort and energy harvesting criteria were simultaneously optimized. Finally, by exploiting the most suitable set of coefficients in the developed control structure, the suspension system showed the ability to recover up to 650 watts of power on rough roads, while leading to a 45% improvement in ride comfort.

Keywords: Artificial Intelligence · Genetic Algorithm · Active Suspension · Ride Comfort · Energy Harvesting · Multi-objective Optimization

1 Introduction

Over the past three decades, active control technologies have been continually developed to enhance vehicle dynamics [1]. Among these technologies, active suspension systems (ASSs) have demonstrated significant potential in improving ride comfort and handling. Despite extensive research, challenges such as high costs, substantial energy consumption, and a lack of functional justification have hindered the practical application of ASSs in the automotive industry [2]. In passenger cars, suspension systems have the potential to harvest energy equivalent to 3% of fuel consumption [3]. In general, the design of any suspension system primarily focuses on achieving ride comfort and road holding [4]. In numerous previous studies, a weighted combination of multiple objectives has

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 1–12, 2024.
https://doi.org/10.1007/978-3-031-56728-5_1

been optimized as a single fitness function [5]. In this research, linear electromagnetic actuators are used as active dampers. These actuators can function in both motor and generator modes. This research aims to optimize ride comfort and energy harvesting by utilizing a genetic algorithm.

1.1 Suspension System and Vehicle Vibration

In determining the damping behavior of a suspension, low damping results in a more comfortable ride but compromises the car's handling. The sky hook and ground hook methods are classical approaches in the control design of semi-active suspension systems. One primary limitation of these methods is that they only consider the vertical coordinates of an axle, neglecting other aspects of vehicle dynamics, which results in a deviation from reality. Another shortcoming is the lack of utilization of speed data and other information reflecting the dynamic state of the car [6]. Therefore, having access to vehicle speed data and an estimate of road quality can significantly enhance their performance [7]. Based on Fig. 1, in passive systems, the damping behavior of a suspension system is represented by a constant curve. In semi-active systems, the damping coefficient can be continuously adjusted within a wide range at any given moment [8]. In the second and fourth areas, the damping coefficient is negative, signifying energy production. An active suspension system is required for operation within these two areas [9].

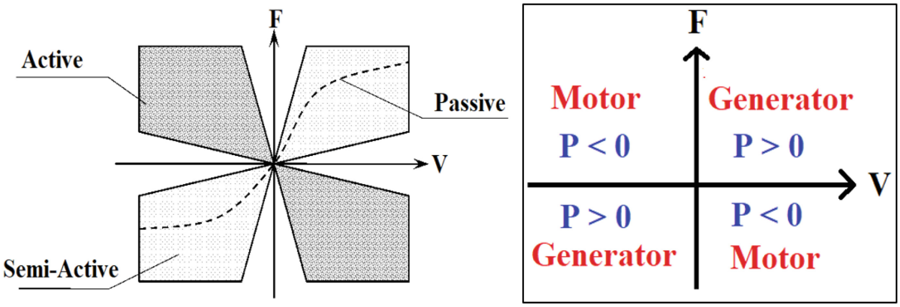


Fig. 1. Damping Behavior and Operational Scope of Renowned Suspension Control Systems

1.2 Ride Comfort Measurement

To evaluate a car's ride comfort, both subjective and objective methods are employed [10]. Various standards exist for the objective measurement of ride comfort [11]. The most prevalent among them is the ISO2631 standard [12]. According to this standard, the weighted mean square acceleration can be calculated using (1) in which $a_w(t)$ represents the weighted acceleration as a function of time, and T denotes the duration of data acquisition, expressed in seconds.

$$a_w = \left[\frac{1}{T}\int_0^T a_w(t)^2 dt\right]^{\frac{1}{2}} \tag{1}$$

1.3 Energy Harvesting Through Suspension System

The vibrations of a vehicle's suspension system have the potential to harvest part of the energy that would otherwise be wasted in the suspension's dampers [13]. The amount of harvested energy varies depending on factors such as vehicle speed, road quality, vehicle class, and the structure of the harvesting system [14]. Generally, the energy harvesting capacity of passenger car suspension systems is equivalent to 3% of the vehicle's fuel consumption, while off-road vehicles can achieve up to 6% [15]. Based on Fig. 1, in the second and fourth areas, negative power necessitates energy injection into the system. Conversely, in the first and third regions, the positive product of damping force and speed results in positive power, enabling energy harvesting [16]. To serve this purpose, an electromagnetic motor operating in both motor and generator mode can be used [17].

1.4 Optimization Using Genetic Algorithm

Optimization methods using Artificial Intelligence (AI) involve using algorithms and techniques to find the best possible solution to a complex problem. These methods are often used in industries such as logistics, manufacturing, and finance to improve efficiency and reduce costs [18]. There are several optimization methods that use AI, including Genetic Algorithms, Particle Swarm, and Ant Colony. These innovative approaches have been extensively utilized in the field of vehicle dynamics [19]. Genetic Algorithm is a type of optimization algorithm inspired by the process of natural selection in biological systems. It starts with a population of potential solutions to a problem, represented as "chromosomes" made up of genes. These chromosomes are evaluated for their fitness, or how well they solve the problem. The fittest chromosomes are then selected to "breed" and produce offspring, which inherit traits from their parents. This process of selection and reproduction continues for several generations, with the hope that the population will converge to a solution that is optimal or near-optimal [20].

Table 1. The vibrational characteristics of the studied

Parameter	Symbol	Unit	Front Axle Value	Rear Axle Value
Sprung Mass	m_s	kg	922	731
Unsprung Mass	m_u	kg	108	94
Axle Distance to Sprung CG	L_s	m	1.303	1.644
Tire Vertical Rate	K_t	N/m	420000	420000
Wheel Vertical Rate	K_w	N/m	105700	106200

2 Software and Analytical Modeling

In vehicle dynamics applications, Adams software offers results that closely resemble reality due to its precise modeling of geometry, kinematics, and dynamic properties of the chassis components [21]. To implement the method, a standard car model defined in

Adams software was utilized and illustrated in Fig. 2. The vibrational characteristics of the studied car are stated in Table 1.

2.1 Analytical Model Extraction

During the initial phase of designing a suspension vibration control system, it is crucial to develop a simplified model that closely resembles real-world conditions, can be easily coded, and incorporated into the optimization loop [22]. Hence, as depicted in Fig. 2, a four-degree-of-freedom model representing the vibrations observed from the side of the car was derived.

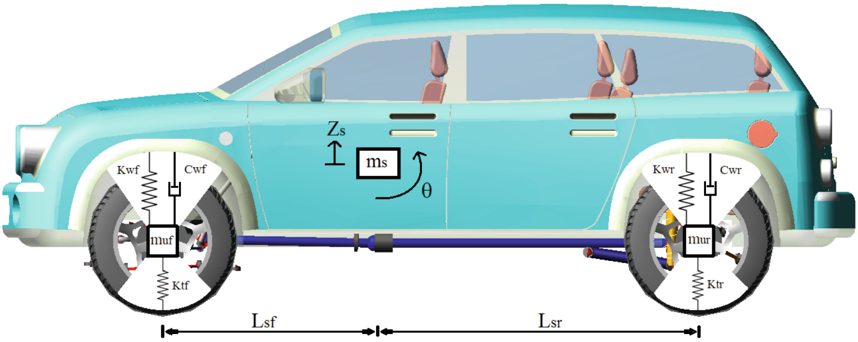


Fig. 2. Side-View Four Degree of Freedom Vibration Model of the Studied Vehicle

As represented in Eqs. 2 through 4, the undamped vibration equations of the analytical model have four generalized coordinates including z_s, θ, z_{uf}, and z_{ur} [23].

$$m_s\ddot{z}_s + K_{wf}\left(z_s - L_{sf}\theta - z_{uf}\right) + K_{wr}(z_s + L_{sr}\theta - z_{ur}) = 0 \tag{2}$$

$$I_s\ddot{\theta} + K_{wf}L_{sf}\left(z_s - L_{sf}\theta - z_{uf}\right) + K_{wr}L_{sr}(z_s + L_{sr}\theta - z_{ur}) = 0 \tag{3}$$

$$m_{uf}\ddot{z}_{uf} - K_{wf}\left(z_s - L_{sf}\theta - z_{uf}\right) - K_{tf}\left(z_{rf} - z_{uf}\right) = 0 \tag{4}$$

$$m_{ur}\ddot{z}_{ur} - K_{wr}(z_s + L_{sr}\theta - z_{ur}) - K_{tr}(z_{rr} - z_{ur}) = 0 \tag{5}$$

To verify the accuracy of the model, a straight-line bump test was conducted on the studied car using the Adams software. The same longitudinal speed and road bump input were then applied to the developed linear model. The results, which compare the vertical position of the front and rear axles of the sprung mass, are illustrated in Fig. 3. The simulation results of the linear half-car model exhibit coherence and similarity with the model developed in the Adams software.

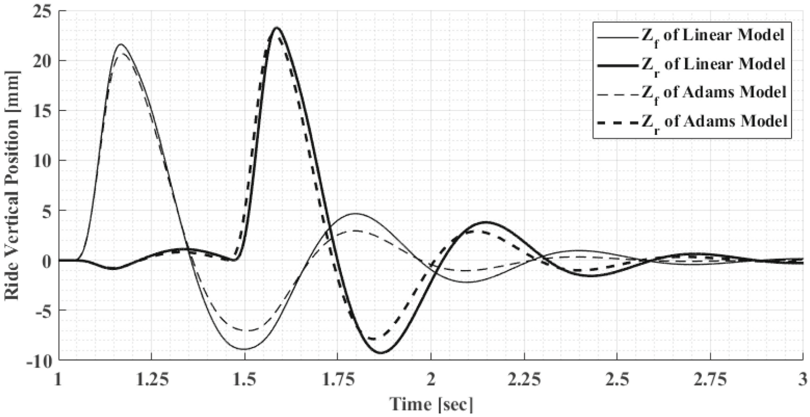


Fig. 3. Comparing Vertical Tip Positions of the Studied Vehicle for the Designed Bump Test

3 Defining the Control Structure and Parameters

Using Genetic Algorithm, this study aims to optimize ride comfort and energy harvesting. To achieve this, an optimal algorithm based on meaningful concepts in vehicle dynamics has been developed. In the 2DOF side-view vibration model, high-frequency vibrations of the unsprung mass can be disregarded, allowing the suspension and tire stiffness to be considered as an equivalent series stiffness.

3.1 Defining the Control Vector

The front and rear vertical forces can be expressed as follows:

$$F_f = F_f^{road} + F_f^u \tag{6}$$

$$F_r = F_r^{road} + F_r^u \tag{7}$$

The excitation force of the road and the suspension's actuator are as follows:

$$\begin{Bmatrix} F_f \\ F_r \end{Bmatrix}^{road} = \begin{bmatrix} K_f + C_f s & 0 \\ 0 & K_r + C_r s \end{bmatrix} \tag{8}$$

$$\begin{Bmatrix} F_z^u \\ F_\theta^u \end{Bmatrix} = \begin{bmatrix} 1 & 1 \\ -L_{sf} & L_{sr} \end{bmatrix} \begin{Bmatrix} F_f^u \\ F_r^u \end{Bmatrix} = \begin{Bmatrix} m_s^u \ddot{z}_s \\ I_s^u \ddot{\theta} \end{Bmatrix} \tag{9}$$

Therefore, the equation of the closed loop system is as follows:

$$\begin{bmatrix} m_s + m_s^u & 0 \\ 0 & I_s + I_s^u \end{bmatrix} \begin{Bmatrix} \ddot{z}_s \\ \ddot{\theta} \end{Bmatrix} + \begin{bmatrix} C_f + C_r & C_r L_{sr} - C_f L_{sf} \\ C_r L_{sr} - C_f L_{sf} & C_r L_{sr}^2 + C_f L_{sf}^2 \end{bmatrix} \begin{Bmatrix} \dot{z}_s \\ \dot{\theta} \end{Bmatrix} + \\ \begin{bmatrix} K_f + K_r & K_r L_{sr} - K_f L_{sf} \\ K_r L_{sr} - K_f L_{sf} & K_r L_{sr}^2 + K_f L_{sf}^2 \end{bmatrix} \begin{Bmatrix} z_s \\ \theta \end{Bmatrix} = \begin{bmatrix} 1 & 1 \\ -L_{sf} & L_{sr} \end{bmatrix} \begin{Bmatrix} F_f \\ F_r \end{Bmatrix}^{road} \tag{10}$$

3.2 Defining the Control Parameters and Their Limitations

To establish the search space, the control coefficients are defined as follows:

$$-0.5 < \sigma_m^u = \frac{m_s^u}{m_s} < 1.5 \quad -0.5 < \sigma_I^u = \frac{I_s^u}{I_s} < 1.5 \tag{10}$$

$$0.1 < \zeta_f = \frac{C_f}{2\sqrt{\frac{m_s(1+\sigma_m^u)L_{sr}}{L}K_f}} < 0.7 \;\; 0.1 < \zeta_r = \frac{C_r}{2\sqrt{\frac{m_s(1+\sigma_m^u)L_{sf}}{L}K_r}} < 0.7 \tag{11}$$

4 Defining the Genetic Algorithm and the Test Procedure

Developing a control algorithm that can simultaneously optimize energy harvesting and ride comfort is the primary challenge of this research. The goal of a genetic algorithm is to find the best solution to this problem by mimicking the process of natural selection. The algorithm starts with a population of potential solutions, and then applies genetic operators such as mutation, crossover, and selection, to evolve the population towards better solutions [24].

4.1 Defining the Search Space and Initial Population

The search resolution was divided into 1024 parts using 10 house chromosomes, resulting in a binary matrix of 10 * 4 for each member of the population. A complete search of the search space would require over 1000 billion searches with this level of accuracy. However, genetic algorithms can efficiently find optimal points in less time by using targeted and intelligent search strategies [25].

$$N = \sum\nolimits_1^{10} a_n * 2^n \;\; N_{max} = 1024 \tag{12}$$

$$\sigma_{u,m} = -0.5 + \frac{N_{u,m}}{N_{max}} * 1.5 \, \zeta_{b,p} = 0.1 + \frac{N_{b,p}}{N_{max}} * 0.6 \tag{13}$$

4.2 Developing a Fitness Function for Evaluating Ride Comfort

To objectively assess the ride comfort for each pair of acceleration feedback coefficients, the frequency gains introduced in the ISO 2631 standard are utilized. The ride comfort criterion and the corresponding fitness function are then calculated as (15).

$$\begin{matrix} \ddot{Z}_s^{\omega} = \ddot{Z}_s(\omega)W_k(\omega) \\ \ddot{\theta}^{\omega} = \ddot{\theta}(\omega)W_e(\omega) \end{matrix} \Rightarrow CFF = RMS\left(\sqrt{\left(\ddot{Z}_s^{\omega}\right)^2 + \left(\ddot{\theta}^{\omega}\right)^2}\right) \tag{14}$$

4.3 Developing a Fitness Function for Evaluating Energy Harvesting

To develop this fitness function, firstly, the feedback force required to control the acceleration of pitch and bounce modes is calculated. Then, using the coordinate transformation matrix, the force corresponding to the acceleration feedback in the front and rear axis is calculated.

$$\begin{aligned} F_b^u &= m_s^u \ddot{z}_s \\ F_p^u &= I_s^u \ddot{\theta} \end{aligned} \tag{15}$$

$$\begin{aligned} F_f^u &= 0.5579 F_b^u - 0.3393 F_p^u \\ F_r^u &= 0.4421 F_b^u + 0.3393 F_p^u \end{aligned} \tag{16}$$

Finally, by considering the damping force, the forces of the front and rear suspension actuator are calculated as follows:

$$\begin{aligned} F_f^d &= C_f \left(\dot{z}_s - L_f \dot{\theta} - \dot{z}_{uf} \right) \\ F_r^d &= C_r \left(\dot{z}_s + L_r \dot{\theta} - \dot{z}_{ur} \right) \end{aligned} \tag{17}$$

$$\begin{aligned} F_f^{act} &= F_f^u + F_f^d \\ F_f^{act} &= F_f^u + F_f^d \end{aligned} \tag{18}$$

After the suspension actuators' power is obtained as (20), by integrating the power, the suspension actuators' energy is computed as follows:

$$\begin{aligned} P_f^{act} &= F_f^{act} * \left(\dot{z}_s - L_f \dot{\theta} - \dot{z}_{uf} \right) \\ P_f^{act} &= F_f^{act} * \left(\dot{z}_s + L_r \dot{\theta} - \dot{z}_{ur} \right) \end{aligned} \tag{19}$$

$$\begin{aligned} E_f^{act} &= \int_t P_f^{act} dt \\ E_r^{act} &= \int_t P_r^{act} dt \end{aligned} \tag{20}$$

Therefore, the total harvested energy is calculated as follows:

$$EFF = E_f^{act} + E_r^{act} \tag{21}$$

4.4 Defining the Total Fitness Function

At this point, for each pair of acceleration feedback coefficient, the overall fitness function is defined as follows:

$$TFF = \left(\frac{\min(CFF)}{CFF} + \frac{EFF}{max(EFF)} \right) / 2 \tag{22}$$

Table 2. Correlation between Vehicle Speed and Road Quality for Ride Test

Road Quality Based on ISO 8608	Vehicle Speed [km/h]
A	90, 100
B	70, 80
C	40, 50, 60
D	10, 20, 30

4.5 Designing the Test Procedure

The vertical dynamics of a vehicle's suspension system are primarily influenced by road's vertical geometry, necessitating a proper model for this application [26]. One precise and reliable technique for simulating a road profile involves using the power spectral density of the road profile in conjunction with inverse Fourier concepts [27]. Consequently, this research employs this method to reproduce the road profile. Based on ISO 8608, the power spectrum density of the road is used [28].

In reality, there exists a logical correlation between the road quality and the car's longitudinal speed [29]. As such, driving speeds on well-maintained roads tend to be higher than on poorly maintained ones [30]. While no standard and comprehensive relationship exists in this field, a review of previous related research has yielded appropriate and reasonable values, as presented in Table 2.

5 Optimization Process Execution and Result Comparison

This research focuses on improving ride comfort and energy harvesting during ride maneuvers. To achieve this goal, at various speeds, the corresponding roads are traversed by the studied vehicle.

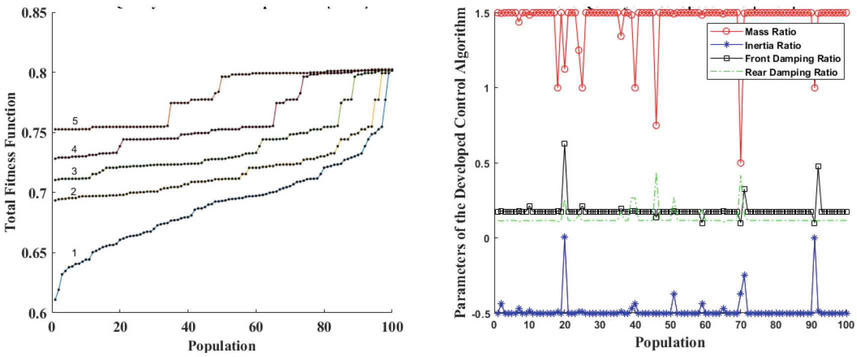


Fig. 4. The convergence of total fitness function at a speed of 25 km/h on a road with a quality rating of D, as well as the best selected individuals

5.1 Optimization Algorithm Convergence and The Best Selection

As a clarifying example, a ride test was conducted at a speed of 25 km/h on a road with a quality rating of D. The total fitness function was calculated and its convergence for first five iterations is depicted in Fig. 4, as well as the final population representing the best selected individuals.

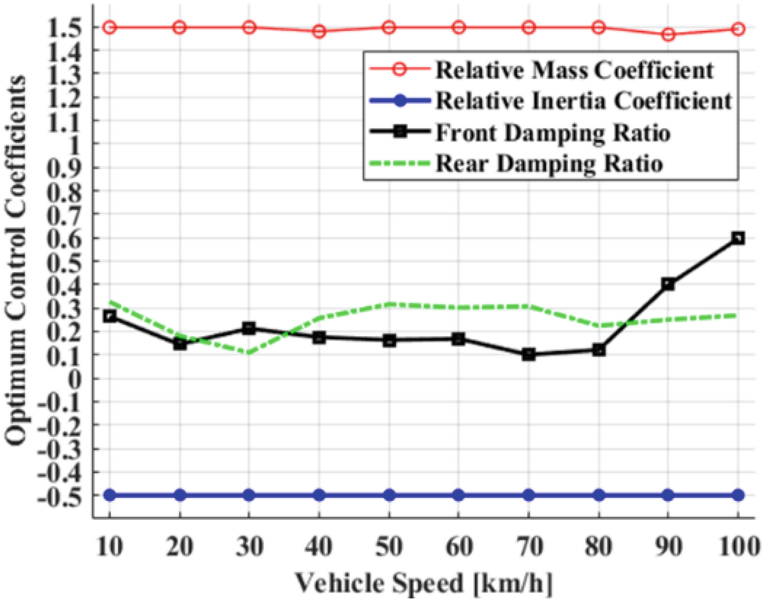


Fig. 5. Optimal control coefficients for various speeds

5.2 Performance Comparison Between Optimized and Base System

By utilizing the proposed control method and optimization algorithm, optimal control coefficients were determined for various speeds and illustrated in Fig. 5. To evaluate the efficacy of the optimized control system and compare its performance with that of the passive suspension system in the base car, a ride test was conducted on the studied vehicle over four different roads, each with varying levels of quality and corresponding speeds. The results, as presented in Table 3, demonstrate that the designed control system provides a 45% improvement in ride comfort.

The optimized control system not only enhances ride comfort but also exhibits significant power harvesting capabilities on rough roads. The study conducted in Table 3, reveals that both vehicle speed and road quality significantly impact the potential of the active suspension system to harvest power. Specifically, higher speeds and lower road quality lead to increased power harvesting potential. Beyond these factors, the policy or control logic of the suspension system also plays a critical role. In this research, the optimized control algorithm for each of the four suspension systems installed in the vehicle's corners has the potential to recover up to 650 watts of power on rough roads.

Table 3. Performance Comparison of the Optimized Control System and the Passive One

Parameter	Unit	Value			
Road Quality Based on ISO 8608	-	A	B	C	D
Vehicle Longitudinal Speed	km/h	100	75	50	25
Selected Mass Coefficient	-	1.5	1.5	1.5	1.5
Selected Inertia Coefficient	-	−0.5	−0.5	−0.5	−0.5
Selected Front Damping Ratio	-	0.6	0.11	0.16	0.18
Selected Rear Damping Ratio	-	0.27	0.25	0.32	0.14
Net Harvested Power	W	60	145	380	650
Comfort of Active System	-	0.23	0.33	0.55	0.67
Comfort of Passive System	-	0.35	0.59	0.99	1.49
Comfort Improvement	%	34	44	44	55

6 Conclusion

Active suspension systems face important challenges and shortcomings, such as high energy consumption, weight considerations, and a lack of economic justification. This research aims to address these imperfections and improve the performance of such systems. Specifically, the goal is to enhance ride comfort while also increasing energy harvesting capabilities. By doing so, this research seeks to contribute to the development of more efficient and effective vehicle suspension systems. In order to implement the method, firstly, a vehicle model available in Adams software was selected. For use in the optimization algorithm, an analytical model corresponding to the Adams model was extracted and validated. Based on meaningful concepts and parameters in vehicle dynamics, a logical search space was introduced, and finally, assuming driving maneuvers, a control structure was developed and optimized, using a genetic algorithm.

The method began by defining the search space, initial population, and fitness functions including ride comfort and energy harvesting. A wide range of ride tests were then conducted to implement the optimization algorithm, considering various road qualities and speeds. For each speed, the most appropriate set of control coefficients was selected. As a result, an optimized control was developed, which has a potential power recovery of up to 650 watts on rough roads. This optimized control leads to a 45% improvement in ride comfort. As a result, the study provides promising results for the development of more energy efficient and comfortable vehicles. In the subsequent stages of the study, the focus is on enhancing the yaw and roll stability of the vehicle in the handling maneuvers. To achieve this, an algorithm will be developed and optimized by considering the effect of the implemented actuators on the vehicle's yaw and roll dynamic behavior.

References

1. Yu, J., Vladimir, V.: Control Applications of Vehicle Dynamics. CRC Press, Boca Raton (2021). https://doi.org/10.1201/9781003134305
2. Sun, W., Gao, H., Shi, P.: Advanced Control for Vehicle Active Suspension Systems, vol. 204. Springer, Heidelberg (2020). https://doi.org/10.1007/978-3-030-15785-2
3. Múčka, P.: Energy-harvesting potential of automobile suspension. Veh. Syst. Dyn. **54**(12), 1651–1670 (2016). https://doi.org/10.1080/00423114.2016.1227077
4. Țoțu, V., Alexandru, C.: Multi-criteria optimization of an innovative suspension system for race cars. Appl. Sci. **11**(9), 4167 (2021). https://doi.org/10.3390/app11094167
5. Ataei, M., et al: Multi-objective optimization of a hybrid electromagnetic suspension system for ride comfort, road holding and regenerated power. J. Vibr. Control **23**(5), 782–793 (2017). https://doi.org/10.1177/1077546315585219
6. Yatak, M.Ö, Şahin, F.: Ride comfort-road holding trade-off improvement of full vehicle active suspension system by interval type-2 fuzzy control. Eng. Sci. Technol. Int. J. **24**(1), 259–270 (2021). https://doi.org/10.1016/j.jestch.2020.10.006
7. Williams, D.E.: Active suspension: future lessons from the past. SAE Int. J. Veh. Dyn. Stab. NVH **2**(10-02-02-0010), 147–165 (2018). https://doi.org/10.4271/10-02-02-0010
8. Wang, R., et al: Switching control of semi-active suspension based on road profile estimation. Veh. Syst. Dyn. **60**(6), 1972–1992 (2022). https://doi.org/10.1080/00423114.2021.1889621
9. Tseng, H.E., Hrovat, D.: State of the art survey: active and semi-active suspension control. Veh. Syst. Dyn. **53**(7), 1034-1062 (2015). https://doi.org/10.1080/00423114.2015.1037313
10. Deubel, C., Ernst, S., Prokop, G.: Objective evaluation methods of vehicle ride com-fort-a literature review. J. Sound Vibr. **548**, 117515 (2022). https://doi.org/10.1016/j.jsv.2022.117515
11. Du, Y., et al: A hierarchical framework for improving ride comfort of autonomous vehicles via deep reinforcement learning with external knowledge. Comput. Aided Civil Infrastruct. Eng. **38**, 1059–1078 (2022). https://doi.org/10.1111/mice.12934
12. ISO 2631: Mechanical vibration and shock -Evaluation of human exposure to whole-body vibration (1997)
13. Darabseh, T., Al-Yafeai, D., Mourad, A.H.I.: Energy harvesting from car suspension system: mathematical approach for half car model. J. Mech. Eng. Sci. **15**(1), 7695–7714 (2021). https://doi.org/10.15282/jmes.15.1.2021.07.0607
14. Lv, X. et al.: Research review of a vehicle energy-regenerative suspension system. Energies **13**(2), 441 (2020). https://doi.org/10.3390/en13020441
15. Abdelkareem, M.A.A., et al.: Vibration energy harvesting in automotive suspension system: A detailed review. Appl. Energy **229**, 672–699 (2018). https://doi.org/10.1016/j.apenergy.2018.08.030
16. Tulsian, N., Dewangan, S.: A discussion on energy harvesting through sus-pension system. Mater. Today Proc. **79**, 189–192 (2023). https://doi.org/10.1016/j.matpr.2022.10.052
17. Zhang, R., Wang, X., John, S.: A comprehensive review of the techniques on re-generative shock absorber systems. Energies **11**(5), 1167 (2018). https://doi.org/10.3390/en11051167
18. Bennis, F., Bhattacharjya, R.K. (eds.): Nature-inspired methods for metaheuristics optimization. MOST, vol. 16. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-26458-1
19. Llopis-Albert, C., Rubio, F., Zeng, S.: Multiobjective optimization framework for designing a vehicle suspension system. A comparison of optimization algo-rithms. Adv. Eng. Softw. **176**, 103375 (2023). https://doi.org/10.1016/j.advengsoft.2022.103375
20. Katoch, S., Chauhan, S.S., Kumar, V.: A review on genetic algorithm: past, present, and future. Multimedia Tools Appl. **80**(5), 8091–8126 (2020). https://doi.org/10.1007/s11042-020-10139-6

21. Bruni, S., et al.: State-of-the-art and challenges of railway and road vehicle dynamics with multibody dynamics approaches. Multibody Sys.Dyn. **49**, 1–32 (2020). https://doi.org/10.1007/s11044-020-09735-z
22. Widner, A, Tihanyi, V., Tettamanti, T.: Framework for vehicle dynamics model validation. IEEE Access 10, 35422–35436 (2022). https://doi.org/10.1109/ACCESS.2022.3157904
23. Rill, G., Castro, A.A: Road Vehicle Dynamics: Fundamentals and Modeling with MATLAB®. CRC Press, Boca Raton (2020). https://doi.org/10.1201/9780429244476
24. Bagchi, K., et al: Designing and optimizing an active car suspension with genetic algorithm. In: 7th International Conference on Environment Friendly Energies and Applica-tions (EFEA). IEEE (2022). https://doi.org/10.1109/EFEA56675.2022.10063744
25. Peng, D., et al: Multiobjective optimization of an off-road vehicle suspension parame-ter through a genetic algorithm based on the particle swarm optimization. Math. Probl. Eng., 1–14 (2021). https://doi.org/10.1155/2021/9640928
26. Hamersma, H.A., Els, P.S.: Vehicle suspension force and road profile pre-diction on undulating roads. Veh. Syst. Dyn. **59**(10), 1616–1642 (2021). https://doi.org/10.1080/00423114.2020.1774067
27. Múčka, P.: Simulated road profiles according to ISO 8608 in vibration analysis. J. Testing Eval. **46**(1), 405–418 (2017). https://doi.org/10.1520/JTE20140493
28. ISO 8608: Second Edition, Mechanical vibration — Road surface profiles — Reporting of measured data (2016)
29. Du, Y., Liu, C., Li, Y.: Velocity control strategies to improve auto-mated vehicle driving comfort. IEEE Intell. Transp. Syst. Mag. **10**(1), 8–18 (2018). https://doi.org/10.1155/2021/9445070
30. Čerškus, A., et al: Identification of road profile parameters from vehicle suspension dynamics for control of damping. Symmetry **13**(7), 1149 (2021). https://doi.org/10.3390/sym13071149

Demystifying Deep Learning Techniques in Knee Implant Identification

Shaswat Srivastava[1], A. Ramanathan[1], Puthur R. Damodaran[2], C. Malathy[1], M. Gayathri[1], and Vineet Batta[3](✉)

[1] School of Computing, SRM Institute of Science and Technology, Chennai, India
[2] Department of Orthopaedics, Apollo Speciality Hospital, OMR, Chennai, Tamil Nadu, India
[3] Department of Orthopaedics and Trauma, Luton, Dunstable University Hospital NHS Trust, Luton, UK

battavineet@doctors.org.uk

Abstract. Accurate identification of an orthopedic implant before a revision surgery is very important and helps both physicians and patients in numerous aspects. The proposed system uses a novel framework to identify five different total knee arthroplasty implants from plain X-ray images using Deep learning techniques. Anterior-Posterior and Lateral images are used together in this study to make identification much more accurate. The proposed system identifies five different knee implants with an accuracy of 86.25% and an Area Under curve of 0.974.

Keywords: Knee Implant · Orthopedics · Biomedical · Deep Learning · Medical Image

1 Introduction

Osteoarthritis (OA) is a condition of disability that is more common among older adults [1]. Symptoms include pain in joints, stiffness, and instability [2]. Total Knee arthroplasty (TKA) or primary total knee arthroplasty is the common surgical procedure to treat the final stage of Osteoarthritis [2, 3]. The total number of patients undergoing knee surgeries across the world is on the rise [4]. The number of primary TKA procedures globally is projected to grow by 85% by 2030 [5]. In Australia as per recorded data by 2019, around 9% of the population suffer from this musculoskeletal problem [2]. In the USA around 400,000 TKA's are performed every year [6] in Korea between 2010 to 2018, TKA increased by 35% [7]. Sometimes these TKA implants need to be revised or replaced completely, demanding a revision surgery. Revision of total joint arthroplasty is more challenging and has a higher risk compared to primary surgery. The volume of revision TKA is likely to increase in accordance with a rise in Primary TKA procedures [8]. The major causes of revision surgery include Periprosthetic fracture, implant loosening and instability [9], and infection [10]. In most countries such as Germany, Australia, the United Kingdom, and the United States of America revision TKA' s is increasing rapidly [10]. Implant Identification is the first and most important step before performing any

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 13–23, 2024.
https://doi.org/10.1007/978-3-031-56728-5_2

revision surgery. About 10% of the implants are difficult to identify before operation and 2% during the time of operation. Failure to identify an implant before revision surgery leads to higher blood and bone loss. It also paves the way for higher medical costs [11].

The proposed approach uses a novel framework to identify the make and model of an orthopedic implant from plain radiographs. X-ray images of implants were used in this study and the authors aim to automate the whole process of implant identification. Data augmentation techniques [12] are used to increase the count of images. Transfer learning-based [13] Deep learning (DL) approaches are incorporated using various pre-trained models to classify total knee arthroplasty implants from their appearance in X-ray images using both Anterior-Posterior (AP) and Lateral (LAT) views for 5 different knee implants.

2 Literature Review

Sukrit et al. classified six knee implant models using deep learning algorithms from 1078 radiograph images of both lateral (LAT) and anterior-posterior (AP) view images. After training all 6 implant models for 30 epochs, the model DenseNet201 achieved 96.38% accuracy [14]. Anjali et al. devised an algorithm to identify 6 different knee implants from X-ray images. Seven unique deep learning models, each trained for a total of 20 epochs. The best results were obtained using VGG16 with a precision of 98.4% and an accuracy of 95.5%. Both AP and LAT images were used for training and testing [15].

Smaranjit et al. proposed a framework for the identification of six such orthopedic knee implants. A dataset consisting of 878 radiographic images of orthopedic knee implants, encompassing both AP and LAT views, was utilized to train Deep Convolutional Neural Networks. The Results were compared for both augmented and unaugmented datasets in training. MobileNetV2 had the best accuracy compared to other deep learning models and showed a high test accuracy of 96.66% [16].

Belete et al. used a Convolutional Neural Network (CNN) trained on seven Total Knee Replacement (TKR) implants to automatically detect their make and model as well as identify the absence of a TKR, using plain-film radiography images. The dataset consisted of a total of 588 knee X-ray images in AP view, which were split randomly into training, testing, and testing sets in the ratio 50:25:25. The utilization of Simple Stochastic Gradient Descent (SGD) with 0.9 momentum and a consistent learning rate of 0.001 brought 100% accuracy was shown by all the models in the classification of prosthesis and no prosthesis [17].

3 Data Set Description

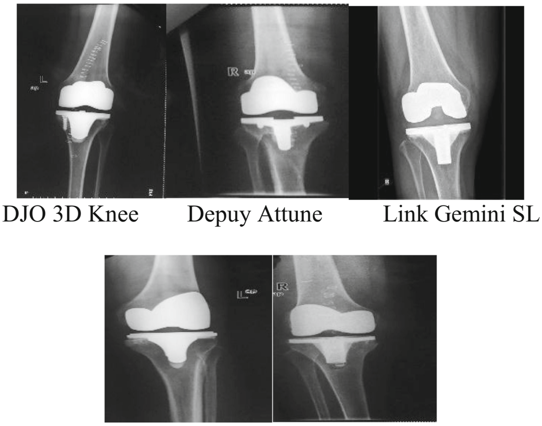
The study proposes to identify 5 different total knee arthroplasty implants from plain X-ray images. The images are collected from individual surgeons practicing in various hospitals in a completely anonymized manner. Patient's details or their health conditions are not present in the images both directly and indirectly. The collected X-ray images are labeled only for their make and model. Table 1 shows the list of implants used in this proposed study and the corresponding number of X-ray images of implants in both Anterior Posterior (AP) and Lateral View (LAT). The study utilized both AP and LAT images for this proposed work.

Table 1. A detailed description of the dataset used in the proposed study

Make of the Implant	Model of the Implant	Images in AP	Images in LAT
DJO	3D Knee	72	69
Depuy	Attune	202	149
Link	Gemini SL	99	97
Zimmer	LPS Flex Knee GSF	91	96
Microport	Medial Pivot	97	89

To avoid the class imbalance problem and to make the model more robust, except Depuy Attune, rest of the four knee implant images in both AP and LAT were initially mirrored to increase their count.

Figure 1 shows the raw AP images of the implants that were used in the study. These implants do not have any patient details.

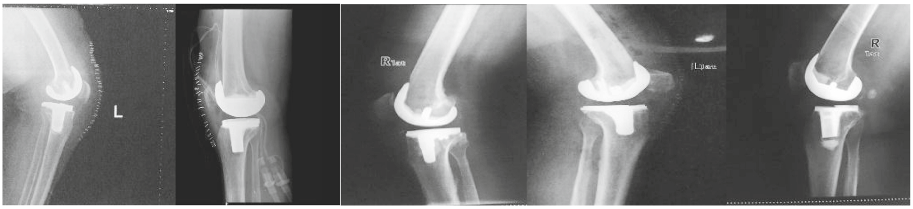


DJO 3D Knee Depuy Attune Link Gemini SL

Microport Medial Zimmer LPS

Fig. 1. AP images of implants used in the proposed work

Figure 2 shows the raw LAT images of the implants that were used in the study along with AP images.



DJO 3D Knee Depuy Attune Link Gemini SL Microport Zimmer LPS

Fig. 2. LAT images of the implants used in the proposed study

4 Methods and Methodology

4.1 Training Set

After finalizing the images in each implant class, a separation was made between the images as train and test. 70% of the images belonging to an implant were used for the training. This 70% is taken from both AP and LAT view images of an implant model randomly. These images further undergo augmentation and are used for the rest of the training.

4.2 Testing Set

The images that were used to measure the quality of training are called internal testing sets or validation set images. These images are 30% of the original dataset (balance images after taking 70%) in both AP and LAT views respectively. These images do not undergo any form of augmentation and remain an untouched set of images.

4.3 Data Augmentation

Augmentation methods are generally performed on training samples to increase their count which enhances the training of the DL model. Image augmentation techniques such as rotation, zoom-in, and zoom-out [12] were applied in a random manner to increase the count of images. The training of images was increased from a mere hundred to a few thousand because of this data augmentation.

4.4 Deep Learning Methods

The proposed work leverages various commonly used pre-trained deep learning models such as VGG16 [18], VGG19 [19], InceptionV3 [20], DenseNet 121 and DenseNet 201 [21]. An in-depth analysis of all the models was performed to obtain the best possible accuracy under each of the deep learning models. All these models underwent extreme hyperparametrs tuning and were tested for various epochs, optimizers, regularizers, learning rates, and batch sizes [22]. Optimizers such as SGD [23], Adam [24], and Adagrad [25] were used in the study with different learning rates to get an in-depth analysis of the performance of all the deep learning models.

4.5 Proposed Model

The best-performing combination was obtained with the VGG 19 deep learning model. The general architectural diagram of VGG19 is shown in Fig. 3. The proposed architecture utilizes VGG 19 as its underlying model, which undergoes fine-tuning based on the input images. VGG 19 is a deep neural network consisting of 19 layers. The model

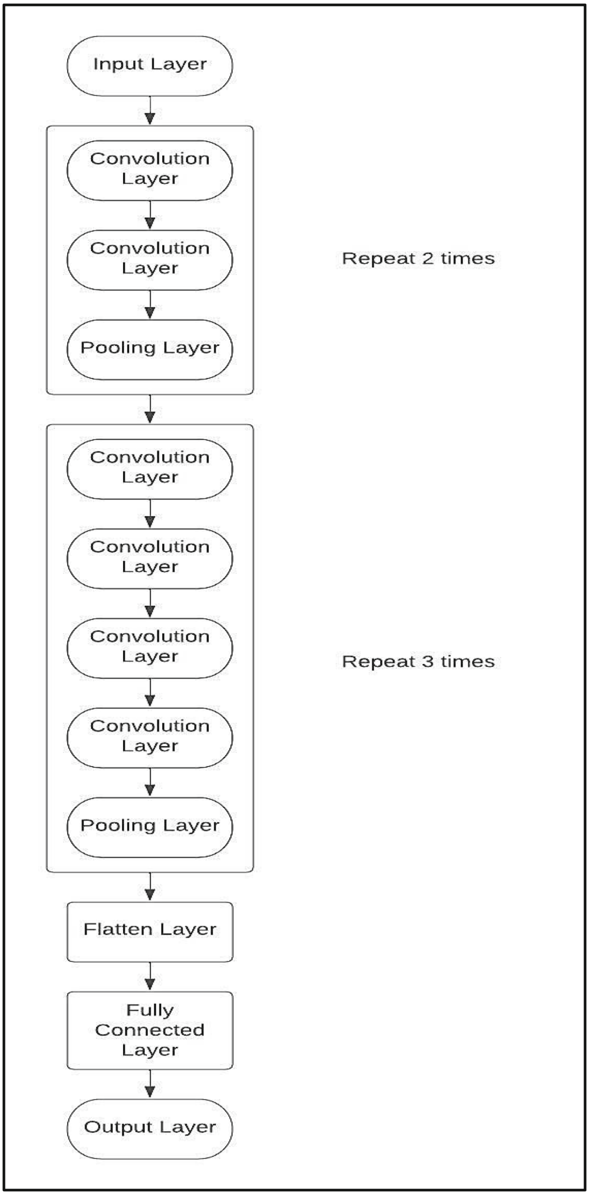


Fig. 3. Block diagram of VGG 19 Architecture

requires input images of size (224, 224). It comprises 2 convolutional layers with a kernel size of 3 × 3 and 64 channels, followed by a pooling layer with a pool size of 2 × 2 and a stride of 2. This sequence is repeated once more.

The output from the pooling layer is then passed through four convolutional layers with a kernel size of 3 × 3 and 64 channels, followed by a pooling layer. This complete cycle repeats itself three more times. The final output is flattened and fed into a pooling layer (Max Pooling) with 512 channels, which is then connected to a fully connected layer. The output is seen in the output layer which denotes the classification of Knee implants. The study employs the 'softmax' activation function and 'Adagrad' as the optimizer with a learning rate of 0.0001.

4.6 Performance Metrics

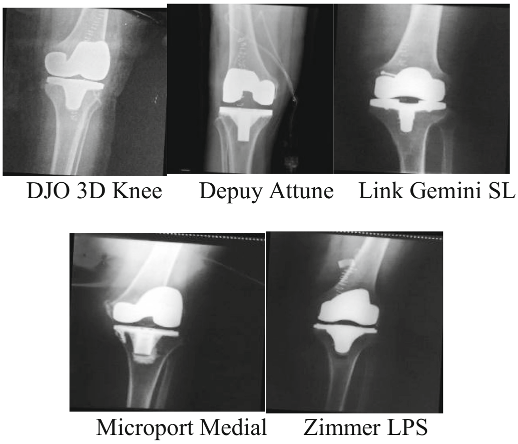
To understand and analyze the individual deep learning model, the study determines the performance metrics of each of the deep learning models. The metrics include precision, recall, F1 score, and Area under Curve (AUC) [26]. These values are obtained from a confusion matrix plot which is supported by true positive, false negative, true negative, and false positive [27].

5 Results and Discussions

The results of data augmentation and deep learning are as follows:

5.1 Data Augmentation

Figure 4 shows the zoomed-out and rotated AP images across 5 different implant classes. These images were used only for training the model.

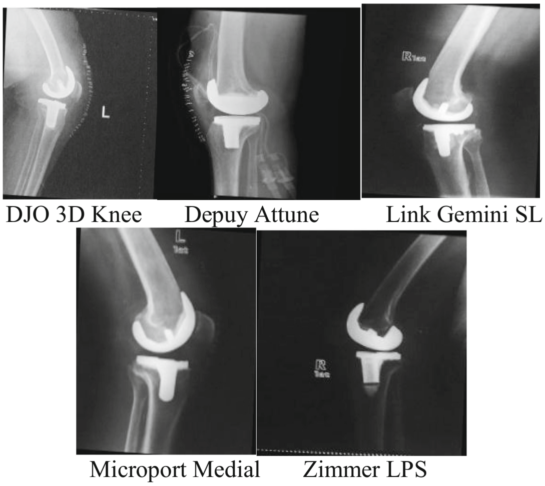


DJO 3D Knee Depuy Attune Link Gemini SL

Microport Medial Zimmer LPS

Fig. 4. Augmented AP images of implants used in the proposed work

Figure 5 shows the zoomed-out and rotated LAT images across 5 different implant classes. These images were used only for training the model along with AP images.



DJO 3D Knee Depuy Attune Link Gemini SL

Microport Medial Zimmer LPS

Fig. 5. Augmented LAT images of implants used in the proposed work

5.2 Deep Learning Results

Table 2. Results obtained using Deep Learning Models

DL Model	Epochs	Loss in Training	Accuracy in Training (%)	Loss in Test	Accuracy in Testing (%)
VGG 16	20	0.2024	97.44%	0.5751	80.00%
VGG 19	20	0.0019	99.98%	0.6911	86.25%
DenseNet 121	20	0.0218	99.94%	0.6255	80.94%
DenseNet 201	20	0.0581	99.65%	0.5563	80.94%
Inception V3	20	0.0190	99.99%	0.7493	78.12%

Table 2 shows the best-obtained results across different deep learning models in classifying five knee implants. The model VGG19 performs better with an accuracy of 86.25% and outperforms the other tested deep learning models.

Table 3 shows the best obtained performance metrics across different deep learning models in classifying five knee implants. The model VGG19 performs better with an F1 score of 0.8648 and an AUC of 0.9747 and outperforms the other deep learning models in terms of performance.

Table 3. Performance Metrics across Various Deep Learning Models

DL Model	Epochs	Precision	Recall	F1 Score	AUC
VGG 16	20	0.8040	0.8000	0.7998	0.9554
VGG 19	20	0.8750	0.8625	0.8648	0.9747
DenseNet 121	20	0.8224	0.8093	0.8070	0.9605
DenseNet 201	20	0.8170	0.8093	0.7955	0.9718
Inception V3	20	0.7998	0.7812	0.7815	0.9506

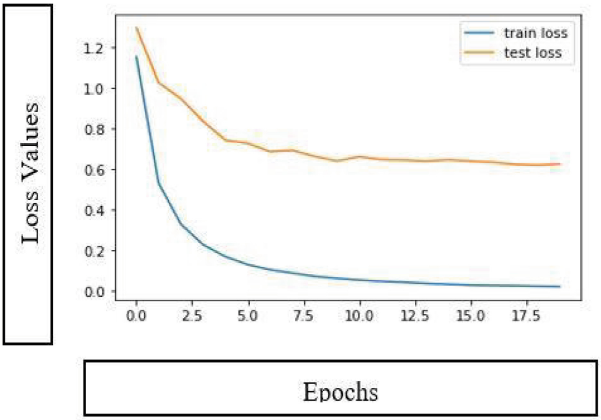


Fig. 6. Plot of Train and Test Loss

Figure 6 shows the graphs of test and train loss for the best-performing VGG19 model. The loss for both training and testing were coming down continuously indicating the model is making good predictions.

Figure 7 shows the graphical representation of both the train and test accuracy for the best-performing VGG19 model. The accuracy for both training and testing was increasing continuously indicating the model is getting trained well.

Figure 8 shows the plot of the confusion matrix for the best-performing VGG19 DL model. Link Gemini SL implant suffers slightly higher misclassification thereby reducing the overall accuracy of the proposed system.

In VGG16, Adam was overfitting [28] within a few epochs in most of the runs and Adagrad was giving lesser accuracy with a learning rate of 0.01 and 0.001. SGD was overfitting with higher epochs and eventually increased the test loss. This was the same with the VGG 19 model. On the other hand, DenseNet121 and 201 and InceptionV3 with Adam gave good accuracy but poor graphs due to overfitting which produces increased test loss. DensNet121 and DenseNet201 showed slightly better performance with SGD

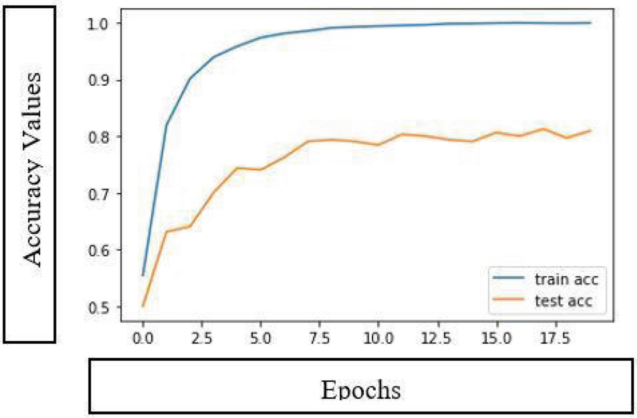


Fig. 7. Plot of Train and Test Accuracy

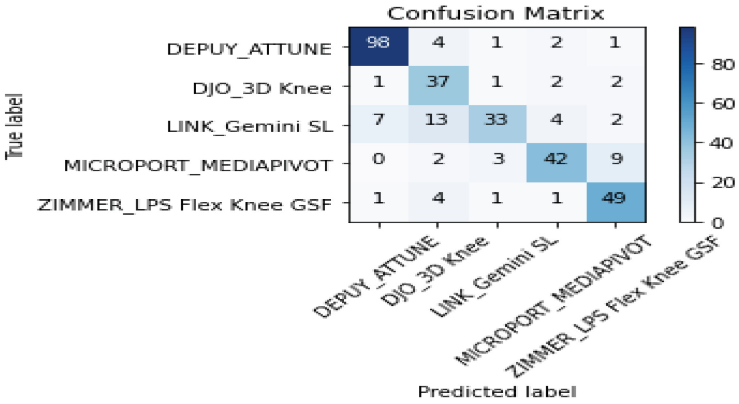


Fig. 8. Plot of Confusion Matrix

optimizer compared to Adagrad with a learning rate of 0.0001. However, for InceptionV3, Adagrad was still best performing. Overall across all deep learning models used in this study, Adagrad optimizer with enhanced fine tuning performs better producing better results in most of the deep learning models.

The best results (VGG19) were obtained with the Adagrad optimizer and with a learning rate of 0.00001 after various trial attempts. The proposed work uses different implants altogether that are not used in a direct single study [14–17] and also gives an in-depth analysis of all the deep learning models used in this study.

6 Conclusion

The proposed work uses deep learning-based classification of implants using various pretrained models. All the models were highly fine-tuned to get the best possible results. The system identifies the five different total knee replacement implant models with an accuracy of 86.25% using both AP and LAT views of X-ray images. The work can be further extended by adding more implant classes and by tuning the model even deeper.

References

1. Feng, J.E., Novikov, D., Anoushiravani, A.A., Schwarzkopf, R.: Total knee arthroplasty: improving outcomes with a multidisciplinary approach. J. Multidisciplinary Healthcare, 63–73 (2018)
2. Ditton, E., et al.: Improving patient outcomes following total knee arthroplasty: identifying rehabilitation pathways based on modifiable psychological risk and resilience factors. Front. Psychol. **11**, 1061 (2020)
3. Malkani, A.L., et al.: The difficult primary total knee arthroplasty. Instr. Course Lect. **65**, 243–265 (2016)
4. Lee, D.W., et al.: Automated detection of surgical implants on plain knee radiographs using a deep learning algorithm. Medicina **58**(11), 1677 (2022)
5. Gao, J., Xing, D., Dong, S., Lin, J.: The primary total knee arthroplasty: a global analysis. J. Orthop. Surg. Res. **15**, 1–12 (2020)
6. Varacallo, M., Luo, T.D., Johanson, N.A.: Total knee arthroplasty techniques (2018)
7. Kim, T.W., Kang, S.B., Chang, C.B., Moon, S.Y., Lee, Y.K., Koo, K.H.: Current trends and projected burden of primary and revision total knee arthroplasty in Korea between 2010 and 2030. J. Arthroplasty **36**(1), 93–101 (2021)
8. Dy, C.J., Bozic, K.J., Padgett, D.E., Pan, T.J., Marx, R.G., Lyman, S.: Is changing hospitals for revision total joint arthroplasty associated with more complications? Clin. Orthop. Relat. Res. **472**, 2006–2015 (2014)
9. Lee, D.H., Lee, S.H., Song, E.K., Seon, J.K., Lim, H.A., Yang, H.Y.: Causes and clinical outcomes of revision total knee arthroplasty (2017)
10. Postler, A., Lützner, C., Beyer, F., Tille, E., Lützner, J.: Analysis of total knee arthroplasty revision causes. BMC Musculoskelet. Disord. **19**, 1–6 (2018)
11. Borjali, A., Chen, A.F., Muratoglu, O.K., Morid, M.A., Varadarajan, K.M.: Detecting total hip replacement prosthesis design on plain radiographs using deep convolutional neural network. J. Orthop. Res. **38**(7), 1465–1471 (2020)
12. Goceri, E.: Medical image data augmentation: techniques, comparisons, and interpretations. Artif. Intell. Rev. **56**, 12561–12605 (2023)
13. Hosna, A., Merry, E., Gyalmo, J., Alom, Z., Aung, Z., Azim, M.A.: Transfer learning: a friendly introduction. J. Big Data **9**(1), 102 (2022)
14. Sharma, S., et al.: Knee Implant Identification by Fine-Tuning Deep Learning Models (2021)
15. Tiwari, A., Yadav, A.K., Bagaria, V.: Application of deep learning algorithm in the automated identification of knee arthroplasty implants from plain radiographs using transfer learning models: are algorithms better than humans? (2022)
16. Ghose, S., Datta, S., Batta, V., Malathy, C., Gayathri, M.: Artificial intelligence-based identification of total knee arthroplasty implants. In: 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), pp. 302–307. IEEE (2020)

17. Belete, S.C., Batta, V., Kunz, H.: Automated classification of total knee replacement prosthesis on plain film radiograph using a deep convolutional neural network. Inform. Med. Unlocked **25**, 100669 (2021)
18. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
19. Sudha, V., Ganeshbabu, T.R.: A convolutional neural network classifier VGG-19 architecture for lesion detection and grading in diabetic retinopathy based on deep learning. CMC-Comput. Mater. Continua **66**(1), 827–842 (2021)
20. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)
21. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708 (2017)
22. Yu, T., Zhu, H.: Hyper-parameter optimization: a review of algorithms and applications. arXiv preprint arXiv:2003.05689 (2020)
23. Ruder, S.: An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747 (2016)
24. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
25. Lydia, A., Francis, S.: Adagrad—an optimizer for stochastic gradient descent. Int. J. Inf. Comput. Sci. **6**(5), 566–568 (2019)
26. Vakili, M., Ghamsari, M., Rezaei, M.: Performance analysis and comparison of machine and deep learning algorithms for IoT data classification. arXiv preprint arXiv:2001.09636 (2020)
27. Ramanathan, A., Christy Bobby, T.: Classification of corpus callosum layer in mid-saggital MRI images using machine learning techniques for autism disorder. In: Saha, S., Nagaraj, N., Tripathi, S. (eds.) MMLA 2019. CCIS, vol. 1290, pp. 78–91. Springer, Singapore (2020). https://doi.org/10.1007/978-981-33-6463-9_7
28. Xiao, M., et al.: Addressing the overfitting problem in deep learning-based solutions for next-generation data-driven networks. Wirel. Commun. Mob. Comput. **2021**, 1–10 (2021)

Artificial Neural Network Model of Nonlinear Behavior of Micro-ring Gyroscopes

Hassan Sayyaadi(✉) and Mohammad Ali Mokhtari Amir Majdi

Department of Mechanical Engineering, Sharif University of Technology, Tehran, Iran
sayyaadi@sharif.edu

Abstract. The investigation is concerned with developing a neural network tomodel a micro-ring gyroscope considering the exact term for the electrical force. To this end, Hamilton's principle alongside Ritz's method has been utilized to obtain the non-linear system of equations governing the dynamics of the micro-ring. The equations are then numerically solved using the fourth-order Runge-Kutta method. It has been observed considering the Taylor series expansion of the electrical force may lead to misleading results in the case of large deformations. Furthermore, gathering a dataset from numerical solutions induces high computational costs and time. So, a fast method is required to obtain a sizable dataset with good accuracy. To this end, the system has been modelled with an artificial neural network. To form the neural network 1720 examples have been gathered with five input features. It is shown that the proposed neural network can perfectly predict the behavior of the micro-ring gyroscope.

Keywords: Neural Network · Ring Micro-Gyroscope · Geometric Nonlinearities · Full Electrical Force Expression

1 Introduction

As the current technologies allow the fabrication of miniature systems, features of micro-electro-mechanical systems (MEMS) such as low power consumption have invited many researchers to study the behavior of those systems [1–5]. Micro-inertial sensors, as one of the many types of MEMS, constitute a considerable portion of these studies [6–9].

A micro-gyroscope is a micro-inertial sensor that measures the angular velocity of a device. Such devices have been recorded to have expanding applications in the market of micro-devices [10]. Micro-gyroscopes have different working principles [11–14]. Vibrating micro-gyroscopes are among the more prevalent types of micro-gyroscopes. Such micro-devices utilize the Coriolis effect to correlate two different vibrating modes of a mechanical element like micro-beams [15] and micro-rings [16].

Micro-rings are a common type of mechanical element that has been implemented in modeling the performance of vibratory micro-gyroscopes. Hence, many researchers have been directed into investigating the behavior of such micro-devices under various working conditions. Barakat et al. [17] propose a model to further exploit the parametric excitation and amplification of micro-ring gyroscopes. To this end, the micro-gyroscope is considered to have two simultaneous phase-lagged parametric excitations.

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 24–34, 2024.
https://doi.org/10.1007/978-3-031-56728-5_3

They obtained the equations of motion by utilizing Hamilton's principle and studied the behavior of both extensible and inextensible micro-rings. They further claim that including the phase lag may lead to new possibilities in designing micro-gyroscopes with higher sensitivity. Polunin and Shaw [18] developed a fully non-linear model for the micro-ring gyroscopes and showed that these dynamical nonlinearities can enhance the performance of previous models. Although they have proposed a fully non-linear model, they have neglected some of the terms in their model to obtain the solution. To completely study the non-linear term introduced by [18], Liang et al. [19] included all of these terms in their model and numerically solved the nonlinear system of equations governing the dynamics of the micro-ring gyroscope.

With the advancement of computing units, many works have been dedicated to implement neural networks to model the behavior of various systems such as MEMS [8, 20]. Chong et al. [21] utilized Elman neural networks to enhance the precision of temperature drift modelling in a tuning fork micro-gyroscope. They also used genetic algorithm to adjust the parameters in the Elman neural network. Han et al. [22] employed a recurrent neural network to reduce the noise levels in a micro-electro-mechanical measurement unit. Shao and Shi [23] developed a feedback control system based off of neural networks to control the disturbances levels exerted on a micro-gyroscope by model uncertainties, outside disturbances, and dynamic couplings.

The current study focuses on developing a neural network model for micro-ring gyroscopes considering the full electrical force expression. To this end, reduced governing equations of micro-ring are presented in Sect. 2. Section 3 is dedicated to the investigation of the results of the model, including validation, time history of the system, and neural network model of the system. It has been found out that the previous works on such systems commonly have utilized the Taylor series expansion of the electrical force that will result in misleading outputs as will be shown further in the study. On the other hand, considering the full for-force terms along with non-linear couplings between drive and sense modes of the micro-ring will impose excessive computational costs. Hence, utilizing neural networks to model the performance of the micro-ring gyroscope can persevere both accuracy and low computational costs. Hence, future investigations can potentially have more freedom in designing a reliable micro-ring gyroscope with varying parameters. Finally, the conclusion of the investigation is presented in Sect. 4.

2 Theoretical Formulation

The current study concentrates on modeling the performance of a micro-ring gyroscope with neural networks. To this end, Fig. 1 shows the schematics of the micro-ring. In the model, the electrical force is applied through electrodes all along the micro-ring. The AC voltage is considered to vary around the ring. As can be observed from Fig. 1a, the ring rotates along its axis with an angular velocity of Ω. Furthermore, the average radius of the micro-ring is denoted by $\overline{R}$. It is noteworthy that the second in-plane bending mode of the micro-ring is excited in the present study. Hence, the drive and sense modes have a phase difference as can be observed from Fig. 1a. From Fig. 1b, one can infer that the thickness and width of the micro-ring are denoted by h, and b, respectively.

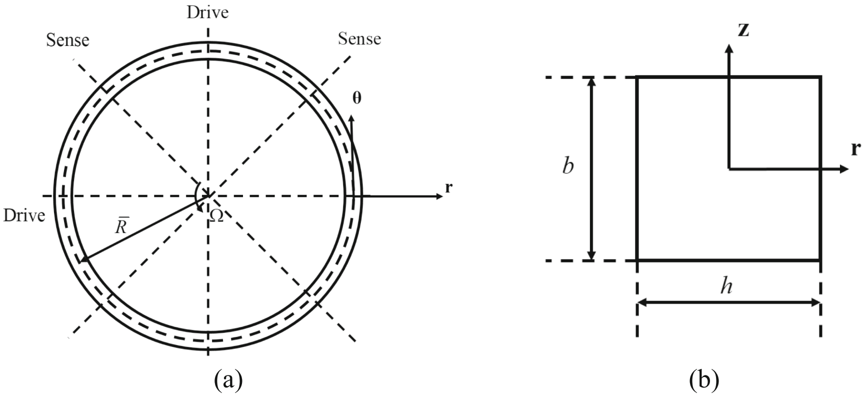


Fig. 1. Schematics of the micro-ring (a) top view, and (b) cross-section

Shaw and Polunin has driven the equations of motion for the elliptical modes of a micro-ring considering the geometric non-linearities [18]. Based on their model, the circumferential strain of the micro-ring can be presented as

$$\varepsilon_{\theta\theta} = \frac{u}{r} + \frac{\partial v}{r\partial\theta} + \frac{1}{2}\left(\frac{\partial u}{r\partial\theta}\right)^2 + \frac{u}{r}\frac{\partial v}{r\partial\theta} \tag{1}$$

where u and v represent the displacement of the neutral axis of the micro-ring in the r and θ directions, respectively. It is worth mentioning that the mid-line stretching is considered to be negligible. Hence, the strain energy can be simplified to [18]

$$U_r = \frac{EI}{2\overline{R}^3}\int_0^{2\pi}\left[u + u'' - \frac{1}{2\overline{R}}\left(u'\right)^2\right]^2 d\theta \tag{2}$$

where the prime sign denotes the differentiation with respect to θ.

The kinetic energy of the micro-ring is [18]

$$T = \frac{1}{2}\int \rho\left[(\dot{u} - v\Omega)^2 + (\dot{v} + (r+u)\Omega)^2\right] r\ dr\ d\theta\ dz \tag{3}$$

in which the symbol ρ denotes the density of the micro-ring material. In addition, to show the differentiation with respect to time, the dot sign has been implemented.

The electrical potential exerted on the micro-ring can be written as [18]

$$U_e = -\frac{\varepsilon_0}{2}\int_0^{2\pi} b\overline{R}\frac{(V_{DC} + V_{AC}(\theta,t))^2}{d-u} d\theta \tag{4}$$

in the Eq. the initial gap between electrodes and electrical permittivity of the air are dented by d and ε_0, respectively. In It is worth mentioning that the thickness of the micro-ring is much lower than its average radius, the average radius is used in (4) instead of the outer radius of the micro-ring.

The AC voltage is considered to vary trigonometrically along the ring and its relation is given in (5)

$$V_{AC}(\theta, t) = V_1 \cos(2\theta)\cos\big(\Omega_f t\big) \tag{5}$$

where V_1 and Ω_f are the AC voltage amplitude and excitation frequency respectively.

The micro-ring is inextensible since the thickness of the micro-ring is much less than the wavelength of the elliptical vibrating mode of the micro-ring [18]. In view of this fact, the mid-line stretching is negligible [18]. As a result, the expressions for the elliptical modes of the micro-ring in the sense and drive directions can be obtained as [19]

$$u = A(t)\cos(2\theta) + B(t)\sin(2\theta) - \Big(A^2(t) + B^2(t)\Big)\Big/\overline{R} \tag{6}$$

$$\begin{aligned} v = {} & [-A(t)\sin(2\theta) + B(t)\cos(2\theta)]\big/2 - \Big\{\Big(A^2(t) - B^2(t)\Big)\sin(4\theta) \\ & -2A(t)B(t)\cos(4\theta)\Big\}\Big/4\overline{R} \end{aligned} \tag{7}$$

in which the elliptical modes of the micro-ring in the drive and sense directions corresponding to the second in-plane bending mode of the ring are denoted by A and B respectively.

Upon substitution of (2), (3), and (4) into Hamilton's Principle; utilizing modal relations from (6) and (7) along with the Ritz method will lead to the governing equations of micro-ring.

$$\begin{aligned} & \ddot{A}\left[1 + \frac{3}{5\overline{R}^2}\Big(11A^2 + B^2\Big)\right] + \dot{A}\left[2C_A^* + \frac{6}{5}\frac{B\dot{B}}{\overline{R}^2}\right] \\ & \quad + A\left[\omega_0^2 + \Omega^2\left(\frac{11}{5} - \frac{37}{10}\frac{B^2}{\overline{R}^2}\right) + \eta\frac{B^2}{\overline{R}^2} + \frac{33}{5}\frac{\dot{A}^2}{\overline{R}^2} + \frac{31}{5}\frac{\dot{B}^2}{\overline{R}^2} + \frac{34}{5}\frac{B\ddot{B}}{\overline{R}^2}\right] \\ & \quad + \frac{A^3}{\overline{R}^2}\left[\eta - \frac{33}{10}\Omega^2\right] - \frac{16}{5}\frac{\dot{B}}{\overline{R}^2}\Omega A^2 = \frac{8}{5}\Omega\dot{B}\left(1 - 2\frac{B^2}{\overline{R}^2}\right) + F_A(A, B, t) \end{aligned} \tag{8}$$

$$\begin{aligned} & \ddot{B}\left[1 + \frac{3}{5\overline{R}^2}\Big(11B^2 + A^2\Big)\right] + \dot{B}\left[2C_B^* + \frac{6}{5}\frac{A\dot{A}}{\overline{R}^2}\right] \\ & \quad + B\left[\omega_0^2 + \Omega^2\left(\frac{11}{5} - \frac{37}{10}\frac{A^2}{\overline{R}^2}\right) + \eta\frac{A^2}{R^2} + \frac{33}{5}\frac{\dot{B}^2}{\overline{R}^2} + \frac{31}{5}\frac{\dot{A}^2}{\overline{R}^2} + \frac{34}{5}\frac{A\ddot{A}}{\overline{R}^2}\right] \\ & \quad + \frac{B^3}{\overline{R}^2}\left[\eta - \frac{33}{10}\Omega^2\right] - \frac{16}{5}\frac{\dot{A}}{\overline{R}^2}\Omega B^2 = -\frac{8}{5}\Omega\dot{A}\left(1 - 2\frac{A^2}{\overline{R}^2}\right) + F_B(A, B, t) \end{aligned} \tag{9}$$

in which

$$\begin{aligned} & \omega_0^2 = \frac{1}{5\rho}\left(3\frac{E\,h^2}{\overline{R}^4}\right), \quad \eta = \frac{6\overline{R}^2}{5\rho}\left(\frac{Eh^2}{\overline{R}^6}\right), \quad M = \frac{5}{4}\pi\,\rho\,b\,h\,\overline{R} \\ & F_A(A, B, t) = \frac{\varepsilon_0}{2M}\int_0^{2\pi} b\overline{R}\frac{(V_{DC} + V_{AC}(\theta, t))^2(\cos(2\theta) + 2A)}{(d - u)^2}d\theta \\ & F_B(A, B, t) = \frac{\varepsilon_0}{2M}\int_0^{2\pi} b\overline{R}\frac{(V_{DC} + V_{AC}(\theta, t))^2(\sin(2\theta) + 2B)}{(d - u)^2}d\theta \end{aligned} \tag{10}$$

In which the structural damping coefficients in the drive and sense directions, and the Young modulus of the micro-ring are denoted by C_A^*, C_B^*, and E, respectively.

The non-dimensional parameters are introduced

$$\hat{A} = \frac{A}{\overline{\overline{R}}},\ \ \hat{B} = \frac{B}{\overline{\overline{R}}},\ \ \hat{t} = \frac{t}{\tilde{t}} = t\omega_0 \tag{11}$$

Substituting normalized parameters from (11) into (8) and (9); and dropping the hats will lead to

$$\begin{aligned}
&\overline{\varsigma}_{A1}\ddot{A} + \overline{\varsigma}_{A2}A + \overline{\varsigma}_{A3}\dot{A} + \overline{\varsigma}_{A4}\dot{B} + \overline{\varsigma}_{A5}A^2\ddot{A} + \overline{\varsigma}_{A6}A\dot{A}^2 + \overline{\varsigma}_{A7}A^3 + \overline{\varsigma}_{A13}\dot{B}B^2 \\
&\quad + \overline{\varsigma}_{A8}\dot{A}B\dot{B} + \overline{\varsigma}_{A9}AB^2 + \overline{\varsigma}_{A10}A\dot{B}^2 + \overline{\varsigma}_{A11}AB\ddot{B} + \overline{\varsigma}_{A12}\dot{B}A^2 = \\
&\quad \frac{\varepsilon_0}{2M\overline{\overline{R}}}\int_0^{2\pi} b\frac{(V_{DC} + V_{AC}(\theta,t))^2(\cos(2\theta) + 2A)}{\left(\frac{d}{\overline{\overline{R}}} - A\cos(2\theta) - B\sin(2\theta) + A^2 + B^2\right)^2}d\theta
\end{aligned} \tag{12}$$

$$\begin{aligned}
&\overline{\varsigma}_{B1}\ddot{B} + \overline{\varsigma}_{B2}B + \overline{\varsigma}_{B3}\dot{B} + \overline{\varsigma}_{B4}\dot{A} + \overline{\varsigma}_{B5}\dot{B}A\dot{A} + \overline{\varsigma}_{B6}BA^2 + \overline{\varsigma}_{B7}B\dot{A}^2 + \overline{\varsigma}_{B8}BA\ddot{A} \\
&\quad + \overline{\varsigma}_{B9}B\dot{B}^2 + \overline{\varsigma}_{B10}\dot{A}B^2 + \overline{\varsigma}_{B11}\dot{A}A^2 = \\
&\quad \frac{\varepsilon_0}{2M\overline{\overline{R}}}\int_0^{2\pi} b\frac{(V_{DC} + V_{AC}(\theta,t))^2(\sin(2\theta) + 2B)}{\left(\frac{d}{\overline{\overline{R}}} - A\cos(2\theta) - B\sin(2\theta) + A^2 + B^2\right)^2}d\theta
\end{aligned} \tag{13}$$

where

$$\begin{aligned}
&\overline{\varsigma}_{A1} = \overline{R}\Big/\tilde{t}^2,\ \ \overline{\varsigma}_{A2} = \left(\omega_0^2 + \frac{11}{5}\Omega^2\right)\overline{R},\ \ \overline{\varsigma}_{A3} = 2\frac{C_A^*\overline{R}}{\tilde{t}},\ \ \overline{\varsigma}_{A4} = -\frac{8}{5}\overline{R}\frac{\Omega}{\tilde{t}} \\
&\overline{\varsigma}_{A5} = \frac{33}{5}\frac{\overline{R}}{\tilde{t}^2},\ \ \overline{\varsigma}_{A6} = \frac{33}{5}\frac{\overline{R}^2}{\tilde{t}^2},\ \ \overline{\varsigma}_{A7} = \eta\overline{R},\ \ \overline{\varsigma}_{A8} = \frac{6}{5}\frac{\overline{R}^2}{\tilde{t}^2},\ \ \overline{\varsigma}_{A10} = \frac{31}{5}\frac{\overline{R}}{\tilde{t}^2} \\
&\overline{\varsigma}_{A9} = -\frac{33}{10}\overline{R}\Omega^2 + \eta\overline{R},\ \ \overline{\varsigma}_{A11} = \frac{34}{5}\frac{\overline{R}^2}{\tilde{t}^2},\ \ \overline{\varsigma}_{A12} = -\frac{16}{5}\frac{\overline{R}}{\tilde{t}}\Omega,
\end{aligned} \tag{14}$$

$$\begin{aligned}
&\overline{\varsigma}_{B1} = \frac{\overline{R}}{\tilde{t}^2},\ \ \overline{\varsigma}_{B2} = \left(\omega_0^2 + \frac{11}{5}\Omega^2\right)\overline{R},\ \ \overline{\varsigma}_{B3} = 2C_B^*\frac{\overline{R}}{\tilde{t}},\ \ \overline{\varsigma}_{B4} = \frac{8}{5}\Omega\frac{\overline{R}}{\tilde{t}}, \\
&\overline{\varsigma}_{B5} = \frac{6}{5}\frac{\overline{R}}{\tilde{t}^2},\ \ \overline{\varsigma}_{B6} = \left(-\frac{37}{10}\Omega^2 + \eta\right)\overline{R},\ \ \overline{\varsigma}_{B7} = \frac{31}{5}\frac{\overline{R}}{\tilde{t}^2},\ \ \overline{\varsigma}_{B8} = \frac{34}{5}\overline{R}\Big/\tilde{t}^2, \\
&\overline{\varsigma}_{B9} = \frac{33}{5}\overline{R}\Big/\tilde{t}^2, \overline{\varsigma}_{B10} = -\frac{16}{5}\frac{\overline{R}}{\tilde{t}}\Omega,\ \ \overline{\varsigma}_{B11} = -\frac{16}{5}\frac{\overline{R}}{\tilde{t}}\Omega
\end{aligned} \tag{15}$$

In the next sections, a neural network has been developed to predict the behaviour of the micro-ring gyroscope with governing equations presented at (14) and (15).

3 Results and Discussions

3.1 Comparison and validation

To check the validity of the model, a comparison between the obtained steady state amplitudes with the available data in the literature has been made. Fig. 2 demonstrates this comparison. From Fig. 2 one can infer that the results of model with Taylor series expansion of the electrostatic force is in good agreement with those of Ref [19]. In view of the full force expression, the steady state amplitudes in the sense directions is in good agreement with reported results in [19] when the initial gap between the electrodes is small but as the amplitude increases the difference between the model with Taylor series expansion and full electrical force becomes considerable. So, it is convenient to assume that considering fourth-order Taylor series expansion of the electrical force may result in misleading values of steady state amplitudes. On the other hand, taking the full electrical force expression into account will lead to high computational costs. Hence, a neural network configuration of the model is needed to capture both the accuracy of the electrical force and low computational costs of Taylor series expansion model. To this end, a detailed explanation on the structure and parameters of the neural network will be presented further in this study.

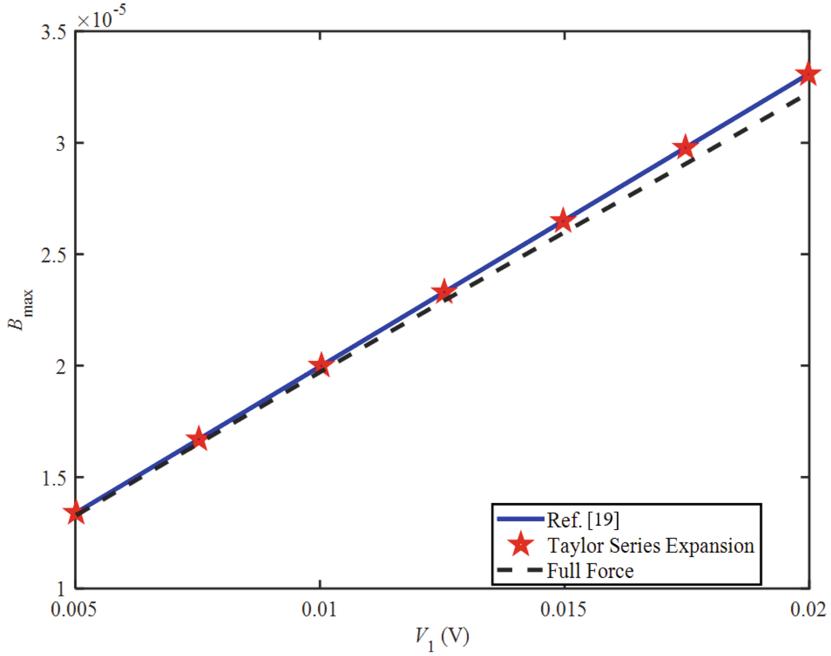


Fig. 2. Comparison of the maximum steady-state amplitude of the micro-ring gyroscope in the sense direction with Ref. [19]

3.2 Time Response of Micro-gyroscope

The time response of micro-ring gyroscope in the sense and drive directions is presented in Fig. 3. To obtain Fig. 3, the parameters $\overline{\varsigma}_{A3}/\overline{\varsigma}_{A1}$, $\overline{\varsigma}_{B3}/\overline{\varsigma}_{B1}$, V_{DC}, V_1, Ω_r, and $\omega_f = \Omega_f/\overline{\omega}_0$ to 0.1, 0.1, 30 V and 1 V, 0.3 and 0.3, respectively. Since the Coriolis effect induces a vibration in the sense direction, one can correlate the maximum amplitude of sense direction to the angular velocity of the micro-ring. A general procedure is to solve the governing equations of the micro-ring for a set of parameters and a specific value of the angular velocity to calculate the maximum steady state amplitude of the micro-ring in the sense direction. By repeating this step for different values of the parameters and angular velocities, one can acquire the table that shows the maximum steady state amplitude with its corresponding rotational velocity. In view of this fact, the next section is dedicated to build a neural network with different input features and maximum steady state amplitude in the sense direction as the output. It is to be noted that the material and geometric properties of the micro-ring are given in the Tables 1 and 2.

Table 1. The mechanical properties of the materials

Material property	Ring
E (GPa)	160
ρ(Kg/m^3)	2320

Table 2. Geometric properties of the system

d (μm)	h (μm)	b (μm)	L (μm)	R (μm)
3	5	30	110	300

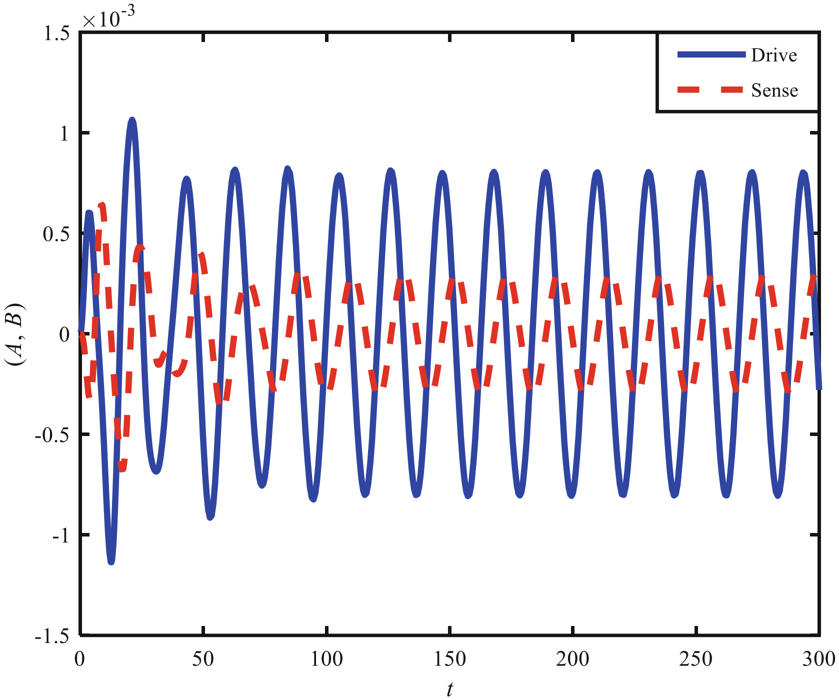


Fig. 3. Time response of the micro-ring in the drive and sense directions

3.3 Neural Network

As a supervised machine learning methods, artificial neural network has been implemented in order to predict the maximum steady state amplitude of the micro-ring gyroscope in the sense direction. The goal of the design is to achieve the weights and biases for each layer of the neural network to minimize the difference between the actual and predicted value of maximum steady state deflection of the micro-ring in the sense direction. To do so, a neural network with five hidden layers have been utilized for the input data with five features. These features are the non-dimensional angular velocity, normalized actuation frequency, DC voltage, AC voltage, and the ratio of thickness h to the average radius $\overline{R}$, which will be denoted by r_1 hereafter. It is to be noted that each of the hidden layers is considered to be fully connected to the previous and next layers. Furthermore, the activation functions for each of the hidden layer and the output layer are set to be tangent sigmoid functions and linear functions, respectively. It is worth mentioning that the number of neurons in each of the layers are 10, 30, 30, 30 and 10, respectively. The range of each of the input features is given in the Table 3.

Table 3. Range of the input features

Parameter	Ω_r	ω_f	V_{DC}(V)	V_{AC}(V)	r_1
Range	[0.01, 0.3]	[0.01, 0.3]	[20, 30]	[0.5, 1.5]	[0.01, 0.02]

In order to avoid numerical error and ill conditioning, each of the input features and the output are scale as follows.

$$\overline{\Omega}_r = \frac{\Omega_r}{\Delta_{\Omega_r}}, \quad \overline{\omega}_f = \frac{\omega_f}{\Delta_{\omega_f}}, \quad \overline{V}_{DC} = \frac{V_{DC}}{\Delta_{DC}}, \quad \overline{V}_{AC} = \frac{V_{AC}}{\Delta_{AC}} \quad \overline{r} = \frac{r_1}{\Delta_r}, \quad \overline{y} = \frac{y}{\Delta_y} \tag{16}$$

in which the bar sign is utilized to show the scaled parameters and $\Delta_i \ \left(i = \Omega_r, \ \omega_f, DC, \ AC, \ r, \ y\right)$ is the scaling number of parameter i. Their values are given below

$$\Delta_{\Omega_r} = 0.3, \quad \Delta_{\omega_f} = 0.3, \quad \Delta_{DC} = 30, \quad \Delta_{AC} = 1.5, \quad \Delta_r = 0.02, \quad \Delta_y = 10^{-5} \tag{17}$$

The error histogram of the neural network is given in Fig. 4. As it is apparent from Fig. 4, all of the dataset containing the train, cross-validation, and test sets have minuscule errors and hence, the neural network has properly tracked the behavior of the model.

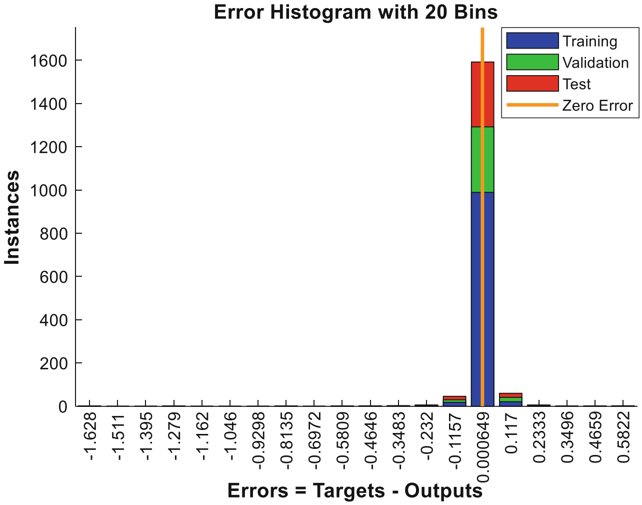


Fig. 4. Error histogram of the neural network

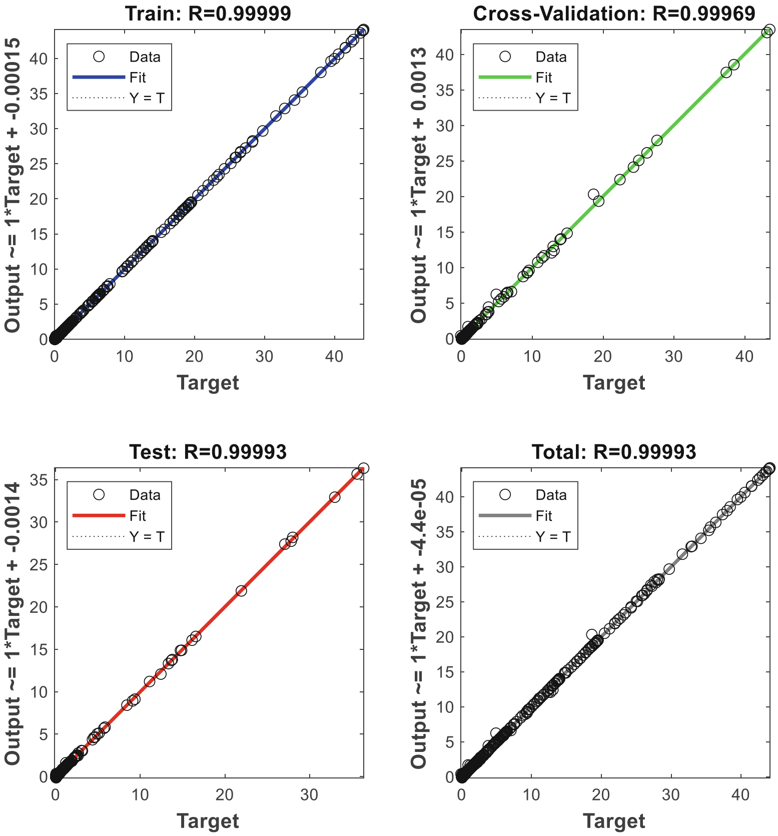


Fig. 5. Regression plot of the train, cross-validation, test, and total data sets

Figure 5 represents the regression plot for the training, cross-validation, test and whole data sets. As can be observed from Fig. 5, the coefficients of correlation *R* for train, cross validation and test sets are extremely close to one which shows an excellent performance for the neural network. Furthermore, it is clear that the neural network exhibits no under fitting nor overfitting. Hence, the model has proven to predict the behavior of micro-ring gyroscopes accurately for both small and large deformations. It is worth mentioning that when the deflections are small (i.e. the Taylor series expansion is valid), the model is capable of predicting the numerical results obtained which are in agreements with various works such as [19, 24].

4 Conclusion

Micro-ring gyroscopes are a type widespread type of micro-gyroscope that measure the angular velocity of a system based on energy transfer between two elliptical modes of a micro-ring. In this study the governing equations of such system have been presented by considering the full electrical force. It has been shown that as the deflection of the micro-ring increases, the accuracy of the Taylor series expansion decreases in comparison with the full electrical force. But taking the exact electrical force will lead to high computational costs. So, a neural network model has been developed to predict the behavior of the micro-system and is show to have a great accuracy and avoids the cases of under and over fitting. This implies that gathering the accurate datasets for different values of the parameters of the micro-ring can be performed in a very short amount of time and can be applied to various designs of micro-ring gyroscopes. To further extend the capabilities of the neural networks, researchers may also consider the pull-in instability in their future neural network designs.

References

1. Ariana, A., Mohammadi, A.K.: Nonlinear dynamics and bifurcation behavior of a sandwiched micro-beam resonator consist of hyper-elastic dielectric film. Sens. Actuator A Phys. **312**, 112113 (2020). https://doi.org/10.1016/j.sna.2020.112113
2. Anjum, N., He, J.H., Ain, Q.T., Tian, D.: Li-He's modified homotopy perturbation method for doubly-clamped electrically actuated microbeams-based microelectromechanical system. FU. Mech. Eng. **19**, 601–612 (2021). https://doi.org/10.22190/FUME210112025A
3. Skrzypacz, P., Ellis, G., He, J.H., He, C.H.: Dynamic pull-in and oscillations of current-carrying filaments in magnetic micro-electro-mechanical system. Commun. Nonlinear Sci. Numer. Simul. **109**, 106350 (2022). https://doi.org/10.1016/j.cnsns.2022.106350
4. Quashie, D., et al.: Magnetic bio-hybrid micro actuators. Nanoscale **14**, 4364–4379 (2022). https://doi.org/10.1039/D2NR00152G
5. Xu, K., Chen, Y., Okhai, T.A., Snyman, L.W.: Micro optical sensors based on avalanching silicon light-emitting devices monolithically integrated on chips. Opt. Mater. Express **9**, 3985–3997 (2019). https://doi.org/10.1364/OME.9.003985
6. Yang, D., Woo, J.K., Lee, S., Mitchell, J., Challoner, A.D., Najafi, K.: A micro oven-control system for inertial sensors. J. Microelectromech. Syst. **26**(3), 507–518 (2017). https://doi.org/10.1109/JMEMS.2017.2692770
7. Ru, X., Gu, N., Shang, H., Zhang, H.: MEMS inertial sensor calibration technology: Current status and future trends. JMM **13**(6), 879 (2022). https://doi.org/10.3390/mi13060879

8. Mohammadzadeh, A., Vafaie, R.H.: A deep learned fuzzy control for inertial sensing: micro electro mechanical systems. Appl. Soft Comput. **109**, 107597 (2021). https://doi.org/10.1016/j.asoc.2021.107597
9. Höflinger, F., Müller, J., Zhang, R., Reindl, L.M., Burgard, W.: A wireless micro inertial measurement unit (IMU). IEEE Tran. Instrum. Meas. **62**(9), 2583–2595 (2013). https://doi.org/10.1109/TIM.2013.2255977
10. N. m. r. 2005-2009. www.nexus-mems.com (2012)
11. Chen, H.Y., Li, W., Yang, H.: Dynamic stability in parametric resonance of vibrating beam micro-gyroscopes. Appl. Math. Model. **103**, 327–343 (2022). https://doi.org/10.1016/j.apm.2021.10.043
12. Shearwood, C., Ho, K.Y., Williams, C.B., Gong, H.: Development of a levitated micromotor for application as a gyroscope. Sens. Actuator A Phys. **83**, 85–92 (2000). https://doi.org/10.1016/S0924-4247(00)00292-2
13. Venediktov, V.Y., Filatov, Y.V., Shalymov, E.V.: Passive ring resonator micro-optical gyroscopes. Quantum Elec. **46**, 437 (2016). https://doi.org/10.1070/QEL15932
14. Soshenko, V.V., et al.: Nuclear spin gyroscope based on the nitrogen vacancy center in diamond. Phys. Rev. Lett. **126**, 197702 (2021). https://doi.org/10.1103/PhysRevLett.126.197702
15. Askari, A.R., Awrejcewicz, J.: Modified couple stress flexural–flexural quasi-static pull-in analysis of large deformable cantilever-based micro-gyroscopes. Commun. Nonlinear Sci. Numer. Simul. **117**, 106933 (2023). https://doi.org/10.1016/j.cnsns.2022.106933
16. Wang, Y., et al.: Quantification of energy dissipation mechanisms in toroidal ring gyroscope. J. Microelectromech. Syst. **30**, 193–202 (2021). https://doi.org/10.1109/JMEMS.2020.3045985
17. Barakat, A.A., Lima, R., Sampaio, R., Hagedorn, P.: Bimodal parametric excitation of a micro-ring gyroscope. PAMM. **20**, 202000153 (2021). https://doi.org/10.1002/pamm.202000153
18. Polunin, P.M., Shaw, S.W.: Self-induced parametric amplification in ring resonating gyroscopes. Int. J. Non Linear Mech. **94**, 300–308 (2017). https://doi.org/10.1016/j.ijnonlinmec.2017.01.011
19. Liang, D.D., Yang, X.D., Zhang, W., Ren, Y., Yang, T.: Linear, nonlinear dynamics, and sensitivity analysis of a vibratory ring gyroscope. Theor. App. Mech. Lett. **8**(6), 393–403 (2018). https://doi.org/10.1016/j.taml.2018.06.001
20. Luo, S., Li, S., Tajaddodianfar, F., Hu, J.: Adaptive synchronization of the fractional-order chaotic arch micro-electro-mechanical system via Chebyshev neural network. IEEE Sens. J. **18**, 3524–3532 (2018). https://doi.org/10.1109/JSEN.2018.2812859
21. Chong, S., et al.: Temperature drift modeling of MEMS gyroscope based on genetic-Elman neural network. MSSP. **72**, 897–905 (2016). https://doi.org/10.1016/j.ymssp.2015.11.004
22. Han, S., Meng, Z., Zhang, X., Yan, Y.: Hybrid deep recurrent neural networks for noise reduction of MEMS-IMU with static and dynamic conditions. Micromachines **12**, 214 (2021)
23. Shao, X., Shi, Y.: Neural-network-based constrained output-feedback control for MEMS gyroscopes considering scarce transmission bandwidth. IEEE Trans. Cybern. **52**, 12351–12363 (2021). https://doi.org/10.1109/TCYB.2021.3070137
24. Liang, F., Liang, D.D., Qian, Y.J.: Nonlinear performance of MEMS vibratory ring gyroscope. Acta Mech. **34**, 65–78 (2021). https://doi.org/10.1007/s10338-020-00195-8

A Framework for Knowledge Representation Integrated with Dynamic Network Analysis

Siraj Munir[1](✉), Stefano Ferretti[1], and Rauf Ahmed Shams Malick[2]

[1] Universit'a degli Studi di Urbino Carlo Bo, Urbino 61029, Italy
s.munir@campus.uniurb.it, stefano.ferretti@uniurb.it

[2] National University of Emerging Sciences, Karachi 74200, Pakistan
rauf.malick@nu.edu.pk

Abstract. Understanding the information residing in any system is of crucial importance. Knowledge Graphs are a tool for achieving such kinds of goals, as they hold the semantic interaction across the entities and, using links, connect them in a better representable way. In this paper, we proposed a dynamic network analysis framework for understanding the evolution of Knowledge Graphs across timelines. To validate our findings, we applied a thorough analysis of the movie recommendation Knowledge Graph, where we considered different snapshots of it. For example, past (historical information), present (current snapshot), and future (predictions based on historical data) information. For the predictions, we employ Graph Neural Network (GNN) modeling. We also compared our recommendation model with the latest related studies and achieved considerable results.

Keywords: Knowledge Graph · Graph Neural Network · Movie Recommendation · Dynamic Network Analysis

1 Introduction

Knowledge Graphs (KG) received massive attention from information science and knowledge based system representation experts in recent years [1–3]. This is because of the inherently more intuitive ability to shape and represent complex forms of knowledge with powerful retrieval and visualization tools [4]. The KG is a higher-order representation of graphs with possible vectors associated with the edges and nodes to represent complex information. The associated nodes and their possibly multi-dimensional relationships can be employed further to infer new kinds of information representation. In several cases, the KG is being used in a dynamic setting where information and knowledge transformation are considered as a continuous process. This demands methods to infer the significance of evaluating new forms of knowledge that should be incorporated into the existing KG. The dynamic representation is powerful and yet can shape a

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 35–49, 2024.
https://doi.org/10.1007/978-3-031-56728-5_4

vulnerable KG because of the lack of knowledge validation and robustness of such KG structures. The presence of a robust knowledge structure will allow the end users to trust the generated knowledge, otherwise, the effectiveness of the KG will be questioned in terms of validation of contextual information and knowledge. With the rise of new KG tools with the ability of meaningful knowledge representation, retrieval, and visualization, it is also important to evaluate the entire process in terms of robustness [35]. The evaluation methods of KG are in the development phase and the state-of-the-art methods are yet to be developed. Because of the avalanche of real-time data, for example, information retrieved from social media [5,6], and [7], sensor-based data [4], or corpus-based information retrieval [5] has the ability to shape knowledge in continuous state space [8]. The continuously updating nature of the basic data and the potential of knowledge transformation from inference methods also demand validation of inferred knowledge. Network Analysis (NA) and Knowledge Graph (KG) are two different ways to understand graphs (networks of interactions with nodes and edges) [9]. Where NA is more focused on the statistical and structural aspects of a given network and KG is more focused on how the network should be interconnected in a better way so that we can extract/ represent information easily [10,11]. In other words, semantically. In this article, we want to combine both approaches (NA and KG) to analyze networks in a dynamic setting. Dynamic networks are self-interesting to explore, as they provide the capability to view the evolution of network(s) transitionally [12]. Through this transitional pattern, we can observe and analyze how any network evolves over the passage of time. Dynamic networks are also very powerful for epistemological studies like COVID-19 and others [13,14]. In this work, we claim that the use of KG, on top of it, could help us to analyze, filter, and query specific types of information residing in that snapshot of the network. By analysis, we observed that a static network or snapshot of KG gives us a partial view of how communities or groups were made or how the interaction took place at first glance. With the help of dynamic network analysis, we can see the evolution pattern of communities and individuals at different levels i.e., edge level, node level, and community level. Further, for predicting future interactions we can resort to machine learning techniques, e.g., Graph Neural Networks. The introduction of GNN over KG and dynamic network analysis allows for observing the past, present, and future of the given KG. For validation of the proposed approach, we took four snapshots of a KG collected from the movie recommendation dataset by IMDb[1] For the implementation of the complete pipeline, we used Neo4j and Cytpscape (a complex network analysis tool for exploiting network attributes)[2]. Neo4j is one of the best available tools for the implementation of graph technology[3]. By integrating the approaches, we reported that the KG has at least two sub-graphs based on activation. Here activation infers information gain (amount of information consisting of a node or edge) and entropy (changes occur in the network

[1] https://www.imdb.com/interfaces/.

[2] https://cytoscape.org/.

[3] https://neo4j.com/.

interactions over time) [12]. One type of sub-graph is where the information is stable i.e., there are no changes in interactions over the period. The other type of sub-graph is entropic in nature i.e., the information is dynamic as we are observing changes in interactions over time. Based on [15,16], and [17], for analyzing and validating the notion of entropy and understanding the evolution of the network, we utilized link analysis along with GNN. The following are the contributions of this research work.

1. Recent literature has explored Knowledge Graph fusion with network analysis [18]. However, this work focuses on the analysis of the Knowledge Graph based on dynamic network analysis.
2. We introduce a dynamic network with Graph Neural Network to predict movie recommendations based on user ratings.
3. We achieve considerable accuracy and f1 measure on the movie recommendation task.

The remainder of this paper is organized as follows. Section 2 discusses the background and state of the art. Section 3 discusses the methodological details and validation of the results. Section 4 discusses the limitations and perspective of this work. Finally, Sect. 4 concludes the paper with future remarks.

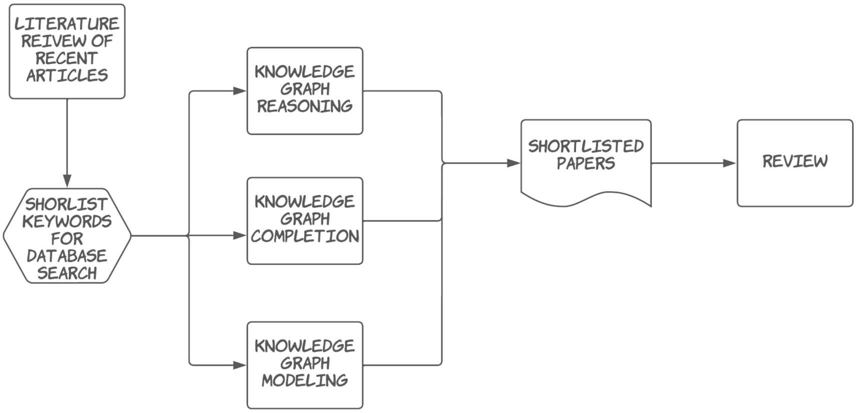

LITERATURE REIVEW OF RECENT ARTICLES
SHORLIST KEYWORDS FOR DATABASE SEARCH
KNOWLEDGE GRAPH REASONING
KNOWLEDGE GRAPH COMPLETION
KNOWLEDGE GRAPH MODELING
SHORTLISTED PAPERS
REVIEW

Fig. 1. Literature Review Pipeline

2 State of the Art

In this section, we will discuss state-of-the-art literature from the domain of KG and dynamic network analysis. KG, as mentioned earlier, is a way to represent real-world concepts in a connected fashion and dynamic networks are a way to analyze network evolution concerning time. So, according to the definition, a KG is a graph G with nodes and edges (V and E) where V represents a relationship and E represents an entity $(person, movie, role, etc.)$. Having the aforementioned definition of KG, a dynamic network is a timestamped version of KG i.e., $\{t1, t2, t3, .tn\}$ where each timestamp itself is a KG representation with different properties like the number of relationships, number of nodes, etc. Figure 1 shows the pipeline for conducting the survey of related work.

2.1 Dynamic Network Analysis

In this section, we will discuss the state-of-the-art literature on dynamic network representation and analysis. The dynamic network is represented as $G(V, E, forg)$, where V is a set of vertices, E is a set of edges u, v, and f represents the cost or weight of the vertices or g represents the cost or weight of the edge. The mentioned representation is used to define either static or dynamic networks. The difference is that in dynamic networks all the parameters V, E, f, g of the graph is dynamic i.e., they can depend on the time variable [19]. In this section, we summarized recent studies highlighting the challenges and introduced different methodologies to cater to them using dynamic network analysis inspired by machine and deep learning [20–25], and [26]. For instance, in [20] authors proposed dynamic metrics for analyzing dynamic networks. Proposed approaches imply machine learning-based deep neural architecture to learn the embedding of temporal growth for the given network. For validation, the authors also tested the proposed methodology for link prediction over real-world datasets. Similarly, in [21] authors introduced an analysis of dynamic network connectivity patterns. The authors investigated different aspects of graph connectivity and analyzed them using statistical parametric and non-parametric features tests. Literature also reported that the proposed approach provides accurate interpretations of dynamic graphs. Literature [22] shed light on neuroimaging and temporal community analysis approaches. The article introduced a detailed survey of available tools for leveraging dynamic graphs as static graphs lack the ability to depict the time-driven features of the network. The authors also introduced a toolbox based on the time-varying structural features for MATLAB (a numerical data analysis tool)[4]. Article [23] introduced a detailed survey on dynamic networks. The article highlighted several aspects of taxonomy and the definition of dynamic networks. In [24] authors studied a dynamic wildlife network that uses GSM-GPS-based module monitoring of mobility. The literature discussed how different animals could be a source of pathogen transmission in epistemological settings. Authors in [25] proposed a detailed survey on data modeling and embedding approaches for dynamic networks. Literature [26] proposed an interesting approach to dealing with GNN. The introduced system utilizes LSTM neural network architecture for learning the temporal features and GNN for learning the structural details of the network. The authors reported the proposed model was able to learn and predict links (new and previously learned) from dynamic representation. For validation, the proposed methodology was tested against

[4] https://www.mathworks.com/products/matlab.html.

state-of-the-art models. The literature above highlights the impact of dynamic networks in different analysis settings. However, more is expected to come, and researchers are exploring it. The next subsection will discuss the state of the art in Knowledge Graph Analysis.

2.2 Knowledge Graph Analysis

Recently, a lot of work has been done on different aspects of dealing with KG. For example, link prediction, neighborhood prediction, community detection, etc. [1–3,18,27–29] and [35]. Authors in [27] presented a detailed study on link prediction tasks. The literature discussed application settings where KG is implemented, and link prediction plays a vital role. The authors considered 18 different methods for comparison of state-of-the-art. In [28] authors discussed the research frontiers for KG as a recommendation system. To compare proposed approaches authors focused on co-occurrence clustering and highlighted the hotspots. The literature in [1] discussed the reasoning mechanism for question-answering systems using KG. The presented approach utilizes GNN for responding to questions. The authors utilized a pre-trained language model and from this model, they described KG that can extract knowledge from the bulk of information. On top of the language model and joint reasoning and relevance scoring, the authors achieved state-of-the-art performance. The authors in [29] presented an information fusion-based method for clustering in KG. Authors in [2] presented a data science-centric toolkit for the analysis of KG. Proposed toolkit integrated different tools to provide end-users a way to transform, extract and enhance KG modeling. [3] presented a detailed view of KG representation over three decades i.e., from 1991–2020. The authors also discovered different contributing facts like top contributing countries and top cited literature, top researchers, etc. Authors in [18] introduced the information fusion-based framework for KG modeling. The authors also conducted a thorough survey of the recent state-of-the-art and discussed crucial future directions. [35] introduced a unique methodology to answer the challenges of KG representation. The methodology presented a two-way semi-autonomous approach i.e., (automated rule extraction and human-in-the-loop rule modeling), to handle raw data and model it into a KG. The next section will discuss the methodological details of the proposed framework.

3 Methodology and Results

In this section, we will discuss the methodological details of the proposed framework for exploring KG as a dynamic network. For the validation of modeled KG, we utilized GNN to predict recommendations based on user rating. For a comprehensive study, we divided our framework into 5 different phases i.e., (i) Dataset collection from raw dump files, (ii) KG modeling, by extracting semantic triplets, (iii) KG embedding for representation of the dynamic network and observing the entropic nature, (iv) Dynamic network analysis to highlight the sub-graphs holding entropic nature within KG, and (v) Use GNN for learning these dynamic patterns of KG and use them for the prediction of movie recommendation. For the modeling of KG, we considered four snapshots of the IMDb movie recommendation dataset. The IMDb dataset contains the interaction between movies and actors along with users rating against the respective movie(s) as shown in Fig. 2. Note that we have also added a temporal node label named location which is a synthetic event. The intuition to introduce the additional node label is to observe KG as a dynamic network. Figure 3 depicts the layered workflow to model the KG as a dynamic network.

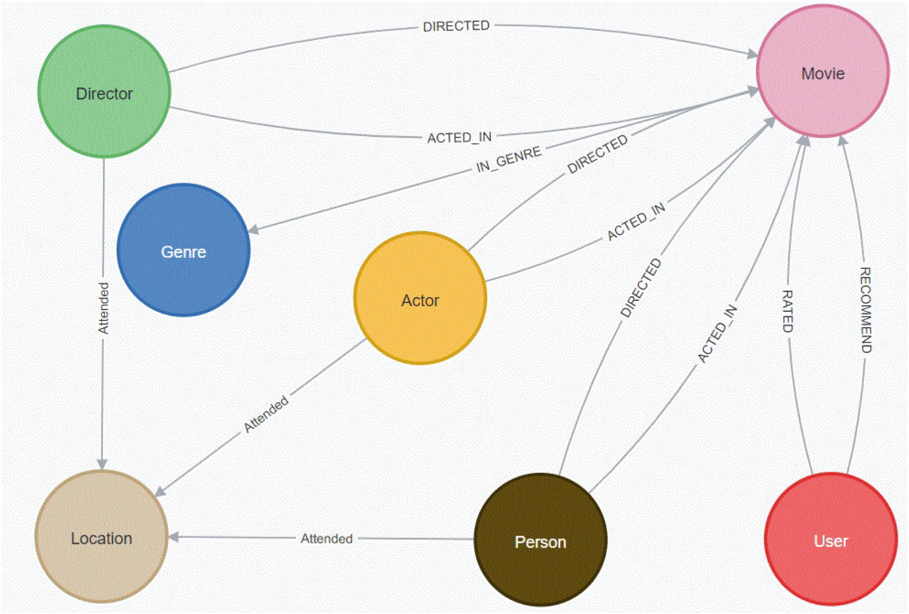


Fig. 2. Graph Data Model of Movie Recommendation Dataset

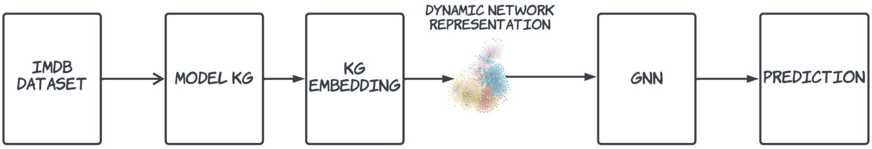


Fig. 3. Framework for Knowledge Representation of Dynamic Network

3.1 Dataset

The version utilized for analysis of KG contains 344604 nodes and 1188564 interactions in the final version. As we worked with different versions of the same dataset, we noticed each version contains approx. a similar number of nodes with some variation of the interaction (relationship) count. A summary of KG is given in Table 1.

Table 1. Knowledge Graph Description

Number of Nodes	344,604,
Number of Relationships	1,188,564,
Types of Relationships	6 (ACTED_IN, DIRECTED, Attended, IN_GENRE, RATED, Recommended)

3.2 Knowledge Modeling

For the modeling of information, we utilized the KG discussed in the previous subsection. First, we collected the open-sourced dump files for the movie recommendation dataset. Later using appropriate syntax, we imported these files into the Neo4j graph database. Finally, we extracted semantic triples that are represented by {*Subject (source), Predicate (relationship), Object (destination)*}. However, we also introduced the location as an event label. These are basically pseudo-random synthetic temporal events that introduce the temporal dynamicity synthetically into the KG. We named this event "A Red Carpet" and by adding new relationship labels we represented the association of nodes i.e., actors and directors. Finally, after applying the mentioned steps we introduced the KG to the FastRP embedding algorithm for generating vector representation so that we can pass it to GNN for generating recommendations.

3.3 Knowledge Graph Embedding

In graph recommendation scenarios, we often need to deal with bipartite graphs for learning associativity patterns. The FastRP algorithm uses the classical bipartite process i.e., (having two disjoint sets) for top-K recommendations. FastRP algorithm is also proven to be a scalable node embedding algorithm that utilizes explicit node similarity and sparse representation to project lower dimensional representation of vector space [30]. Figure 4 depicts the FastRP process for recommendations. In this work, we utilized the FastRP algorithm for generating embedding for user and movie recommendations based on rating. For training the embedding model we used a 56-dimensional vector representation. Further, we used movies and user nodes to learn similarities and utilized them for the recommendation of novel relationships or links in the future. Finally, we passed this network to the dynamic network analysis phase to learn the evolution patterns.

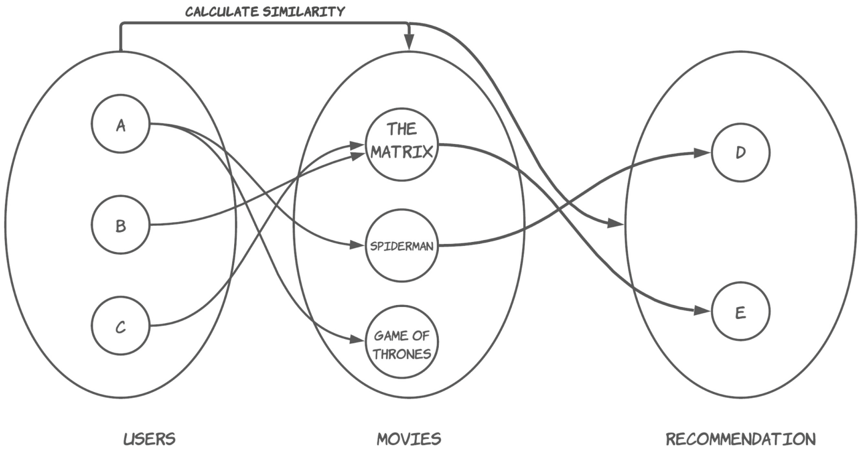


Fig. 4. Bi-partite Recommendation Process used by FastRP

3.4 Dynamic Network Analysis

In this work, we considered the notion of network dynamics (entropy or information flow) as defined in [16]. In dynamic networks, it is not necessary that the structure of the network is dynamic. Rather, the dynamicity of the network could be due to changes in the state of network nodes that can be mathematically represented by Eq. 1:

$$x_i = f_1\left(x_i\right) + \sum_{j=1}^{n} A_{ij} f_2\left(x_1, x_2\right),\ \forall\ i\ \epsilon\ 1, 2, 3, ..., n, \tag{1}$$

where x_i represents the current state of the node (i), which is expected to evolve over time. However, the state of the node(i) at each time step could be observed as x_i (t), where t represents a time step. $A_i j$ is the adjacency matrix representing network nodes and edges. Here f_1 represents the inherent dynamics of the node at state x_i and f_2 represents a pair of neighboring nodes. For visualization and understanding of evolution patterns of network dynamics, we introduced dynamic network analysis. This helped us to observe different patterns across the network timeline i.e., the creation, merging, and breaking of communities. Figure 5 shows the comparison of these timestamps. This analysis showed us a new viewpoint of the network i.e., within our network, we have two types of information or entropic flows. That is some portion of the network holds concrete information and tends it to keep the same structure of communities and node associations i.e., this information is stabilized or concrete. Hence nothing changes over the span. On the other hand, some parts of the network i.e., relationships and entities are dynamic and are changing at every timestamp of the

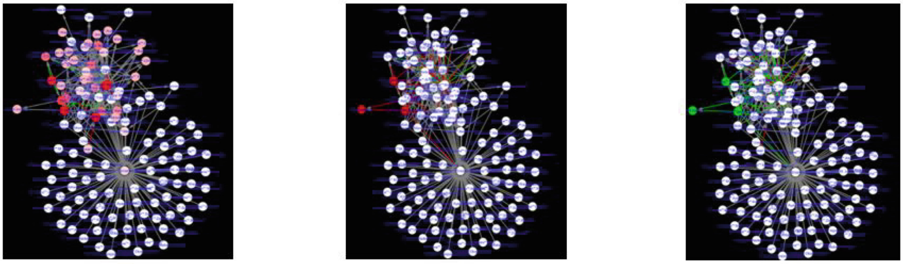


Fig. 5. Figure 5: Comparative analysis of different timestamped versions. Figure 5(a) represents the changes held between the first and last snapshots. While Fig. 5(b) represents the changes that occurred in contrast to the third snapshot. Figure 5(c) represents the changes that occurred between the second and third snapshots.

network. This means the information is not stabilized and hence we are observing new community patterns and node interactions as shown in Fig. 4. Usually, with single timestamped network analysis this point of view is ignored. However, with dynamic network analysis, we can easily visualize the changes happening within the network. To dig deeper into entropy and its evolution, we introduced the GNN architecture proposed in [15]. With the help of GNN, we were able to learn the dynamic part of the network and predict future interactions i.e., movie recommendations. The following subsection highlights the details of the prediction.

3.5 Prediction

In recent literature works, researchers have used GNN, and other machine learning based approaches to understand and learn recommendation tasks [17] and [31]. However, in this study, our goal is to understand the GNN learning and recommendation from the perspective of KG represented as a dynamic network. For attaining this goal, we conducted conserved learning i.e., we used only a sub-part of KG that we obtained by dynamic network analysis. Therefore, for the validation of dynamic patterns within our KG network, we introduced GNN architecture [15]. Using GNN we modeled a top-K recommendation system for movies. For learning patterns of interest, GNN uses neighborhood-inspired collaborative filtering and historic user behavior i.e., rating of movies and interests [17]. For the evaluation of the proposed approach, we used error, accuracy, and F1-measure metric. The details of the evaluation are discussed in the later section. As mentioned earlier, the GNN implementation utilized in this work leverages aggregated neighborhood learning i.e., multi-hop neighborhood. Then, based on learned patterns, GNN tries to predict future node labels and relationships. The model learning can mathematically be represented by Eq. 2:

$$Learning = AGGREGATE_k \left(\left(h_u^{k-1} \right), \ \forall u \in N\left(v\right) \right), \tag{2}$$

where h_u^{k-1} is the node representation of node u over the N is the number of neighborhood vectors with k depth [15]. Moreover, we adopted Nadam (Nesterov Adaptive Momentum) as an activation function with RMSE (Root Mean Squared Error) analysis for model validation [8]. That is given by,

$$RMSE = sqrt\frac{\sum_{i-1}^{N}(x_i - x_i)}{N}, \tag{3}$$

where x_i and x_i represent actual and predicted output over N number of observations of the model. A comparative study conducted in [32] has validated that the Nadam converges optimal gradient quicker as compared to its variants. The mathematical formulation for Nadam is as follows,

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\nu_t} + \varepsilon}\left(\beta_1 m_t + \frac{(1-\beta_t)\,g_t}{1-\beta_1^t}\right), \tag{4}$$

where θ_{t+1} represents the update over the given timestep or iteration, θ_t represents the input parameter at a given timestep or iteration, ν_t represents momentum decay (direction of the gradient (a measure of all weights and error to optimize the learning of model)), m_t represents the parameter update of the current timestep, g_t represents parameter update of next timestep, η and β are hyperparameters for model learning [32]. Figure 6 shows the RMSE based error analysis of the GNN model trained on the presented KG. On the x-axis, we have the number of iterations, and on the y-axis, we have the error rate whose average is [1.08–1.1]. The proposed model achieved good results i.e., 58.23 % accuracy and 53% F1-score in recommendation tasks. Furthermore, recent and related studies like [33] and [24] did not consider dynamic network analysis metrics and used a single snapshot of the dataset for evaluation. Table 2 shows the comparison of the presented model with recent studies on movie recommendations.

Table 2. Comparison of Model Performance

Paper	Accuracy	F1-Score
[33]	48	Not Evaluated
[34]	60	Not Evaluated
Ours	58.23	53

Finally, based on the learned GNN model we predicted top 10 movies against each user based on his/her rating and incorporated this information into KG for knowledge enrichment or link prediction. This knowledge enrichment phase is crucial as it enhances the informational view of knowledge. This knowledge enrichment mechanism can be represented by Eq. 5.

$$h_v^k = f_{update}\left(a_v, h_v^{k-1}\right), \tag{5}$$

where k is the number of neighbors and a_v is the representation of aggregated information. This complete procedure for link prediction is presented in Algorithm 1.

Algorithm 1. Link Prediction Mechanism

```
  Begin(data, userId, movieId):
  Procedure importPredictions(data):

 while $data \neq null$ do

     for each row in data: do
         Retrieve the user with userId equal to row.user and store it as u

         for each movieId in row.movies: do
             Retrieve the movie with movieId equals to row.movies and store it as m
             if there is no existing rated relationship between u and m: then
                 Create a new recommendations relationship between u and m
             end if
         end for
     end for
 end while
 data = results
 Call importPredictions(data)
```

4 Discussion and Limitations

In this work, we introduced the integrated approach for the representation of KG as a dynamic network in an entropic setting. As a validation of the study, we utilized the IMDb dataset for movie recommendations. A survey of related research works was studied to determine the problem foundations. For conducting this study, we considered four snapshots of the aforementioned dataset i.e., the last four releases of the respective dataset[5]. Using Cytoscape and Neo4j as graph analysis tools we modeled and analyzed KG as a dynamic network. Lastly, we introduced GNN to learn the identified dynamic portion of KG for recommendations of movies based on user rating. The combinatorial study conducted in this work is novel, as best to our knowledge. However, there are similar studies conducted by other researchers but either they are determined towards GNN learning [1,26], and [27] or KG focused like [28,29] or just dynamic networks [20–26]. In [18] authors introduced a framework that uses KG fusion and complex networks, but they did not use or considered dynamic network analysis and its metrics. This study opens a new venue to explore and leverage the power of KG and GNN along with the metrics of dynamic networks. However, still, there are several aspects that need attention. For instance, there is a requirement for the identification of a baseline model for the evaluation of such combinatorial studies. Another limitation is the testing and validation of this kind of study with more real-world datasets for robustness. Moreover, there is a requirement for analysis, evaluation, and formalization of the entropy notion referred to in this

[5] https://github.com/neo4j-graph-examples/recommendations/tree/main/dataDatasetisdownloadablefromthementionedlink..

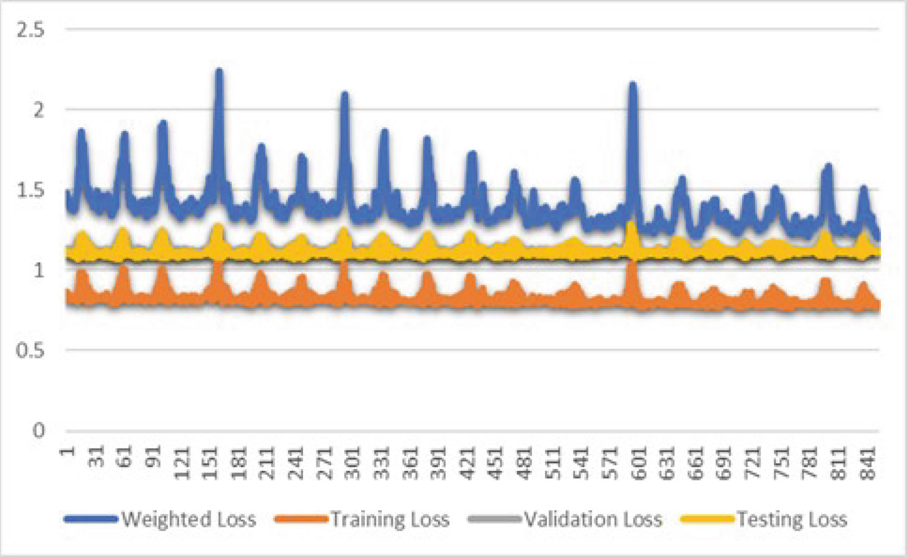


Fig. 6. RMSE based GNN Train Test Analysis

work [16]. Other possible venues worth exploring are the fusion of dynamic networks for federated querying or learning or understanding of dynamic networks in epistemological settings like COVID-19 analysis or protein-protein analysis, etc. This study highlighted the possibility of integrating graph technologies with entropy concepts and leveraging them in useful use cases like recommendations and others. These results can also be applied to the scenarios of surveillance-based networks to analyze the dynamic creation and evolution of nodes in different communities. Also, these results are fruitful for the studies where GNN is utilized for learning the specific pattern of interest. For example, link prediction, neighborhood detection, interaction networks, etc.

5 Conclusion and Future Work

In this paper, we presented a framework for the analysis of the Knowledge Graph as a dynamic network. We commented in detail on how different timestamps of the same Knowledge Graph can hold and represent different information. The static version lacks this vision of information representation. We observed that dynamic networks are very crucial and possess a high impact on understanding network behavior. We also discussed entropy could be a very useful metric to analyze network dynamics. Finally, we integrated the dynamic networks with Knowledge Graph based movie recommendation dataset and trained a Graph Neural Network model to predict future interactions. Note that this study was intended to represent the integration of a Graph Neural Network with a Knowledge Graph modeled as a dynamic graph. We will try to use other prediction methods to improve the learning of Graph Neural Networks. In future work, we will also explore in depth hyperparameter tunning for Graph Neural Networks in entropic networks. We will also explore different Graph Neural Network architectures and compare their performances on dynamic networks. Further, we will explore different use case settings where the presented framework fits and improve the information representation flow and analysis for end users. For example, decentralized searching and querying (search blockchain network or smart contracts information over the network represented as Knowledge Graph), the federation of Knowledge Graph (fusion of Knowledge Graphs) as a dynamic network, and federated querying (querying the fused Knowledge Graph engine in an efficient way). These venues will further open new directions for the research community to expand and leverage the power of Graph Neural Networks and Knowledge Graphs.

Acknowledgement. This research is funded by the PON Program and the University of Urbino Carlo Bo under matriculation number: 317840.

References

1. Yasunaga, M., Ren, H., Bosselut, A., Liang, P. and Leskovec, J. QA-GNN: reasoning with language models and knowledge graphs for question answering. arXiv preprint arXiv:2104.06378 (2021)
2. Ilievski, F., Pan, J.Z., et al.: KGTK: a toolkit for large knowledge graph manipulation and analysis. In: The Semantic Web-ISWC (2020)
3. Chen, X., Xie, H., Li, Z., Cheng, G.: Topic analysis and development in knowledge graph research: a bibliometric review on three decades. Neurocomputing **461**, 497–515 (2021)
4. Nguyen, H.L., Vu, D.T., Jung, J.J.: Knowledge graph fusion for smart systems: a survey. Inf. Fusion **61**, 56–70 (2020)
5. Gao, L., Wang, Y., Li, D., Shao, J., Song, J.: Real-time social media retrieval with spatial, temporal and social constraints. Neurocomputing **253**, 77–88 (2017)
6. Yao, W., et al.: Early and late fusion of multiple modalities in sentinel imagery and social media retrieval. In: Del Bimbo, A., et al. (eds.) ICPR 2021. LNCS, vol. 12667, pp. 591–606. Springer, Heidelberg (2021). https://doi.org/10.1007/978-3-030-68787-8_43
7. Wise, C., et al.: COVID-19 knowledge graph: accelerating information retrieval and discovery for scientific literature. In: AACL-IJCNLP (2020)
8. Bosselut, A., Le Bras, R., Choi, Y.: Dynamic neuro-symbolic knowledge graph construction for zero-shot commonsense question answering (2021)
9. de Haan, P., Cohen, T.S., Welling, M.: Natural graph networks. Adv. Neural. Inf. Process. Syst. **33**, 3636–3646 (2020)
10. Singhal, A.: Introducing the knowledge graph: things, not strings. Official google blog. https://blog.google/products/search/introducing-knowledge-graph-things-not/. Accessed 13 June 2023
11. Chiesi, A.M.: Network Analysis. Pergamon (2001)
12. Carley, K.M.: Dynamic network analysis (2003)

13. Chu, A.M., Chan, T.W., So, M.K., Wong, W.K.: Dynamic network analysis of COVID-19 with a latent pandemic space model. Int. J. Environ. Res. Public Health **18**(6), 3195 (2021)
14. Chakrabarti, P., Jawed, M.S., Sarkhel, M.: COVID-19 pandemic and global financial market interlinkages: a dynamic temporal network analysis. Appl. Econ. **53**, 2930–2945 (2021)
15. Hamilton, W., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. Adv. Neural. Inf. Process. Syst. **30**, 1–11 (2017)
16. Lambiotte, R., Schaub, M.T.: Modularity and Dynamics on Complex Networks. Cambridge University Press, Cambridge (2021)
17. Zhang, S., Yao, L., Sun, A., Tay, Y.: Deep learning based recommender system: a survey and new perspectives. ACM Comput. Surv. (CSUR) **52**, 1–38 (2019)
18. Lü, J., Wen, G., Lu, R., Wang, Y., Zhang, S.: Networked knowledge and complex networks: an engineering view. IEEE/CAA J. Automatica Sinica **9**, 1366–1383 (2022)
19. Harary, F., Gupta, G.: Dynamic graph models. Math. Comput. Model. **25**, 79–87 (1997)
20. Goyal, P., Chhetri, S.R., Canedo, A.: dyngraph2vec: capturing network dynamics using dynamic graph representation learning. Knowl.-Based Syst. **187**, 104816 (2020)
21. Khambhati, A.N., Sizemore, A.E., Betzel, R.F., Bassett, D.S.: Modeling and interpreting mesoscale network dynamics. Neuroimage **180**, 337–349 (2018)
22. Sizemore, A.E., Bassett, D.S.: Dynamic graph metrics: tutorial, toolbox, and tale. Neuroimage **180**, 417–427 (2018)
23. Skarding, J., Gabrys, B., Musial, K.: Foundations and modeling of dynamic networks using dynamic graph neural networks: a survey. IEEE Access **9**, 79143–79168 (2021)
24. Triguero-Ocaña, R., Martínez-López, B., Vicente, J., Barasona, J.A., Martínez-Guijosa, J., Acevedo, P.: Dynamic network of interactions in the wildlife-livestock interface in mediterranean Spain: an epidemiological point of view. Pathogens **9**(2), 120 (2020)
25. Xie, Y., Li, C., Yu, B., Zhang, C., Tang, Z.: A survey on dynamic network embedding. arXiv preprint arXiv:2006.08093 (2020)
26. Chen, J., Wang, X., Xu, X.: GC-LSTM: graph convolution embedded LSTM for dynamic network link prediction. Appl. Intell. **52**, 7513–7528 (2022)
27. Rossi, A., Barbosa, D., Firmani, D., Matinata, A., Merialdo, P.: Knowledge graph embedding for link prediction: a comparative analysis. ACM Trans. Knowl. Disc. Data (TKDD) **15**, 1–49 (2021)
28. Shao, B., Li, X., Bian, G.: A survey of research hotspots and frontier trends of recommendation systems from the perspective of knowledge graph. Expert Syst. Appl. **165**, 113764 (2021)
29. Tang, J., Liu, Y., Lin, K.-Y., Li, L.: Process bottlenecks identification and its root cause analysis using fusion-based clustering and knowledge graph. Adv. Eng. Inf. **55**, 101862 (2023)
30. Chen, H., Sultan, S. F., Tian, Y., Chen, M., Skiena, S.: Fast and accurate network embeddings via very sparse random projection (2019)
31. Fan, W., et al.: Graph neural networks for social recommendation (2019)
32. Dozat, T.: Incorporating nesterov momentum into adam (2016)
33. Saraswat, M., Chakraverty, S., Kala, A.: Analyzing emotion based movie recommender system using fuzzy emotion features. Int. J. Inf. Technol. **12**, 467–472 (2020)

34. Lim, B., Bansal, S., Buru, A., Manthey, K.: A multimedia recommendation model based on collaborative graph. arXiv preprint arXiv:2205.14931 (2022)
35. Munir, S., Ferretti, S.: Towards symbolic representation-based modeling of Temporal Knowledge Graphs. In: International Conference on Smart Applications, Communications and Networking (SmartNets), pp. 1–8 (2023). https://doi.org/10.1109/SmartNets58706.2023.10215541

Time Series Forecasting Using Parallel Randomized Fuzzy Cognitive Maps and Reservoir Computing

Omid Orang[1,3](✉), Hugo Vinicius Bitencourt[1], Petrônio Cândido de Lima e Silva[2], and Frederico Gadelha Guimarães[1,3]

[1] Graduate Program in Electrical Engineering, Machine Intelligence and Data Science (MINDS) Laboratory, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
omid.orang2009@gmail.com, {hugovynicius,fredericoguimaraes}@ufmg.br

[2] Federal Institute of Education Science and Technology of Northern Minas Gerais, Januária Campus, Belo Horizonte, Brazil
petronio.candido@ifnmg.edu.br

[3] Department of Computer Science, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
https://minds.eng.ufmg.br

Abstract. Fuzzy Cognitive Maps (FCMs) have been widely employed as nonlinear forecasting methods that are easily interpretable. They have a remarkable capability to enhance accuracy and are well-equipped to handle uncertainty and emulate the dynamics of complex systems. The main goal of this article is to present a new randomized multiple-input multiple-output (MIMO) FCM-based forecasting technique named M-PRFCM to forecast real-world high-dimensional time series in Internet of Things (IoT) applications. M-PRFCM is a first-order forecasting method integrating the concepts of randomized FCMs, Echo State Networks (ESNs), and Kernel Principle Components Analysis (KPCA). The training process of M-PRFCM is accelerated as a result of utilizing the ESN weight initialization trick, which randomly selects weights. The obtained results show the efficacy and validation of the proposed technique in terms of accuracy when compared with other existing approaches.

Keywords: Fuzzy Cognitive Map · Echo State Network · Kernel Principal Component Analysis · Multiple-Input Multiple-Output

1 Introduction

Fuzzy Cognitive Maps (FCMs) as popular weighted fuzzy time series (FTS) approaches were proposed by Kosko et al. [1] and used widely in the area of time series forecasting. Structurally, FCMs are interpretable recurrent neuro-fuzzy networks composed of nodes and causal relations among pairs of nodes. FCMs can effectively model the dynamic behavior of complex systems and deal with uncertainty [2,3]. Thus, a range of FCM-based forecasting methods has been

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 50–61, 2024.
https://doi.org/10.1007/978-3-031-56728-5_5

introduced to predict univariate and multivariate time series [4]. Constructing a proper structure of FCMs and training the weight matrices are key components of using FCMs for time series forecasting [5]. Granularity, membership values representation, Fuzzy c-means clustering, wavelet transformation, and empirical mode decomposition (EMD) are commonly used methods for constructing the structure of FCMs [5].

Based on the literature [6], most research on weight learning has focused on the use of population-based methods. Although Genetic Algorithm (GA) [7] and Particle Swarm Optimization (PSO) [8] have been widely used, they have recently been replaced by regression-based methods that are more time-efficient [9–11].

Although FCMs have demonstrated impressive achievements in the domain of time series forecasting and analysis, there are still gaps and challenges that need to be tackled. Multiple output prediction is one of these challenges. As such, the main aim of this paper is to implement a randomized Multiple-Input Multiple-Output (MIMO) FCM-based forecasting approach to predict high dimensional time series named M-PRFCM (MIMO-Parallel Randomized FCM). To clarify, M-PRFCM is a hybrid method that integrates the concepts of FCM, Echo State Network (ESN) [13], reservoir computing, and Kernel Principal Component Analysis (KPCA).

Thus, M-PRFCM draws inspiration from ESN reservoir computing and offers a substantially lower computational cost than population-based learning algorithms. Two high-dimensional datasets in IoT (Internet of Things) applications are applied to assess the performance of the model when compared to other implemented baseline MIMO techniques. The results obtained confirm the superiority of M-PRFCM in terms of efficacy and accuracy.

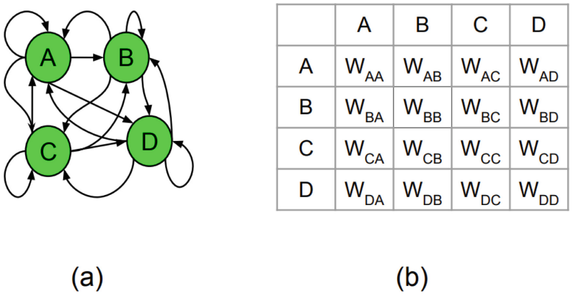
The following is an outline of the structure for the rest of this paper: Sect. 2 presents a brief description of the Fuzzy Cognitive Map; Sect. 3 introduces the proposed method in details; Sect. 4 describes two IoT case studies used to test our methodology; the experimental results and corresponding discussion are presented in Sect. 5; and 6 will contain the final remarks, conclusions, and suggestions for future work.

2 Fuzzy Cognitive Maps

FCMs are interpretable recurrent neural networks with a high ability to deal with uncertainties as well as an excellent ability to model the dynamics of complex systems [17]. FCMs are formulated by a set of concepts (nodes) and the directed signed arrows (weights) connecting pairs of concepts, which represent the influence each node has on the others.

Each FCM is distinguished by four elements, $(\mathbf{C}, \mathbf{W}, \mathbf{a}, f)$, so that $\mathbf{C} = [c_1, \ldots, c_n]$ signifies the collection of n concepts. The activation levels of these concepts at any given time t are expressed as follows:

$$\mathbf{a}(t) = (a_1, \ldots, a_n) \tag{1}$$



	A	B	C	D
A	W_{AA}	W_{AB}	W_{AC}	W_{AD}
B	W_{BA}	W_{BB}	W_{BC}	W_{BD}
C	W_{CA}	W_{CB}	W_{CC}	W_{CD}
D	W_{DA}	W_{DB}	W_{DC}	W_{DD}

(a) (b)

Fig. 1. Simple FCM with 4 nodes (a) graphical structure (b) Weight matrix

where $a_i \in [0, 1]$, $i = 1, 2, ..., n$. The core element is a weight matrix with size $n \times n$ that represents the connections between the nodes, defined as follows:

$$\mathbf{W} = \begin{pmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \dots & w_{nn} \end{pmatrix} \tag{2}$$

Here, $w_{ij} \in [-1, 1]$ $(i, j = 1, 2, \dots, n)$ denotes the influence of node i on node j and can be categorized into three groups (positive, negative, and zero). A simple structure of an FCM with four concepts is exemplified in Fig. 1.

The activation state of each node at the $(t + 1)$-th iteration is calculated using Kosko's updating rule. Thus, the dynamics of an FCM can be expressed using the equation provided below.

$$a_i(t + 1) = f\left(\sum_{j=1}^{n} w_{ji} a_j(t)\right) \tag{3}$$

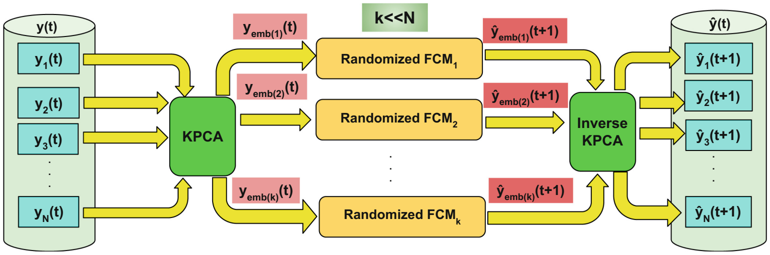


Fig. 2. General structure of the proposed method.

where $a_i(t)$ represents the state value of concept c_i at time step t and $a_i(t+1)$ indicates the state value of concept c_i at time step $t+1$. The activation function (f) is the final element used to map the activation degree of a node into the state space. According to the Eq. 3, the value of each concept at time $t+1$ depends on the activation level of all connected concepts at time t.

3 Proposed Methodology

This section presents our proposed M-PRFCM method which is an expansion of the earlier proposed univariate R-HFCM technique [12]. R-HFCM is a group of randomized HFCM-FTS proposed in [14] while the weights are randomly chosen using the ESN weight initialization technique. Figure 2 demonstrates the fundamental configuration of the M-PRFCM technique.

As depicted in Fig. 2, the M-PRFCM method is a combination of KPCA and randomized-based FCM. To elaborate, the original time series is passed through KPCA to produce k user-defined principal components, defined by $(\mathbf{y}_{emb} = y_{emb(1)}, y_{emb(2)}, ..., y_{emb(k)})$, in a new feature space such that $k \ll N$. After that, the obtained components are fed as inputs to each randomized FCM block independently. Subsequently, the predicted values $(\hat{y}_{emb(1)},...,\hat{y}_{emb(k)})$ from k models at time $t+1$ are fed to the inverse KPCA unit to calculate the final predicted value for each original feature.

Based on Fig. 3, the internal structure of M-PRFCM. Each principal component $y_{emb(i)}(t)$, $i = \{1, 2, ..., k\})$ is injected into each block individually. Thus, for each component, there will be a corresponding randomized FCM block. Each block consists of three layers: the input layer, the intermediate or reservoir layer, and the output layer. Thus, M-PRFCM is a kind of ESN where solely the output layer of each randomized FCM unit is trainable and the reservoir parameters are randomly initialized and remain unchanged throughout the training procedure. It is noteworthy that each randomized FCM block represents the first-order version of the R-HFCM model.

The M-PRFCM method can be divided into two main phases: the Training process and the Forecasting process. The details of these steps are explained in the following sub-sections.

3.1 Training Procedure

The main objective of this procedure is to find the least squares coefficients training the output layer in each randomized FCM, given a crisp training embedded data set. The steps of the method are listed below:

(1) **Pre-processing**: Prior to training, the input data sets are refined through the removal of outliers and missing values, since M-PRFCM lacks the capability to forecast missing values. Addressing this limitation could be considered for future research.

(2) **Embedding**: In this stage, RBF kernel PCA is applied to extract k principal components, which better present the components of high-dimensional time series. Firstly, the Eq. 4 is applied to compute the kernel similarity matrix.

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \parallel \mathbf{x}_i - \mathbf{x}_j \parallel_2^2) \tag{4}$$

where γ is the kernel coefficient. Then, $\tilde{k} = k - 1_T k - k 1_T + 1_T k 1_T$ is used to determine the centered kernel matrix. 1_T is an $T \times T$ matrix with all values equal to $1/T$. Third, eigenvalues are computed based on $\tilde{k}\alpha_i = \lambda_i T \alpha_i$. Finally, the eigenvectors associated with the highest eigenvalues are extracted.

(3) **Weight initialization**: This step aims to initialize the weight matrix between the concepts C_i and C_j at each time step for each sub-reservoir in Fig. 3. The novelty is that weight matrices are selected randomly using the ESN weight initialization trick according to the following formula:

$$\mathbf{W}^t = \mathbf{W}^{\mathbf{rand}} \cdot \left(\frac{\epsilon}{\rho_{\mathbf{max}}(\mathbf{W}^{\mathbf{rand}})} \right) \tag{5}$$

where $\mathbf{W}^{\mathbf{rand}}$ is chosen randomly with a uniform distribution in the range [–1, 1] and $\rho_{\mathbf{max}}(\mathbf{W}^{\mathbf{rand}})$ is the maximum eigenvalue of $\mathbf{W}^{\mathbf{rand}}$. $\epsilon \in (0, 1)$ is the scaling parameter and is set to 0.5 to preserve the members of the weight matrix in the FCM in the range [–1, 1]. Also, Eq. 6 is utilized to initialize the bias weight vector in our method.

$$\mathbf{w}^0 = \mathbf{w}^{\mathbf{0}}_{\mathbf{rand}} \cdot \left(\frac{\epsilon}{\mathbf{S}} \right) \tag{6}$$

Here, ($\mathbf{S}$) stands for the maximum singular value of $\mathbf{w}^{\mathbf{0}}_{\mathbf{rand}}$, $\mathbf{w}^{\mathbf{0}}_{\mathbf{rand}}$ is chosen randomly from a uniform distribution in the range [–1, 1] and $\epsilon = 0.5$.

(4) **Partitioning**: Randomized FCM units share common linguistic variables, represented by fuzzy sets defined over $U = [\min(Y), \max(Y)]$. In this stage, U is partitioned into κ overlapping intervals with the same length. Next, the midpoint of each interval (mp_i, $i \in \{1, \ldots, \kappa\}$) and the membership function (μ_{C_i}) are used to define a fuzzy set C_i (i.e. a concept). In each sub-reservoir, the linguistic variable C is created by the set of the κ concepts where $C_i \in C$, $\forall i \in \{1, \ldots, \kappa\}$. M-PRFCM uses grid partitioning where the number of concepts matches the number of partitions.

(5) **Fuzzification**: The aim of this step is to transform numerical time series (Y_{emb}) into a fuzzy series(A) indicating the state value $a_i(t)$ of each concept ($C_i \in C$) for each sample. Then, the set of activations $a_i(t) \in a(t)$ represents the associated fuzzified value $a(t) \in A$ for each sample $y_{emb}(t) \in Y_{emb}$ such that $a_i(t) = \mu_{C_i}(y_{emb}(t))$, for $i \in \{1, \ldots, \kappa\}$.

(6) **Activation**: The activation level of each concept at time $t + 1$ inside each sub-reservoir can be stated via the following equation:

$$\mathbf{a}_j(t+1) = f\left(\mathbf{w}^0 + \mathbf{W}^l \cdot \mathbf{a}(t)\right) \tag{7}$$

(7) **Defuzzification**:
After updating the state value of each concept, defuzzification is applied to generate outputs corresponding to each sub-reservoir. The numerical prediction value for each sub-reservoir at time $t+1$ can be computed using Eq. 8.

$$\hat{y}_j(t+1) = \frac{\sum_{i=1}^{k} a_{ji}(t+1) \cdot mp_i}{\sum_{i=1}^{k} a_{ji}(t+1)} \tag{8}$$

where $a_j(t+1)$ is obtained from the activation step and mp_i is the center of each concept $C_i \in C$.

(8) **Least Squares coefficients determination**:
For the outputs from each sub-reservoir ($\hat{y}_{sub(1)}(t+1)$, $\hat{y}_{sub(2)}(t+1)$,..., $\hat{y}_{sub(L)}(t+1)$ where $t = 1, ..., T$), a matrix $X \in \mathbb{R}^{L \times T}$ is established to illustrate the linear system $Y = \lambda X$, where $\lambda = [\lambda_0, ..., \lambda_j]$ is the coefficient vector. least squares approach is applied to solve this linear system and determine the λ coefficient vector that minimizes the Mean Squared Error.

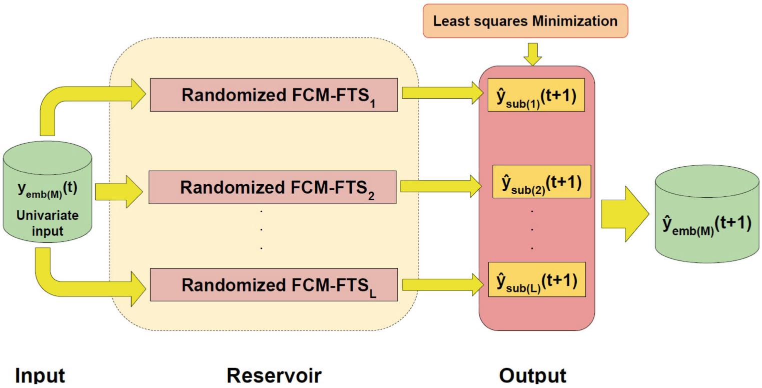


Fig. 3. Internal structure of each randomized FCM unit.

3.2 Forecasting Procedure

In this stage, the goal is to forecast the final values of all variables ($\hat{y}_1(t+1)$, $\hat{y}_2(t+1)$,..., $\hat{y}_N(t+1)$), given the linguistic variable C, weight matrices $\mathbf{W}^t$, activation function f, and an input $Y(t)$. The following elucidates the complete procedure of forecasting steps.

(1) **Pre-processing**: The same as Step 1 in Training procedure.
(2) **Embedding**: The same as Step 2 in Training procedure.
(3) **Fuzzification**: The same as Step 5 in Training procedure.
(4) **Activation**: The same as Step 6 in Training procedure.

(5) **Defuzzification**:

A. **Sub-reservoir deffuzification**: The same as Step 6 in the Training procedure.

B. **Defuzzification for each principal component:** Each component at time $t + 1$ is predicted through the linear combination of the outputs from sub-reservoirs and the least squares coefficients using Eq. 9.

$$\hat{y}_{emb(k)}(t+1) = \lambda_0 + \sum_{j=1}^{L} \lambda_j \cdot \hat{y}_{sub(j)}(t+1) \tag{9}$$

C. **Final predicted crisp values for original datasets:** In the last phase, as indicated by Fig. 2, the final predicted values ($\hat{y}_1(t+1)$, $\hat{y}_2(t+1)$, ..., $\hat{y}_N(t+1)$) are generated from the obtained predicted values for each component using inverse KPCA.

4 Computational Experiments

This section is dedicated to evaluating the accuracy of our proposed approach. We conducted all experiments using Python open-source packages including Scikit-Learn, Keras, Tensorflow, PyTorch, Pandas, Numpy, and pyFTS [15].

4.1 Case Studies

We assess our method using two IoT applications: smart buildings and air quality monitoring. Accordingly, two high-dimensional public data sets are employed to examine our proposed technique as detailed in the following.

The first case study is the "Kaggle Smart home with Weather Information" data set (SH-DS) [18]. SH-DS includes 500,910 instances and 29 variables comprising the power consumption of household appliances in KW along with weather information from January 2016 to December 2016 at a frequency of 1 min. It should be noted that the frequency of the original time series is changed from 1 to 10 min in this work.

The "UCI Beijing Multi-Site Air-Quality Data Data Set" (AQB-DS) [19] is the second case study. AQB-DS includes sets of hourly meteorological and air pollutants data from 12 nationally controlled air-quality monitoring stations. It contains 35,065 instances and 132 variables in 12 stations over a 4-year period from March 1st, 2013 to February 28th, 2017.

4.2 Experimental Methodology

The performance accuracy of our proposed approach is evaluated using the Normalized Root Mean Squared Error (NMRSE), which is described in Eq. 10, where $y(t)$ and $\hat{y}(t)$ stand for the actual and forecast values, respectively.

$$NRMSE = \frac{\sqrt{\frac{1}{T}\sum_{i=1}^{T}(y_i - \hat{y}_i)^2}}{Y_{max} - Y_{min}} \tag{10}$$

The sliding window cross-validation technique is utilized to calculate NRMSE, where 75% of each window is used as a training set and 25% for testing. For each variable $y_i(t) \in Y$ of each data set, the samples were divided into 30 windows. For each window, we train and test the methods using respectively the training and test subsets. Noteworthy that the accuracy of the models is evaluated over the test data computing the average values of accuracy metrics obtained for all 30 windows.

To conduct a statistical comparison of the model's performance, we performed the Kruskal-Wallis test with a confidence level of $\alpha = 0.05$. We compared the average NRMSE of all variables in each data set for each window. The proposed method considers the null hypothesis (H_0) as the equality of the average NRMSE errors of all the methods, while the alternative hypothesis (H_1) states that at least one of the means differs from the others. If H_0 is rejected, it becomes necessary to conduct *post hoc* tests to compare the equality of each pair of means. The Wilcoxon test was selected as the *post hoc* test for this study.

5 Results

This section showcases the experimental outcomes of our proposed method and compares its performance against other forecasting algorithms that were tested on the same case studies.

5.1 Parameter Setting

The aim of this part is to analyze the influence of hyper-parameters (HP) on the accuracy of the proposed method as well as baseline algorithms. A well-performing HP configuration is needed to reach the best performance of our proposed model. The number of concepts (κ), the number of components (k), the number of layers or sub-reservoir (L), and the kernel coefficient of KPCA (γ) are the most influential parameters on the model forecasting accuracy. In this case, the activation function (f) is ReLU. Table 1 summarizes the optimal combinations of the HP adjusted separately for each case study. Based on this table, the best performance of our proposed model is obtained with the minimum values of HP as the strengths of our approach. These values are obtained through trial and error, but in the future, the intention is to use random search or Bayesian optimization to better tune the values of HPs.

Also, Table 1 presents the optimal values of HP for competitor models including Vanilla Random Forest (RF), Support vector regression (SVR), XGBoost (XGB), PCA-MO-ENSFTS (PFTS)[16] and KPCA-MO-ENSFTS (KFTS) [16] using randomized search CV. The best HP for RNN methods was obtained with layers = 1, epochs = 300, batch size = 64, learning rate = 0.001, weight decay = 0.1, and optimizer = Adam. Also, the number of hidden neurons is determined considering the number of dimensions and instances of each data set, varying from 3 to 467. In addition, the optimal outcomes for SLSTM could be achieved with epochs = 25, batch size = 32, neurons = 200, and optimizer = Adam, after various settings.

Table 1. Optimal HPs of each data set for the proposed M-PRFCM method and competitor models

Model	HP	SH-DS	AQB-DS
M-PRFCM	κ	3	3
	L	5	2
	k	2	2
	γ	0.9	0.9
RF	Max depth	15	30
	Max leaf nods	20	30
	Min samples leaf	3	2
SVR	C	17.07	25.71
	ϵ	0.15	0.05
KFTS	k	3	10
	κ	60	60
PFTS	k	8	10
	κ	60	60

5.2 Comparison Against Baselines

The key goal of this section is to make a comparison between the obtained results of our proposed model and the baseline methods in terms of average NRMSE over all variables for every data set as exhibited in Table 2. As suggested in this table, KFTS performs very closely to our proposed method in this study. However, KFTS has been equipped with non-stationary fuzzy sets (NSFS). Hence, exploring the integration of NSFS into M-PRFCM could be an intriguing avenue for future research. In addition, the results indicate that M-PRFCM outperforms RNNs in terms of accuracy. To the best of the authors' knowledge, a potential hypothesis is that RNNs excel in accuracy when dealing with data sets with a large number of samples. For instance, GRU, LSTM, and RNN demonstrate promising accuracy performance close to M-PRFCM for SH-DS. Overall, the results highlight the excellent predictive capability of the M-PRFCM model in comparison to the other baseline algorithms for the given data sets.

Table 2. Evaluation of the models' accuracy in terms of average NRMSE

Methods	M-PRFCM	KFTS	PFTS	XGB	RF	SVR	SLSTM	LSTM	GRU	RNN
SH-DS	**0.086**	0.09	0.126	0.126	0.266	1.596	0.639	0.111	0.105	0.106
AQB-DS	**0.107**	0.109	0.126	0.159	0.176	0.274	0.171	0.193	0.189	0.228

Besides, the number of trainable parameters in our proposed method is considerably lower than KFTS, PFTS, and deep models as listed in Table 3. The reason is that the number of trainable parameters in our approach depends only on the number of least squares coefficients which is equal to the number of sub-reservoirs plus one $(L+1)$. Thus, M-PRFCM outperforms the other methods in terms of parsimony. In other words, the other strength of our proposed method is that M-PRFCMM is more parsimonious, cheaper, and less complex than other competing methods. In summary, it can be said that our proposed method is robust and effective with a high ability to dealing with high-dimensional IoT time series.

Table 3. The parsimony of our model against KFTS, PFTS and deep learning methods

data set	M-PRFCM	KFTS	PFTS	LSTM	GRU	RNN
SH-DS	**6**	169	376	2687612	707532	244268
AQB-DS	**3**	423	454	2172	1761	939

5.3 Statistical Testing

As discussed in Sect. 4.2, our proposed method is statistically compared with other baseline approaches using the Kruskall-Wallis test. Table 4 highlights that, for all cases, the null hypothesis (H_0) was rejected based on the corresponding test statistics and p-values. Consequently, Wilcoxon test is utilized as *post-hoc* test to compare the algorithms against each other. Table 5 summarizes the statistical ranking of the models for each data set. The results demonstrate that our proposed approach is superior to other baseline models. More clearly, the results indicate that the first rank belongs to M-PRFCM, followed by KFTS.

Table 4. Kruskal-Wallis mean comparison test results

data set	Statistic	p-value	Result
SH-DS	184.724	5.219e-35	H_0 is rejected
AQB-DS	217.731	6.276e-42	H_0 is rejected

Table 5. The summary of the ranking of the forecasting models

data set	**M-PRFCM**	SLSTM	LSTM	GRU	RNN	RF	SVR	KFTS	PFTS	XGB
SH-DS	**1**	8	5	3	3	3	10	**1**	6	6
AQB-DS	**1**	5	8	7	10	3	7	2	3	5

6 Conclusion

This study develops an FCM-based predictive method called M-PRFCM to forecast high-dimensional time series. M-PRFCM extends the univariate R-HFCM method by mixing the concepts of ESN, FCM, and KPCA to predict multiple outputs considering $\Omega = 1$. The training process of M-PRFCM is accelerated by employing the ESN weight initialization technique to randomly select weights inside each sub-reservoir.

We tested the validity of our proposed method against some popular deep learning and machine learning algorithms using two high-dimensional IoT data sets with 29 and 132 variables. The obtained results verify the superior performance of our proposed model in terms of accuracy and parsimony. Future research intends to upgrade our proposed model to forecast multiple steps ahead and design the model with the ability to handle outliers and missing values.

Acknowledgments. This work has been supported by the Brazilian agencies (i) National Council for Scientific and Technological Development (CNPq), Grants no. 306850/2016-8 and 312991/2020-7; (ii) Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) through the Academic Excellence Program (PROEX) and (iii) Foundation for Research of the State of Minas Gerais (FAPEMIG, in Portuguese), Grant no. APQ-01779-21.

References

1. Kosko, B.: Fuzzy cognitive maps. Int. J. Man-Mach. Stud. **24**, 65–75 (1986)
2. Yang, S., Liu, J.: Time-series forecasting based on high-order fuzzy cognitive maps and wavelet transform. IEEE Trans. Fuzzy Syst. **26**(6), 3391–3402 (2018)
3. Papageorgiou, K.I., Poczeta, K., Papageorgiou, E., Gerogiannis, V.C., Stamoulis, G.: Exploring an ensemble of methods that combines fuzzy cognitive maps and neural networks in solving the time series prediction problem of gas consumption in Greece. Algorithms **12**(11), 235 (2019)
4. Felix, G., Nápoles, G., Falcon, R., Froelich, W., Vanhoof, K., Bello, R.: A review on methods and software for fuzzy cognitive maps. Artif. Intell. Rev. **52**, 1707–1737 (2019)
5. Gao, R., Du, L., Yuen, K.: Robust empirical wavelet fuzzy cognitive map for time series forecasting. Eng. Appl. Artif. Intell. **96**, 103978 (2020)
6. Orang, O., Silva, P., Guimarães, F.: Time series forecasting using fuzzy cognitive maps: a survey. Artif. Intell. Rev. **56**, 7733–7794 (2022)
7. Lu, W., Yang, J., Liu, X.: The linguistic forecasting of time series based on fuzzy cognitive maps. In: 2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), pp. 649–654 (2013)
8. Homenda, W., Jastrzebska, A., Pedrycz, W.: Joining concept's based fuzzy cognitive map model with moving window technique for time series modeling (2014)
9. Yuan, K., Liu, J., Yang, S., Wu, K., Shen, F.: Time series forecasting based on kernel mapping and high-order fuzzy cognitive maps. Knowl.-Based Syst. **206**, 106359 (2020)
10. Wu, K., Liu, J., Liu, P., Yang, S.: Time series prediction using sparse autoencoder and high-order fuzzy cognitive maps. IEEE Trans. Fuzzy Syst. **28**, 3110–3121 (2019)

11. Vanhoenshoven, F., Nápoles, G., Froelich, W., Salmeron, J., Vanhoof, K.: Pseudoinverse learning of Fuzzy Cognitive Maps for multivariate time series forecasting. Appl. Soft Comput. **95**, 106461 (2020)
12. Orang, O., e Silva, P.C.D.L., Silva, R., Guimarães, F.: Randomized high order fuzzy cognitive maps as reservoir computing models: a first introduction and applications. Neurocomputing **512**, 153–177 (2022)
13. Jaeger, H.: The "echo state" approach to analysing and training recurrent neural networks-with an erratum note. Bonn, Germany: German Natl. Res. Center Inf. Technol. GMD Techn. Rep. **148**(34), 13 (2001)
14. Orang, O., Silva, R., e Silva, P.D.L., Guimarães, F.: Solar energy forecasting with fuzzy time series using high-order fuzzy cognitive maps. In: 2020 IEEE International Conference On Fuzzy Systems (FUZZ-IEEE), pp. 1–8 (2020)
15. Silva, P., et al.: PYFTS/pyFTS: Stable version 1.7. - Type hints - New methods - Performance improvements - Bugfixes. Zenodo (2019)
16. Bitencourt, H., Orang, O., Souza, L., Silva, P., Guimarães, F.: An embedding-based non-stationary fuzzy time series method for multiple output high-dimensional multivariate time series forecasting in IoT applications. Neural Comput. Appl. **35**, 9407–9420 (2022). https://doi.org/10.1007/s00521-022-08120-5
17. Papageorgiou, E.: A new methodology for Decisions in Medical Informatics using fuzzy cognitive maps based on fuzzy rule-extraction techniques. Appl. Soft Comput. **11**, 500–513 (2011)
18. Kaggle Smart Home Data Set with weather Information (2021). https://www.kaggle.com/taranvee/smart-home-dataset-with-weather-information. Accessed 28 Ago 2021
19. Dua, D., Graff, C.: UCI Machine Learning Repository. University of California, Irvine, School of Information (2017). http://archive.ics.uci.edu/ml

Review of Offensive Language Detection on Social Media: Current Trends and Opportunities

Lütfiye Seda Mut Altın(✉) and Horacio Saggion

Department of Information and Communication Technologies, Pompeu Fabra University, C/Tànger, 122, 08018 Barcelona, Spain
lutfiyeseda.mut01@estudiant.upf.edu, horacio.saggion@upf.edu

Abstract. Offensive language is defined as derogatory or obscene language that has various forms such as hate speech or cyberbullying. Automated detection of offensive language gains traction due to the high and growing scale of social media user input. In this paper, we provide an overview of the field including background and recent research with a focus on natural language processing. We present a synopsis on the ambiguity in definition and categorization of offensive language, application areas of an automated system, shared tasks organized in this field, dataset creation, model evolution in time through machine learning and deep learning algorithms. Finally challenges and gaps in research are discussed.

Keywords: Natural Language Processing · Social Media · Offensive Language

1 Introduction

At present, the active social media population is reported as more than 4.5 billion worldwide[1]. As the amount of social media content produced by users increases, it emerges the need for better moderation techniques for unwanted content. Therefore, automated detection of offensive text gains a lot of traction focusing on concepts around aggression, hate speech, trolling activities, misogyny, cyberbullying and so on. Offensive language is considered as degrading language that has a negative impact. Examples of offensive (OFF) and not offensive (NOT) texts are given in Table 1, from Semi-Supervised Offensive Language Identification Dataset (SOLID) [54].

There are numerous research and a number of previous reviews on the field of offensive language detection [28,37,64]. Advancements in natural language processing also led to improvements and an increase in the variety of research in this field. Use of machine learning and deep learning algorithms for accurate classification of offensive language and further classification of fine-grained types

[1] https://www.statista.com/topics/1164/social-networks/dossierKeyfigures.

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 62–76, 2024.
https://doi.org/10.1007/978-3-031-56728-5_6

Table 1. Example texts from SOLID dataset

Class	Text
OFF	Somebody come get her she's dancing like a stripper
OFF	@USER We are a country of morons
NOT	This account owner asks for people to think rationally
NOT	Hate the sin not the sinner

of offense are widely researched. Moreover, creating high-quality datasets to train and test the models, as well as methods for evaluating dataset annotation have been studied.

In this paper, we present an overview of the background and current state of offensive language detection on social media. In Sect. 2, we describe our methodology for article search and selection. In Sect. 3, we provide a background around terminology, variations, and definitions, application areas, shared tasks organized on the topic, existing datasets with their differences in classes, and differences on creation steps such as annotation agreement and finally model evolution over time. In Sect. 4, we discuss challenges, gaps, and potential opportunities in the area.

2 Methodology of the Literature Review

While forming the methodology, guidelines from Kitchenham and Templier for writing literature reviews were used for best practice [32,61]. The following resources were considered to do the search on the topic:

Conference Proceedings: According to conference rankings (by Google Scholar on Computational Linguistics) the following top 3 conferences for the last 2 years were examined: Meeting of the Association for Computational Linguistics (ACL), Conference on Empirical Methods in Natural Language Processing (EMNLP), Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (HLT-NAACL).

Digital Libraries: Among digital libraries Web of Science has been chosen for key term search since it is considered more reliable and has information on citation. Keyword groups have been searched on the platform and the results of the searches have been collected. Time interval between 2017-01-01 and 2023-05-31 was considered for the recent papers while no time constraint imposed for the background. Furthermore, additional resources have been included for relevant sections including areas other than computer science.

As search terms, firstly, keywords that can describe the offensiveness in various ways have been selected. The European Commission against Racism and

Intolerance (ECRI) Glossary is also benefited from for the search keyword selection.[2] These keywords are: **"offensive", "hate speech", "racism", "sexism", "cyberbullying"**. For the conference papers, these keywords were used directly in the search within the titles of the accepted papers. For the digital library search these keywords were used along with complementary terms to help discriminate articles from fields such as social science. For this purpose, the following search term list was applied on Web of Science: **1. "offensive" "text classification" 2. "hate speech" "text classification" 3. "cyberbullying" "detection" 4. "racism" "text classification" 5. "sexism" "text classification"** for the recent papers (2017–2022); **6. "offensive" "language detection"**, **7."hate speech" "detection"** and **8. "cyberbullying" "detection"** for a fundamental background. Papers were sorted by number of citations in descending order.

After obtaining the search results, overlapping papers were excluded, publications only written in English have been taken into account and publications on other fields were excluded such as social sciences.

3 Background

3.1 Definition and Variations

Offensive language is defined as the term that is applied to hurtful, derogatory or obscene comments made.[3] Whereas, the United Nations indicates that **hate speech** is used in common language as loosely referring to "offensive discourse targeting a group or an individual based on inherent characteristics - such as race, religion or gender - and that may threaten social peace".[4]

On the other hand, different terms are used in the literature for automatic text detection to refer to the same concept as offensive language such as ***aggression*** [56,58], ***toxic*** [38,40,62], ***abusive*** [66] or ***threatening*** [35] language. Another specific term that is commonly used in the field is cyberbullying. **Cyberbullying** is a generic term which is defined as "bullying that takes place over digital devices and includes sending, posting or sharing negative, harmful, false, or mean content about someone else".[5]

Furthermore, more specific concepts under the umbrella "hate speech" have been considered. These concepts are usually based on the target group. They include but are not limited to particular concepts such as racism, sexism, homophobia, and so on. Additionally, very few examples of solely ideological hate speech identification towards the right wing in Germany [29]. Figure 1 shows a hierarchical schema of our attempt to clarify the relations of terms around the concept.

[2] https://www.coe.int/en/web/european-commission-against-racism-and-intolerance/ecri-glossary.

[3] https://thelawdictionary.org/offensive-language/.

[4] https://www.un.org/en/hate-speech/understanding-hate-speech/what-is-hate-speech.

[5] https://www.stopbullying.gov/cyberbullying/what-is-it.

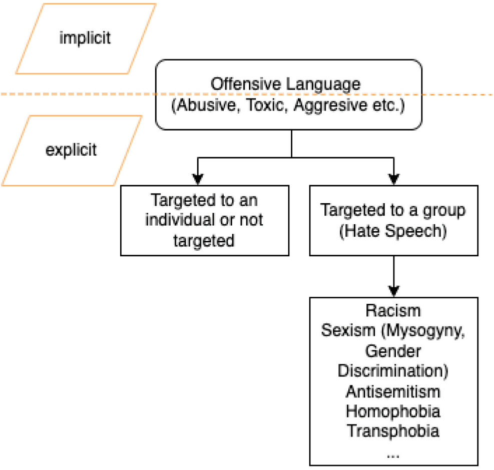

implicit
explicit
Offensive Language
(Abusive, Toxic, Aggresive etc.)
Targeted to an individual or not targeted
Targeted to a group (Hate Speech)
Racism
Sexism (Mysogyny, Gender Discrimination)
Antisemitism
Homophobia
Transphobia
...

Fig. 1. Hierarchical therminology schema

Moreover, Wiegand et al. [65] drew attention to the lack of good performance on detection of implicit abusive language (i.e. not conveyed by explicit offensive words) and presented a list of **sub-types of implicit abusive language** with divide-and-conquer idea behind it. The sub-types they recommended are '**stereotypes'**, **perpetrators**' (meaning a person committing an illegal, criminal or evil act), **'comparisons'** (e.g. *"You sing like a dying bird"*), **'dehumanization'** (as the act of perceiving people as less than human, e.g. *"I own my wife and her money."*), **'euphemistic constructions'** (e.g. *"You inspire my inner serial killer."* actually being an equivalent of *"I want to kill you."*), **'call for action'** (the author asking something, typically some form of punishment), **'multimodal abuse'** (i.e. harmful content of a micropost is hidden in the nontextual components or results as an interplay of text and image/video), **'phenomena requiring world knowledge and inferences'** (as sub-types: **jokes**, **sarcasm**, **rhetorical question**) and finally, **other implicit abuse** to cover further cases.

3.2 Motivation and Application Areas

The increasing amount of social media input makes human moderation impossible while traditional rule-based (e.g. black lists of words) systems are insufficient to provide good coverage. Therefore the need for an efficient automated detection mechanism gains a lot of traction.

Previous research has shown a strong negative relation between cyberbullying and young people's mental health [37]. Earlier studies claim that derogatory language aiming at minority groups leads to political radicalization and worsens intergroup interactions [4].

Considering ethical, sociological and psychological impacts, the demand for an efficient mechanism is quite high in various application areas. In private sector, tech companies and user input based platforms want to increase audience engagement and protect their brand by removing unwanted content as efficiently and quickly as possible. The 2018 Content Moderation report indicated that 27% of respondents of the Digital Trust Survey stated that they would stop using a social platform if it continued to allow harmful content.[6] As stated in the Business Journal from the Wharton School of the University of Pennsylvania (Jan, 2022), Facebook alone has committed to allocating 5% of the firm's revenue, $3.7 billion, to content moderation (please note that it is overall content moderation including text, image, video etc.)[7].

As Klonick [34] summarized the development of online speech moderation, the major social media platforms such as Facebook and Youtube did not even have clear public policies and community standards until the late 2000s, and yet since then they have been developing and improving the scope and definition of their policies, user feedback mechanisms and internationality of the moderation. Moreover, the increase in the usage of streaming platforms (e.g. Twitch) emphasized the need for real-time content moderation which requires speeds that is not always possible with manual moderation.

All in all, due to its scalability and speed, AI-based content moderation is on an increasing demand, while accuracy remaining the biggest challenge at the moment.

3.3 Shared Tasks

Shared tasks are challenges or competitions organized by research community that enable teams of researchers submit systems to solve specific tasks. Escartin et al., conducted a survey and reported that, in the NLP community, shared tasks are generally celebrated as an important factor for advancement of the field [17]. Among various specific tasks in the field of NLP, from news article similarity[8] to patronizing language detection[9], identification of various forms of offensive language is quite popular.

As shown in Fig. 2, the main data source of datasets used in the previous shared tasks is Twitter. In terms of the language of the datasets; the most common is English with 11 dataset, followed by Spanish with 7, German and Hindi with 4, Arabic with 3, Italian with 2 and then the others including Bengali, Danish, Greek,Marathi, Turkish, Urdu and Vietnamese with only 1.

In terms of the participant and winner models, it is seen that throughout the years, submissions trained on neural network models have increased compared to non-neural networks. Support Vector Machine (SVM) [12], Logistic Regression, Random Forest, Naive Bayes, Decision Trees were the popular non-neural

[6] https://store.businessinsider.com/products/the-content-moderation-report.

[7] https://knowledge.wharton.upenn.edu/article/social-media-firms-moderate-content/.

[8] https://competitions.codalab.org/competitions/33835.

[9] https://competitions.codalab.org/competitions/34344.

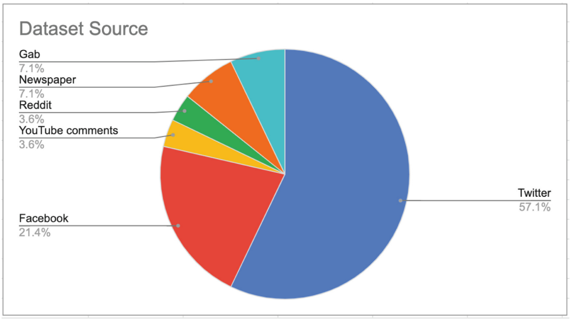


Fig. 2. Source of the datasets used in the shared tasks

approaches while recurrent neural network (RNN) [55], convolutional neural network (CNN) [21], long short-term memory (LSTM) [25], bi-LSTM, GRU were the popular deep learning architectures. Also, ensemble classification systems are in general highly preferred among participants. For the tasks on more than one language, some participants also submitted their results with a multilingual approach where they trained their model with multiple languages.

For example, in the earlier tasks such as GermEval, TRAC, EVALITA in 2018, the ratio of total participant submission models were around 48% non-Neural (mostly SVM and Logistic Regression) and 52% Neural networks [6,36,67]. Whereas in a recent task, the latest EXIST [52], it is reported that all participants used some kind of transformer-based system except one team; more specifically, majority used Bidirectional Encoder Representations from Transformers (BERT)[16] or versions of BERT including multilingual BERT - mBERT, Spanish version of BERT called BETO, RoBERTa, DeBERTa, multilingual version of RoBERTa called XLM-R or other transformer versions. Also in OSACT (2022), it is reported that the participant teams used different fine-tuned transformer versions such as AraBERT, mBert, XLMRoberta etc. where the highest ranking submissions used an ensemble of different transformers [24].

3.4 Datasets

Reviewing the recent datasets it is seen that as well as the differences in labeling and classification schema, annotation mechanisms have also differences such as different numbers of annotators and evaluation of the agreement between anotators. For example, usually three or more annotators annotated each instance in the datasets. For the annotation agreement calculations *Fleiss'sKappa* was considered for some datasets [19,71], *Krippendorf's* was used in another one [38]. Moreover, annotator's profile was also often in consideration. In some datasets it is stated that variety has been maintained among annotators, the details have, however, been kept private. In some datasets it is revealed, as an example, for the Levantine hate speech dataset by Mulki et al., genders of the 3 annotators has been chosen as one male and two females [46].

In terms of the data sources, the most common one appears to be social media platforms such as Twitter due to the limited short text structure. However, also sources such as newspapers and platforms known as more liberal such as 'gab.com' have also been considered. Data collection is usually based on certain keywords and hashtags; however at times data is collected based on searches following important events that have gone viral. A different example of data collection strategy is seen from the hate speech dataset by Mubarak et al. [45] which they collected tweets using emojies existed in offensive text which are extracted from previous datasets by Zampieri et al. [72] and Chowdhury et al. [10].

Rosa et al. [53] examined earlier cyberbullying datasets (from 2011 to 2018) and reported that the majority of the datasets are in English, mainly labelled by 3 annotators, with variety of size from 2K to 85K instances and from data sources including not only Twitter, YouTube, Instagram but also Formspring, AskFM, MySpace.

Regardless of which the classes are considered, the majority of the datasets we are aware of have a binary approach to labeling; however Hada et al. created the first dataset based on the degree of offensiveness, where each instance has a score between -1 (maximally positive) and 1 (maximally offensive) [23].

More recently, counter-narrative generation started to be researched as an alternative solution [11,60,73]. Counter-narratives are texts that withstand hate speech with fact-bound arguments or alternative viewpoints. Also in dataset creation, examples of hate speech/counter narrative datasets started to emerge [18]. In a recent study, GPT-2 was utilized to generate synthetic training data for the model [47,50].

Further examples of dataset creation have concentrated on aspects of racial bias. Sap et al. examined annotators' insensitivity to differences in dialect and showed that when annotators are made explicitly aware of an African-American English tweet's dialect they are significantly less likely to label the tweet as offensive [57]. Davidson et al. also examined racial bias by training models on different datasets and finally referred that racial bias exists in datasets, as classifiers trained on them tend to predict that tweets written in African-American English are abusive at substantially higher rates [13].

The area of research in various languages also keep expanding with new datasets as in most recent Korean, Chinese, Turkish studies [15,30,31].

3.5 Methodology and Model Evolution

In general, a detection system starts with pre-processing of the data, splitting the data as train and test (typically as 80% to 20%), feature extraction, model training and finally the classified output.

Pre-processing depends on the data format and involves punctuation removal, stopword removal, tokenization, stemming, lemmatization, smiley and slang conversion.

Features used vary from lexical features such as Bag of Words (BoW), n-grams, use of offensive word dictionaries to syntactic features such as use of identifiers i.e. second person pronouns and user level features [9]. Common techniques for feature extraction includes TF-IDF, BoW, sentiment, Part of Speech (PoS) tag, word embeddings (GloVe [48], Word2Vec [42], FastText [5], ELMO [49]). Jahan et al. reported in a systematic review that the most used features were word embeddings and TF-IDF; additionally as for the word embeddings most commonly used ones were Word2Vec and FastText [28].

Additionally, further features are explored by researchers. Galan et al. proposed an approach for cyberbullying detection based on the idea that every trolling profile is followed by the real profile of the user behind the trolling one and generated the hypothesis that it is possible to link an account of fake profile to the real profile and analyse different features of the profile including text [22]. McGillivray et al., recently draw attention to the fact that the meaning of a word may change over time and proposed a time dependent lexical feature approach, meaning that they applied an algorithm to detect semantic change over a 2 years' time to detect words whose semantics has changed and either acquired or lost offensiveness [41]. Casavantes et al. used metadata features such as tweet creation time of the day or account age and reported that utilizing metadata gave better results. While doing this, they utilized three different learning models as their consideration being classical (BoW), advanced (Glove) and state-of-the-art (BERT) text representations and reported statistically significant difference with use of metadata [8].

In terms of models, some of the earliear research have utilized WEKA (standard machine learning software developed at University of Waikato) [26] and applied traditional algorithms as in Rezavi et al. [51]. They selected a classifier and created an abusive language dictionary and assigned a weight (1–5) to the entries in it then applied models from multi-level classifiers boosted by the dictionary. Akhter et al., performed a series of experiments and reported that character n-grams outperformed word n-grams [2].

As Rosa et al. noted, earlier research on cyberbullying detection that dates back from 2011 to 2017; detection mechanisms were mainly based on traditional machine learning algorithms such as SVM [53]. Van Hee et al. also showed that Support Vector Machine had better results than baseline systems that are based on keywords and word unigrams [63]. Similarly, on hate speech specific detection SVM has commonly been experimented [39]. Fortuna et al. also stated in their review back in 2018 that most chosen algorithms in the hate speech detection as traditional algorithms being the most frequent as SVM and followed by Random Forest, Decision Tree, Logistic Regression and Naive Bayes [20].

More recently, deep neural network models gained great attention [1]. In 2017, Badjatiya et al. reported that deep learning methods significantly outperformed state-of-the-art char/word n-gram methods on hate speech detection utilizing CNN, LSTM and FastText [3]. Mozafari et al. obtained high F1 scores on hate speech detection with their transfer learning approach including BERT [16] and CNN [44]. In their systematic review on hate speech, Jahan et al. identified 96 documents with deep learning approaches and noted that among deep learning algorithms, BERT is the most commonly used one with 38% although it has been released quite recently, then LSTM, CNN, bi-LSTM, GRU and combination of these followed respectively by the percentage of usage [28].

Lately, language generation has also been involved in hate speech detection. Chung et al. proposed methods for improving counter narrative generation for hate speech detection [11]. Another example is that Wullach et al. used GPT for generating synthetic hate speech data from available labeled examples [68]. Depending on the dataset size and class distribution, data augmentation is commonly utilized on imbalanced datasets to improve performance and prevent issues such as overfitting. Ilan et al. reported they improved the performance with an augmentation method that they introduced as input real unlabelled data unlike real labeled or synthetic data (using a generative model), which their approach made use of online platforms in which people are specifically asked to be bullying (such as subreddit r/RoastMe/ platform) [27]. Another recent data oriented approach was reported by Yang et al. in which they asked people to generate offensive arguments deliberately aiming to be less sensitive to lexical overlap [70]. With researchers drawing attention to the scarcity of labelled data, Sarracén et al. presented a study in which their model was composed of Convolutional Graph Neural Network (GNN) and reported that it performed better then state-of-the-art models on small datasets [14]. Besides, Tanvir et al. reported that with their GAN-BERT approach, they obtained a promising result with a small sized dataset in Bengali language [59]. Breazzano et al. experimented a transformer-based architecture combining BERT with multi-task and generative adversarial learning (MT-GAN-BERT) for six different abusive language classification tasks enabling semi-supervised learning and reported conclusions as decrease in computational costs without a considerable decrease in prediction quality [7].

From a more generic perspective, Minaee et al. reviewed deep learning approaches for various text classification tasks through more than 150 deep learning models and 40 datasets. They emphasized the fast progress on text classification over the recent years thanks to contributions such as neural embedding, attention mechanism, self attention, transformer as well as showing again that deep learning models resulted in significant improvements compared to non-deep learning models. They also discussed that to choose the best neural network for a classification task, there is not one single solution and it varies depending on things such as nature of the domain, application areas, availability of the labels [43].

4 Discussion and Conclusions

In this paper, we presented an overview of detection of offensive language in social media text by natural language processing. We tried to bring light on the ambiguity in terminology and classification along with different data classes given in the datasets provided through shared tasks that accepts many experiment inputs each.

For this review we have not taken into account the research that considers multi-modals such as including image and video into consideration along with text and author metadata.

4.1 Challenges

As in the 'garbage in garbage out' principle, issues around data constitutes an important part of the challenges in the field [64]. Complexity of the definition as mentioned in previous sections sometimes causes ambiguity in dataset classes, annotation and combining similar datasets. Moreover, due to the nature of social media, most of the text contains slang, variety of smileys and grammatically incorrect sentences therefore they are hard to predict structures. In addition, context switch, use of different dialects, not enough available data source for all languages are also other language related challenges.

Furthermore, as social media is our main consideration; social media entries are subject to change in time relatively quick. For example, a new term or an acronym which did not exist in a language before or exist with neutral or positive meaning before, might emerge in a new incident such as a new released television advertisement or a political scandal or a viral video and after that people might start using it with a secondary negative meaning.

4.2 Gaps in the Research

In spite of the traction on the field, some of the gaps in the research can be identified as follows:

- Although there are numerous datasets and experiments on different languages, the majority of the research is on English language. However, we are not aware of any research with comparison of the perception on '*native speaker*' and '*non-native speaker*' point of view.
- The amount of research on more particular topics such as migration or refugee status, disability etc. is relatively low in comparison to more generic classification such as offensive/non-offensive. There is an opportunity to create datasets and work on classification on more specialized areas of hate speech.
- Even though recently there are more research on multi-lingual classification, still the research so far is limited and there is opportunity to study languages other than English and research on multi-lingual models.
- Dataset sources are mostly social media user input sometimes along with user metadata and there are very few examples of other sources such as: Song lyrics, movie or TV show dialogs (sub-titles) (Informative documents such as Wikipedia are not in consideration).
- Images and videos are important parts of social media. For instance, Instagram users share over one million memes daily[10]. However, research that combines text with image or video is quite limited. Although there has been

[10] https://webtribunal.net/blog/meme-statistics/gref.

some research, such as (Yang et al. 2019)'s multi-modal [69] and (Kiela et al. 2020)'s 'Hateful Meme Challenge' [33] there is still opportunity on this aspect.

Acknowledgments. We acknowledge support from the Maria de Maeztu Units of Excellence Programme CEX2021-001195-M, funded by MICIU/AEI /10.13039/50110001103 and with the support from Departament de Recerca i Universitats de la Generalitat de Catalunya (ajuts SGR-Cat 2021).

References

1. Agrawal, S., Awekar, A.: Deep learning for detecting cyberbullying across multiple social media platforms. In: Pasi, G., Piwowarski, B., Azzopardi, L., Hanbury, A. (eds.) ECIR 2019. LNCS, vol. 10772, pp. 141–153. Springer, Heidelberg (2018). https://doi.org/10.1007/978-3-319-76941-7_11
2. Akhter, M.P., Jiangbin, Z., Naqvi, I.R., Abdelmajeed, M., Sadiq, M.T.: Automatic detection of offensive language for urdu and roman urdu. IEEE Access **8**, 91213–91226 (2020)
3. Badjatiya, P., Gupta, S., Gupta, M., Varma, V.: Deep learning for hate speech detection in tweets. In: Proceedings of the 26th International Conference on World Wide Web Companion, pp. 759–760 (2017)
4. Bilewicz, M., Soral, W.: Hate speech epidemic. The dynamic effects of derogatory language on intergroup relations and political radicalization. Polit. Psychol. **41**, 3–33 (2020)
5. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. Trans. Assoc. Comput. Linguist. **5**, 135–146 (2017)
6. Bosco, C., Felice, D., Poletto, F., Sanguinetti, M., Maurizio, T.: Overview of the evalita 2018 hate speech detection task. In: EVALITA 2018-Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian, vol. 2263, pp. 1–9. CEUR (2018)
7. Breazzano, C., Croce, D., Basili, R.: Multi-task and Generative Adversarial Learning for Robust and Sustainable Text Classification. In: Bandini, S., Gasparini, F., Mascardi, V., Palmonari, M., Vizzari, G. (eds.) AIxIA 2021. LNCS, vol. 13196, pp. 228–244. Springer, Heidelberg (2022). https://doi.org/10.1007/978-3-031-08421-8_16
8. Casavantes, M., Aragón, M.E., González, L.C., Montes-y Gómez, M.: Leveraging posts' and authors' metadata to spot several forms of abusive comments in twitter. J. Intell. Inf. Syst. **61**, 519–539 (2023)
9. Chen, Y., Zhou, Y., Zhu, S., Xu, H.: Detecting offensive language in social media to protect adolescent online safety. In: 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing, pp. 71–80. IEEE (2012)
10. Chowdhury, S.A., Mubarak, H., Abdelali, A., Jung, S., Jansen, B.J., Salminen, J.: A multi-platform arabic news comment dataset for offensive language detection. In: Proceedings of the 12th Language Resources and Evaluation Conference, pp. 6203–6212 (2020)
11. Chung, Y.L., Tekiroglu, S.S., Guerini, M.: Towards knowledge-grounded counter narrative generation for hate speech. arXiv preprint arXiv:2106.11783 (2021)

12. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)
13. Davidson, T., Bhattacharya, D., Weber, I.: Racial bias in hate speech and abusive language detection datasets. arXiv preprint arXiv:1905.12516 (2019)
14. De la Peña Sarracén, G.L., Rosso, P.: Convolutional graph neural networks for hate speech detection in data-poor settings. In: Natural Language Processing and Information Systems: 27th International Conference on Applications of Natural Language to Information Systems, NLDB 2022, Valencia, Spain, June 15–17, 2022, Proceedings, pp. 16–24. Springer, Heidelberg (2022). https://doi.org/10.1007/978-3-031-08473-7_2
15. Deng, J., et al.: Cold: a benchmark for chinese offensive language detection. arXiv preprint arXiv:2201.06025 (2022)
16. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
17. Escartín, C.P., Lynn, T., Moorkens, J., Dunne, J.: Towards transparency in nlp shared tasks. arXiv preprint arXiv:2105.05020 (2021)
18. Fanton, M., Bonaldi, H., Tekiroglu, S.S., Guerini, M.: Human-in-the-loop for data collection: a multi-target counter narrative dataset to fight online hate speech. arXiv preprint arXiv:2107.08720 (2021)
19. Fortuna, P., da Silva, R.R., Wanner, L., Nunes, S., et al.: A hierarchically-labeled Portuguese hate speech dataset. In: Proceedings of the Third Workshop on Abusive Language Online, pp. 94–104 (2019)
20. Fortuna, P., Nunes, S.: A survey on automatic detection of hate speech in text. ACM Comput. Surv. (CSUR) **51**(4), 1–30 (2018)
21. Fukushima, K., Miyake, S.: Neocognitron: a self-organizing neural network model for a mechanism of visual pattern recognition. In: Amari, S., Arbib, M.A. (eds.) Competition and Cooperation in Neural Nets, pp. 267–285. Springer, Heidelberg (1982). https://doi.org/10.1007/978-3-642-46466-9_18
22. Galán-García, P., de la Puerta, J.G., Gómez, C.L., Santos, I., Bringas, P.G.: Supervised machine learning for the detection of troll profiles in twitter social network: application to a real case of cyberbullying. Logic J. IGPL **24**(1), 42–53 (2016)
23. Hada, R., Sudhir, S., Mishra, P., Yannakoudakis, H., Mohammad, S.M., Shutova, E. Ruddit: norms of offensiveness for english reddit comments. arXiv preprint arXiv:2106.05664 (2021)
24. Hassan, S., Samih, Y., Mubarak, H., Abdelali, A., Rashed, A., Chowdhury, S.A.: Alt submission for osact shared task on offensive language detection. In: Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection, pp. 61–65 (2020)
25. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
26. Holmes, G., Donkin, A., Witten, I.H.: Weka: a machine learning workbench. In: Proceedings of ANZIIS'94-Australian New Zealand Intelligent Information Systems Conference, pp. 357–361. IEEE (1994)
27. Ilan, T., Vilenchik, D.: Harald: augmenting hate speech data sets with real data. In: Findings of the Association for Computational Linguistics: EMNLP 2022, pp.2241–2248 (2022)
28. Jahan, M.S., Oussalah, M.: A systematic review of hate speech automatic detection using natural language processing. arXiv preprint arXiv:2106.00742 (2021)
29. Jaki, S., De Smedt, T.: Right-wing german hate speech on twitter: analysis and automatic detection. arXiv preprint arXiv:1910.07518 (2019)

30. Jeong, Y., et al.: Kold: Korean offensive language dataset. arXiv preprint arXiv:2205.11315 (2022)
31. Karayiğit, H., Akdagli, A., Aci, Ç.İ: Homophobic and hate speech detection using multilingual-bert model on Turkish social media. Inf. Technol. Control **51**(2), 356–375 (2022)
32. Keele, S., et al.: Guidelines for performing systematic literature reviews in software engineering. Technical report, ver. 2.3 ebse technical report. ebse (2007)
33. Kiela, D., et al.: The hateful memes challenge: competition report. In: NeurIPS 2020 Competition and Demonstration Track, pp. 344–360. PMLR (2021)
34. Klonick, K.: The new governors: the people, rules, and processes governing online speech. Harv. L. Rev. **131**, 1598 (2017)
35. Kumar, A., Saumya, S., Roy, P.K.: Abusive and threatening language detection from urdu social media posts: a machine learning approach (2021)
36. Kumar, R., Ojha, A.K., Malmasi, S., Zampieri, M.: Benchmarking aggression identification in social media. In: Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018), pp. 1–11 (2018)
37. Kwan, I., et al.: Cyberbullying and children and young people's mental health: a systematic map of systematic reviews. Cyberpsychol. Behav. Soc. Netw. **23**(2), 72–82 (2020)
38. Leite, J.A., Silva, D.F., Bontcheva, K., Scarton, C.: Toxic language detection in social media for brazilian portuguese: new dataset and multilingual analysis. arXiv preprint arXiv:2010.04543 (2020)
39. MacAvaney, S., Yao, H.-R., Yang, E., Russell, K., Goharian, N., Frieder, O.: Hate speech detection: challenges and solutions. PLoS ONE **14**(8), e0221152 (2019)
40. Makhnytkina, O., Matveev, A., Bogoradnikova, D., Lizunova, I., Maltseva, A., Shilkina, N.: Detection of toxic language in short text messages. In: Karpov, A., Potapova, R. (eds.) SPECOM 2020. LNCS, pp. 315–325. Springer, Heidelberg (2020). https://doi.org/10.1007/978-3-030-60276-5_31
41. McGillivray, B., et al.: Leveraging time-dependent lexical features for offensive language detection. In: Proceedings of the 1st Workshop of Ever Evolving NLP, EMNLP 2022 (2022)
42. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)
43. Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., Gao, J.: Deep learning-based text classification: a comprehensive review. ACM Comput. Surv. (CSUR) **54**(3), 1–40 (2021)
44. Mozafari, N., Farahbakhsh, R., Crespi, N.: A bert-based transfer learning approach for hate speech detection in online social media. In: Cherifi, H., Gaito, S., Mendes, J., Moro, E., Rocha, L. (eds.) COMPLEX NETWORKS 2019, vol. 881, pp. 928–940. Springer, Heidelberg (2020). https://doi.org/10.1007/978-3-030-36687-2_77
45. Mubarak, H., Al-Khalifa, H., Al-Thubaity, A.M.: Overview of osact5 shared task on arabic offensive language and hate speech detection. In: Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools with Shared Tasks on Qur'an QA and Fine-Grained Hate Speech Detection, pp. 162–166 (2022)
46. Mulki, H., Haddad, H., Ali, C.B., L-hsab, H.A.: A levantine twitter dataset for hate speech and abusive language. In: Proceedings of the Third Workshop on Abusive Language Online, pp. 111–118 (2019)
47. Nouri, N.: Data augmentation with dual training for offensive span detection. In: Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 2569–2575 (2022)

48. Pennington, J., Socher, R., Manning, C.D.: Glove: global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543 (2014)
49. Peters, M.E., et al.: Deep contextualized word representations. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, New Orleans, Louisiana, vol. 1 (Long Papers), pp. 2227–2237. Association for Computational Linguistics (2018)
50. Radford, A., Jeffrey, W., Child, R., Luan, D., Amodei, D., Sutskever, I., et al.: Language models are unsupervised multitask learners. OpenAI Blog **1**(8), 9 (2019)
51. Razavi, A.H., Inkpen, D., Uritsky, S., Matwin, S.: Offensive language detection using multi-level classification. In: Farzindar, A., Keselj, V. (eds.) Canadian AI 2010. LNCS, vol. 6085, pp. 16–27. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13059-5_5
52. Rodríguez-Sánchez, F., et al.: Overview of exist 2022: sexism identification in social networks. Procesamiento del Lenguaje Natural **69**, 229–240 (2022)
53. Rosa, H., et al.: Automatic cyberbullying detection: a systematic review. Comput. Human Behav. **93**, 333–345 (2019)
54. Rosenthal, S., Atanasova, P., Karadzhov, G., Zampieri, M., Nakov, P.: Solid: a large-scale semi-supervised dataset for offensive language identification. arXiv preprint arXiv:2004.14454 (2020)
55. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. Nature **323**(6088), 533–536 (1986)
56. Sadiq, S., Mehmood, A., Ullah, S., Ahmad, M., Choi, G.S., On, B.W.: Aggression detection through deep neural model on twitter. Future Gener. Comput. Syst. **114**, 120–129 (2021)
57. Sap, M., Card, D., Gabriel, S., Choi, Y., Smith, N.A.: The risk of racial bias in hate speech detection. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 1668–1678 (2019)
58. Si, S., Datta, A., Banerjee,S., Naskar, S.K.: Aggression detection on multilingual social media text. In: 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1–5. IEEE (2019)
59. Tanvir, R., et al.: A gan-bert based approach for bengali text classification with a few labeled examples. In: Omatu, S., Mehmood, R., Sitek, P., Cicerone, S., Rodriguez, S. (eds.) DCAI 2022. LNCS, vol. 583, pp. 20–30. Springer, Heidelberg (2022). https://doi.org/10.1007/978-3-031-20859-1_3
60. Tekiroglu, S.S., Chung, Y.L., Guerini, M.: Generating counter narratives against online hate speech: data and strategies. arXiv preprint arXiv:2004.04216 (2020)
61. Templier, M., Paré, G.: A framework for guiding and evaluating literature reviews. Commun. Assoc. Inf. Syst. **37**(1), 6 (2015)
62. Van Aken, B., Risch, J., Krestel, R., Löser, A.: Challenges for toxic comment classification: an in-depth error analysis. arXiv preprint arXiv:1809.07572 (2018)
63. Van Hee, C., Jacobs, G., Emmery, C., Desmet, B., Lefever, E., Verhoeven, B., De Pauw, G., Daelemans, W., Hoste, V.: Automatic detection of cyberbullying in social media text. PLoS ONE **13**(10), e0203794 (2018)
64. Vidgen, B., Derczynski, L.: Directions in abusive language training data, a systematic review: Garbage in, garbage out. PLoS ONE **15**(12), e0243300 (2020)
65. Wiegand, M., Ruppenhofer, J., Eder, E.: Implicitly abusive language–what does it actually look like and why are we not getting there? In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 576–587. Association for Computational Linguistics (2021)

66. Wiegand, M., Ruppenhofer, J., Kleinbauer, T.: Detection of abusive language: the problem of biased datasets. In: Proceedings of the 2019 conference of the North American Chapter of the Association for Computational Linguistics: human language technologies, vol. 1 (long and short papers), pp. 602–608 (2019)
67. Wiegand, M., Siegel, M., Ruppenhofer, J.: Overview of the germeval 2018 shared task on the identification of offensive language. In: Overview of the Germeval 2018 Shared Task on the Identification of Offensive Language (2018)
68. Wullach, T., Adler, A., Minkov, E.: Fight fire with fire: fine-tuning hate detectors using large samples of generated hate speech. arXiv preprint arXiv:2109.00591 (2021)
69. Yang, F., et al.: Exploring deep multimodal fusion of text and photo for hate speech classification. In: Proceedings of the Third Workshop on Abusive Language Online, pp. 11–18 (2019)
70. Yang, K., Jang, W., Cho, W.I.: Apeach: attacking pejorative expressions with analysis on crowd-generated hate speech evaluation datasets. arXiv preprint arXiv:2202.12459 (2022)
71. Zampieri, M., Malmasi, S., Nakov, P., Rosenthal, S., Farra, N., Kumar, R.: Predicting the type and target of offensive posts in social media. arXiv preprint arXiv:1902.09666 (2019)
72. Zampieri, N., et al. Semeval-2020 task 12: multilingual offensive language identification in social media (offenseval 2020). arXiv preprint arXiv:2006.07235 (2020)
73. Zhu, W., Bhat, S.: Generate, prune, select: a pipeline for counterspeech generation against online hate speech. arXiv preprint arXiv:2106.01625 (2021)

Text Mining and Sentimental Analysis to Distinguish Systems Thinkers at Various Levels: A Case Study of COVID-19

Mohammad Nagahisarchoghaei[1], Morteza Nagahi[2], and Harun Pirim[3](✉)

[1] Mississippi State University, Mississippi State, MS 39762, USA
[2] DiversityInc Media LLC, West Palm Beach, FL 33405, USA
[3] North Dakota State University, Fargo 58102, USA
harun.pirim@ndsu.edu

Abstract. Limited research exists on how experts' Systems Thinking (ST) skills can be linked to their tweets and sentiments. This study employs text mining and social media analysis to explore the relationship between experts' ST and their tweets, specifically focusing on COVID-19. Twitter is crucial for information dissemination, but misinformation can spread during a pandemic like COVID-19. By analyzing the emotional and sentimental aspects of tweets from 55 COVID-19 experts, we identified three distinct clusters with significant differences in emotions and sentiments. This study introduces a novel framework using NLP, text mining, and sentiment analysis to assess the systems thinking skills of COVID-19 experts.

Keywords: Social Media Analytics · Systems Thinking · Sentiment Analysis

1 Introduction

Social Media Analysis (SMA) gathers and examines information from individuals and users engaged on various social media platforms in order to tackle intricate societal issues. SMA offers a comprehensive examination of media data to cultivate meaningful understandings. In this study, the SMA was utilized to examine the individual's capabilities as the level of Systems Thinking (ST) and to investigate the relationship between COVID-19 experts' systems thinking and their tweets and sentiments. The current literature review reveals research gaps relevant to evaluating if the ST skills of experts could be correlated with their tweets as well as their level of emotion and sentiment. This article is organized as follows: Sect. 1 presents an overview of the current methods and tools used in the Twitter analysis, Sect. 2 describes the procedure of using Text mining and the natural language processing (NLP) methods in extracting data to reach our research goal based on the research gaps. Section 3 interprets the results of the analysis. Section 4 concludes by discussing the summary of the results, implications, limitations, and future work.

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 77–88, 2024.
https://doi.org/10.1007/978-3-031-56728-5_7

1.1 Text Mining of Twitter Data

Social networking and communication platforms such as Facebook, Twitter, and Google+ empower individuals to distribute and articulate their viewpoints, share messages on a global scale, and interact with diverse networks. There are many recent studies employing sentiment analysis using Twitter data [1–3] across various contexts. In their study, Kharde and Sonawane [4] performed an assessment of sentiment analysis techniques, which encompassed vocabulary-based approaches as well as machine learning methods. They observed that machine learning techniques like SVM and Bayes demonstrated remarkable precision and can be regarded as conventional learning methods. Sailunaz and Alhajj [5] investigated the utilization of Twitter users' posts to generate recommendations by scrutinizing emotions and sentiment. They found that even if tweets and comments seemed to have contradictory sentiments, they were actually consistent with the text. Teresa and Syed [6] explored scientists' perceptions of interdisciplinarity on Twitter, emphasizing positive encounters that could foster motivation, innovation, development, and prosperity. Plunz et al. [7] utilized social media to analyze the fluctuating emotions in metropolitan green spaces, revealing that in-park tweets in Manhattan express more negative sentiments compared to tweets from areas farther from parks. Schumaker et al. [8] posed a hypothetical question regarding whether sentiment in tweets can serve as a crucial intermediary to predict outcomes and if the magnitude of results can be anticipated based on the level of opinion. The authors found that concealed data within tweet sentiment can indeed have predictive implications for the design of automated betting systems. Arora and Kansal [9] proposed a text standardization model using deep convolutional neural networks for sentiment analysis of unstructured Twitter data, demonstrating its effectiveness in standardization and opinion analysis. Wakade et al. [10] demonstrated the benefits of using WEKA data mining tools to extract useful information and classify sentiment in tweets related to iPhone and Microsoft, showing that decision tree classifiers outperformed naive Bayes algorithms. Lazard et al. [11] employed text mining to explore important arguments and reactions to e-cigarette regulations on Twitter, uncovering predominantly negative or mixed responses from the public. Twitter media was perceived as one of the channels for accessing credible information about the pandemic [3].

2 Methodology and Data Collection

2.1 The Research Goal Based on the Research Gap

Since Scopus is one of the most extensive scientific sources, it was used to expand more on the literature systematically. The goal of this endeavor was to find articles that relied on Twitter analysis in order to assess the level of systems thinking capabilities based on individuals using NLP, text mining, emotional, and sentiment analysis to distinguish systems thinkers. A comprehensive search thread to get hold of the relevant research articles with relation to these subject areas had

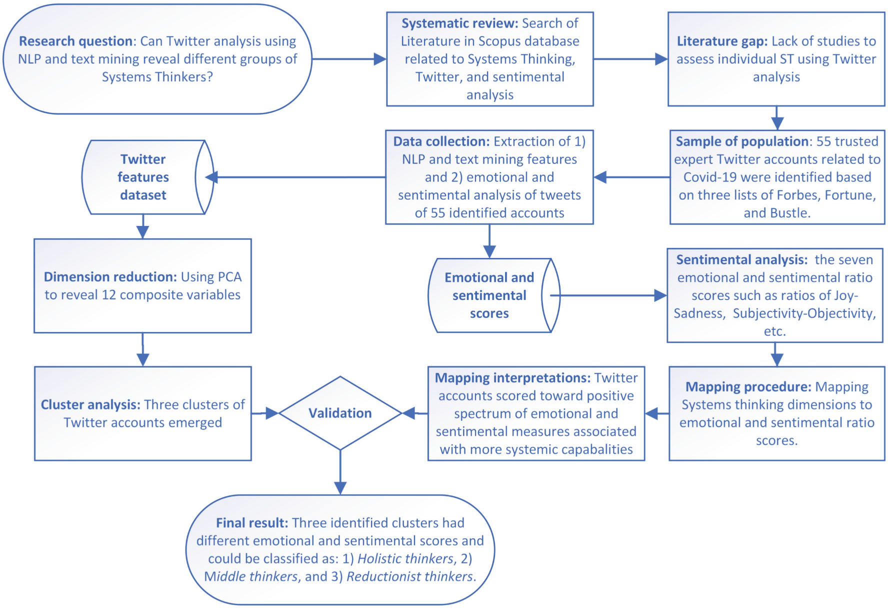

Research question: Can Twitter analysis using NLP and text mining reveal different groups of Systems Thinkers?
Systematic review: Search of Literature in Scopus database related to Systems Thinking, Twitter, and sentimental analysis
Literature gap: Lack of studies to assess individual ST using Twitter analysis
Sample of population: 55 trusted expert Twitter accounts related to Covid-19 were identified based on three lists of Forbes, Fortune, and Bustle.
Data collection: Extraction of 1) NLP and text mining features and 2) emotional and sentimental analysis of tweets of 55 identified accounts
Twitter features dataset
Dimension reduction: Using PCA to reveal 12 composite variables
Emotional and sentimental scores
Sentimental analysis: the seven emotional and sentimental ratio scores such as ratios of Joy-Sadness, Subjectivity-Objectivity, etc.
Mapping procedure: Mapping Systems thinking dimensions to emotional and sentimental ratio scores.
Mapping interpretations: Twitter accounts scored toward positive spectrum of emotional and sentimental measures associated with more systemic capabalities
Validation
Cluster analysis: Three clusters of Twitter accounts emerged
Final result: Three identified clusters had different emotional and sentimental scores and could be classified as: 1) Holistic thinkers, 2) Middle thinkers, and 3) Reductionist thinkers.

Fig. 1. The research framework.

to be achieved. As a result, the extensive review of these articles revealed that the literature lacks studies inspecting the level of systems thinking of people in relation to sentiment analysis and NLP. A critical case study was developed based on the literature gap. A vital and complex problem (COVID-19 pandemic) emerged as a sensitive case to investigate the relationship between the systems thinking capabilities of individuals and their Twitter responses. The relationship between tweets of trusted Twitter accounts and the level of systems thinking skills and sentimental ratio scores in the case of the COVID-19 pandemic was the basis of current research. The research framework is presented in Fig. 1. The procedure of identifying these 55 Twitter accounts and corresponding data collections is introduced in the following section.

2.2 Sample of Population

The COVID-19, initially originating in China in December 2019, rapidly spread globally. The World Health Organization (WHO) declared it a worldwide pandemic due to its severe threat to human health, leading to government-imposed quarantines. “As reported by the BBC, false information ranging from suggested medical care to conspiracy theories have been seen widely around the web, shared, and reposted by thousands of people on social media” [12]; increased the number of casualties and pandemic fear. As a result of these threats, some credible media such as Forbes [13], Fortune [14], and Bustle [12] made up-to-date lists of trusted Twitter accounts of public health officials, researchers, epidemiologists, virus experts, family doctors, among others to assist in spreading right information, news, and recommendation to mass people to enhance society knowledge regarding COVID-19. To classify and identify credible individuals and organizations on Twitter, we have used the three trusted lists. Thus, a list of 55 Twitter accounts consists of 12 organizations and 43 individuals have been chosen for data collection.

2.3 Data Collection Procedure and Twitter Features

To collect the data, we used Twitter's API. Twitter, by default, has a few restrictions regarding the number of tweets, access to timelines, and historical data. We extracted tweets from 55 identified Twitter accounts. First, we extracted potentially important information from each Twitter account, such as the account name, ID, screen name, location, description, follower count, friends count, listed count, favorites count, status count, and other relevant data.

2.4 Text Mining and NLP Features Engineering and Extraction

Opposite to tabular data, textual data is unstructured and to build the machine learning model on top of them, it needs to convert the textual data to numeric format. There are plenty of methods are available to help us from simple vectorization methods including frequency-based methods (TF-IDF, LDA, ...), simple word-embedding methods (word2vec, skipgram, GloVe, ...) to advanced contextual-embedding methods such as the pre-trained transformed-based language models (BERT, RoBERTa, XLNet, ...). Each above mentioned methods has unique characteristics, and it enable us to have better vector representation of text. For example the TF-IDF or LDA methods can cover global statistics over the frequency of word distribution. Word2vec can learn the local relationship between a word and surrounding words with limited window-size. The BERT enables to learn complex relationship between the words from syntactic to semantic ones. Although the simple vectorization and embedding methods are fast, they does not take into account the semantic meaning or context of the words. For example the word "Free" in "I am free to go" and "this product is free" has different meaning regarding different context. So, we utilized some of above mentioned methods to get a more informative features and better feature engineering. This research is designed to explore the possibility of analyzing only textual data from Twitter users to clustering the systems thinking of them. For this purpose, we mainly focused on the textual data rather than other features which are provided by Twitter. For the sake of this research, we needed to explore the tweets of each user to extract useful features that may help us to cluster them. We assumed that NLP features could provide enough information to map the systems thinking dimensions to textual features. We captured different features such as:

A. Average similarity measure, B. Hashtag related features, C. Twitter-length features, D. Topic modeling related features, E. Top "N-gram" word distribution,

F. Word and Sub-word named entity recognition features, G. Ten emotional related features, H. Four sentimental related features, I. Other features.

3 Results and Discussion

3.1 Twitter Feature Engineering and Extraction

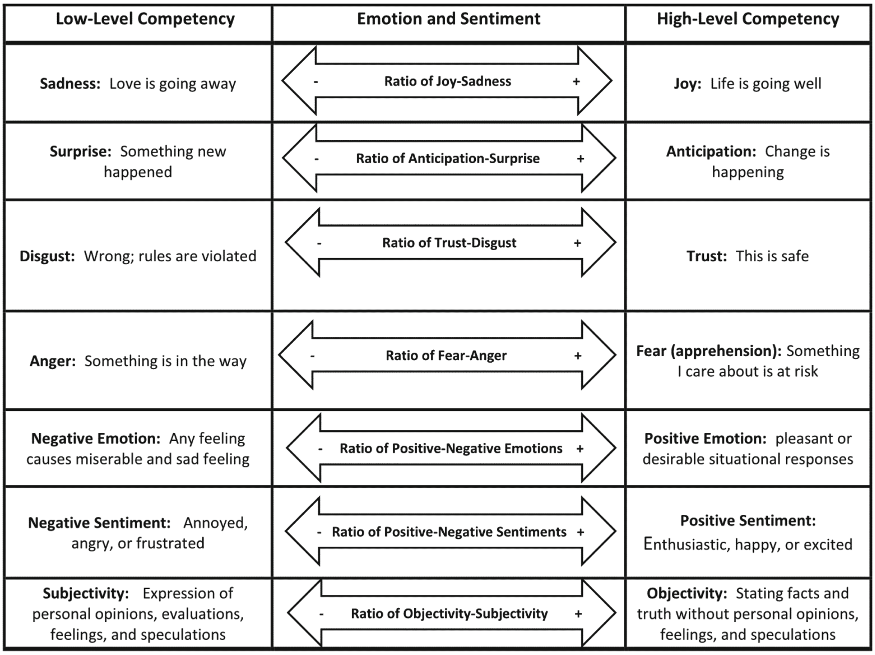
We constructed 60 features related to text mining and NLP techniques extracted from 55 twitter accounts of COVID-19 experts. The first eight features belong to the average similarity measure between systems thinking dimensions and users' tweets for each of the seven dimensions of systems thinking in addition to all dimensions together. The next ten features were extracted based on top-10 hashtags embedded by GloVe for each user. These ten top hashtags include COVID19, coronavirus, Ebola, DRC, 2019nCoV, covid19, flu, HIV, Covid19, and SARSCoV2. The next nine features are relevant to eight Twitter-length features such as token-count, word-count, sentence-count, avg-word-length, avg-token-length, avg-sentence-length, and avg-user-mentions, and one feature for Covid-19 synonym frequency for each user's tweets. The next twelve features are 12 topics that emerged by applying topic modeling of 55 users' tweets-these twelve topics were found with a cohesion score of 0.49 and a perplexity score of -22.25. The next eight features contain top words in tweets with "N-grams." The emerged words found in targeted 55 COVID-19 experts' tweets are "public health," "health care," "global health," "social distancing," "novel coronavirus," "confirmed cases," "vaccine," and "health care workers." Finally, we used two different pre-trained BERT models with word-level tokenizer and sub-word tokenizer. Each model's output was consisting six entity-labels such as I-organization, I-location, I-Miscellaneous (e.g., events, nationalities, products, or works of art), I-person, B-location, B-Miscellaneous entities (B- denotes the beginning and I- inside of an entity). In the end, 12 NER features for each user account were produced (I-ORG-word, I-LOC-word, I-MISC-word, I-PER-word, and B-LOC-word, B-MISC-word, I-ORG-word, I-LOC-word, I-MISC-word, I-PER-word, and B-LOC-word). After extracting the NLP and text mining features from 55 COVID-19 experts' tweets, a 60-features dataset was analyzed. Principle Component Analysis (PCA) with Varimax rotation was performed for the extracted features' dataset. Out of 60 extracted features, twelve composite variables emerged that cumulatively explain about 75 percent of the variability in the data.

Then, we performed K-means clustering to group Twitter users with similar tweets features. The intent was to distinguish COVID-19 experts' tweets based on the extracted NLP and text mining features introduced above.The "SK-Learn" package with the default setting in Python is utilized to perform the clustering task. Grid search cross-validation is employed to tune the hyper-parameters of K-means clustering algorithms with 10-fold cross-validation sorting by the Silhouette score for validation of consistency within clusters of data. Then the best model is utilized to group Twitter users. K-means result shows three distinct clusters emerge. The first cluster consisted of 19 Twitter accounts of COVID-19 experts; the second cluster contained 30 Twitter accounts; the third cluster had six Twitter accounts. The center distance of K-means clustering between the first and second clusters was 2.723. The center distance between the second and third clusters was 1.824. The center distance between the first and third clusters was 2.744. In Sect. 3.3. Cluster analysis and validation, the validation and interpretation of these three clusters of COVID-19 experts based on the emotional and sentimental analysis and corresponding mapping will be presented.

3.2 Sentimental Analysis and Systems Thinking Mapping

We utilized a well-established tool to evaluate the Systems Thinking (ST) proficiency of COVID-19 experts. This tool, created by Jaradat [15], was crafted within the framework of systems science and systems theory, particularly in the domain of complex systems. The ST skills instrument encompasses seven distinct aspects, gauging inclinations towards seven systemic competencies essential for effective engagement with intricate systemic issues. We correlated these seven ST dimensions with emotional and sentiment metrics derived from the analysis of tweets posted by COVID-19 experts. This examination aimed to identify potential associations between emotional and sentiment scores and an individual's level of ST skills.

Ten emotional features extracted using R version 4.0.0. and "syuzhet" library, including Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, Trust, Negative, Positive. Ten scores were assessed based on analyzing the entire tweets of a user. Additionally, we used a pre-trained BERT in Python for extracting two positive and negative sentiment scores for each user and then average them for each user. We also use "textblob" library in Python to extract two objectivity and subjectivity scores of each user's tweets. A review of the literature on sentimental and emotional analysis revealed that the eight primary emotions can be described and interpreted using Plutchik's Wheel of Emotions [16]. Plutchik explored the fundamental origins of these human emotions, as well as their cognition, interaction, and causal relationships. Additionally, he proposed a revolutionary framework linking emotions with the concept of functional thinking. This framework suggests a potential connection between functional and systemic thinking, which represents the mental capacity for problem-solving. Thus, we hypothesize there is a potential relationship between emotional and sentimental measures and the level of systemic thinking for individuals. As a result, we mapped the seven dimensions of the ST skills instrument to emotional and sentimental scores. To make the mapping and interpretation of emotional and sentimental scores to systems thinking dimensions easier, we constructed seven emotional and sentimental ratio scores. Initially, we analyzed all the tweets of each user and calculated 14 emotional and sentimental scores. Then, we scaled 14 emotional and sentimental scores between zero to one. Since each emotion and sentiment has a polar opposite (e.g., joy versus sadness, anticipation versus surprise, positive versus negative, etc.), we constructed seven pairs of emotions



Low-Level Competency	Emotion and Sentiment	High-Level Competency
Sadness: Love is going away	- Ratio of Joy-Sadness +	**Joy:** Life is going well
Surprise: Something new happened	- Ratio of Anticipation-Surprise +	**Anticipation:** Change is happening
Disgust: Wrong; rules are violated	- Ratio of Trust-Disgust +	**Trust:** This is safe
Anger: Something is in the way	- Ratio of Fear-Anger +	**Fear (apprehension):** Something I care about is at risk
Negative Emotion: Any feeling causes miserable and sad feeling	- Ratio of Positive-Negative Emotions +	**Positive Emotion:** pleasant or desirable situational responses
Negative Sentiment: Annoyed, angry, or frustrated	- Ratio of Positive-Negative Sentiments +	**Positive Sentiment:** Enthusiastic, happy, or excited
Subjectivity: Expression of personal opinions, evaluations, feelings, and speculations	- Ratio of Objectivity-Subjectivity +	**Objectivity:** Stating facts and truth without personal opinions, feelings, and speculations

Fig. 2. Two Extremes of Each Emotion and Sentiment.

and sentiments, including ratios of Joy-Sadness, Anticipation-Surprise, Trust-Disgust, Fear-Anger, Positive-Negative emotions, Positive-Negative sentiments, and Objectivity-Subjectivity. For calculating the ratio score of each emotion and sentiment, we subtract the positive scale from the negative scale for each pair (for example, the ratio of Joy-Sadness score is obtained by subtracting the Joy scaled score from the Sadness scaled score).

Since both the ST skills instrument and emotional/sentimental measures utilize pair scales with two extremes (e.g., Level of Complexity in Fig. 2 or Ratio of Joy-Sadness in Fig. 2), we can establish the directional mapping between systems thinking and emotional/sentimental measures. For example, a person who scores toward more systemic extreme in some dimensions of ST skills might score toward positive extremes of emotional/sentimental measures. To connect various degrees of systems thinking skills with emotional and sentimental indicators, we initiated the process by examining each of the seven dimensions of systems thinking. Subsequently, we formulated hypotheses regarding the correlation between each particular systems thinking dimension and every emotional or sentimental metric. These hypotheses were established in accordance with the operational definitions of the systems thinking tool and the interpretations of emotional and sentimental measurements.

According to the ST instrument, if a person scores high in the Level of Complexity dimension, they are inclined to anticipate uncertainty, tend to tackle multidimensional problems, prefer practical solutions, and have a proclivity for exploring their environment. We hypothesize that such individuals are more likely to score toward the positive end of the spectrum in the ratios of Joy-Sadness, Anticipation-Surprise, Positive-Negative emotions, Positive-Negative sentiments, and Objectivity-Subjectivity. In simpler terms, these individuals tend to ignite creativity, foster connections, and generate positive energy (leaning toward Joy), exhibit anticipation for future developments (leaning toward Anticipation), favor favorable situational responses (leaning toward Positive Emotion), lean toward enthusiastic reactions (leaning toward Positive Sentiment), and emphasize facts and truths while avoiding personal opinions, feelings, and speculations [16].

If an individual scores high in the Level of Integration dimension, they are inclined to prioritize global integration, lean toward interdependent decision-making, and aim for global effectiveness. Consequently, we hypothesize that such individuals may lean towards the positive ends of the ratios for $Trust - Disgust$, $Positive - Negative$ emotions, $Positive - Negative$ sentiments, and $Objectivity - Subjectivity$. In other words, they are likely to prefer openness and connection (leaning toward Trust), favor joyful situational responses (leaning toward Positive Emotion), prefer enthusiastic reactions (leaning toward Positive Sentiment), and focus on facts and truths while minimizing personal feelings [16].

If an individual scores more toward the Level of Interconnectivity, he/she is inclined to global interactions, follows a general plan, works within a team, and is interested less in identifiable cause-effect relationships. This person could tend toward the positive extremes of the ratios of Joy-Sadness, Trust-Disgust, Positive-Negative emotions, Positive-Negative sentiments, and Objectivity-Subjectivity, i.e., this person has a tendency to creativity, connection, and energy (more toward Joy extreme), inclined to buildsalliance (more toward Trust extreme), favors happy situational responses (more toward Positive Emotion), prefer exciting responses (more toward Positive Sentiment) and concentrate on facts and truths without involving personal feelings [16].

If a person scores high in the Level of Change, he/she is inclined to taking multiple perspectives into consideration, underspecifies requirements, focuses more on external forces, likes long-range plans, keeps options open, and works best in a changing environment. This person might score more toward positive spectrums of the ratios of Anticipation-Surprise, Fear-Anger, Positive-Negative emotions, Positive-Negative sentiments, and Objectivity-Subjectivity. This individual might prefers looking ahead (more toward Anticipation extreme), protects what he/she cares about (more toward Fear/Apprehension extreme), has positive emotion and sentiment, and has an inclination toward being objective [16].

If a character is more inclined toward the Emergence extreme (high Level of Uncertainty), he/she reacts to situations as they occur, focuses on the whole, is comfortable with uncertainty, believes the work environment is difficult to control, enjoys non-technical problems. Thus, he/she potentially scores more toward Joy, Anticipation, Positive emotions, Positive sentiments, and Subjectivity extremes.

If an individual has a high Level of Systems Worldview (Hierarchical View), he/she prefer to focus on the whole, interested more in the big picture, interested in concepts and abstract meaning of ideas. Consequently, this person might Joy, Anticipation, Trust, Positive emotions, Positive sentiments, and Subjectivity extremes.

If he/she is more inclined toward the Flexibility extreme, he/she more likely to accommodate to change, likes a flexible plan, opens to new ideas, and is unmotivated by routine. Therefore, he/she potentially tends to be more toward Joy, Anticipation, Positive emotions, Positive sentiments, and Subjectivity extremes.

3.3 Cluster Analysis and Validation

The NLP and text mining features dataset has been used to perform the K-means clustering analysis for 55 COVID-19 experts' Twitter accounts. The intent was to group these 55 Twitter accounts based on the tweets' features collected from their Twitter accounts. Since clustering is an unsupervised learning method that needs validation, the emotional and sentimental analysis' results are used to validate the K-means clustering analysis. The 55 COVID-19 experts' Twitter account group in three clusters with different sizes (n_1=19, n_2=30, and n_3=6).

Interestingly, Twitter accounts in cluster 1 have a relatively more positive spectrum of emotional scores, including Joy (versus sadness), Anticipation (versus surprise), trust (versus disgust), and positive emotion (versus negative emotion) than the other two clusters. Additionally, Twitter accounts in cluster 1 have relatively less negative sentiment, less subjectivity (versus objectivity), and less fear (versus anger) than those in the other two clusters. According to mapping the relationships between ST dimensions and emotional and sentimental ratio scores of Twitter accounts, we inferred high scores in emotions and sentiments (except fear-anger scale) of Twitter users associated with high ST capability of individuals. As a result, we concluded since Twitter accounts in cluster 1 have relatively higher emotional and sentimental ratio scores, the individuals corresponding to these Twitter accounts might have higher ST capability skills than others. Consequently, we called Twitter users in cluster 1 holistic thinker clusters. Moreover, the Twitter accounts in cluster 3 called reductionist thinker Twitter accounts due to low scores (more negative score) pertaining to emotional and sentimental ratio scores of their tweets. Since Twitter accounts in cluster 2 have emotional and sentimental ratio scores almost between clusters 1 and 3, they called middle thinker Twitter users. In summary, the overarching outcome of this study suggests that, in the context of COVID-19 pandemic experts, the use of NLP and text mining functionalities, along with emotional and sentimental analysis, could potentially group individuals based on their systemic thinking abilities as reflected in their tweets.

3.4 Randomization Experiment

To assess the influence of having a small cluster size, specifically Cluster 3 comprising six users, on the multiple group T-test analysis to detect significant differences in emotional and sentimental scores among the three clusters, we devised a randomized experiment. We re-clustered the user for 100 trials. Random built-in library in Python utilized to randomly re-assign labels to data with same proportions that observed labels were such as cluster one with 19 users, cluster two with 30 users, and cluster three with six users. Then we performed the ANOVA test to measure the between-group significant differences for emotional and sentimental variables for each of 100 trials. The intent was to make sure cluster size does not impact our main result of the study, which is the meaningful emotional and sentimental score differences between three original identified clusters of COVID-19 experts.

4 Conclusion

Twitter plays a pivotal role in disseminating information, knowledge, and news. Unfortunately, it also serves as a platform where erroneous or insufficient messages can spread within social communities. For instance, during the global COVID-19 pandemic, we witnessed numerous false and unsupported claims regarding the virus's origin, transmission, potential prevention methods, and treatments. In the midst of this confusion, having access to reliable sources with a holistic approach becomes essential for the public to make informed and systematic decisions. Hence, COVID-19 emerged as a compelling use case to explore the correlation between tweets from COVID-19 experts and their levels of systems thinking capabilities. Given the interdisciplinary nature of our research, we conducted a systematic review, encompassing a comprehensive global scientific source like Scopus databases, to ensure we encompassed all relevant literature in Systems Thinking (ST), pertinent Twitter Analysis, and NLP/text mining.

After reviewing the articles, we found that there is a literature gap connecting experts' systems thinking level with the way they tweets. Therefore, we identified 55 trusted expert Twitter accounts based on the published lists of Forbes, Fortune, and Bustle to investigate their tweets with their ST capabilities. We constructed the tweets of these experts based on 60 NLP and text mining features extracted from their Twitter accounts, including the average similarity measures between systems thinking dimensions and users' tweets, top-10 hashtags embedded by GloVe, top-10 hashtags by different COVID terminologies (coronavirus, SARSCoV2, etc.), Twitter-length features (token-count, word-count, avg-sentence-length, etc.), 12 emerged topics by applying topic modeling, top words distribution in tweets with "N-grams," and word and sub-word named entity recognition features. PCA as a method of dimension reduction unfolded 12 composite variables explaining more than 75 percent of the variance in the initial features. After clustering the 12 composite variables related to tweets' features, we found that three distinct groups of experts emerged.

For validating the result, we went further and analyzed the emotional and sentimental scores of experts. To make the interpretation easier and also to connect to the origin of emotional and sentimental analysis ("Plutchik's Wheel of Emotions"), we have generated ratio scores for emotional and sentimental measures for each user based on their tweets. Then, we mapped the systems thinking dimensions to the seven emotional and sentimental ratio scores, such as ratios of Joy-Sadness, Anticipation-Surprise, Trust-Disgust, Fear-Anger, Positive-Negative emotions, Positive-Negative sentiments, and Objectivity-Subjectivity. By comparing the emotional and sentimental characteristics of 55 experts, we found that three identified clusters had meaningful differences in the level of emotions and sentiments.

Moreover, the first cluster with scores more toward the positive spectrum of emotions and sentiments (e.g., trust, anticipation, objectivity, positive, etc.) associated with higher systems thinking capability, which can be categorized as a holistic thinker. On the other hand, the third cluster with scores more toward the negative spectrum of emotions and sentiments (e.g., anger, sadness, subjectivity, negative, etc.) can be classified as reductionist thinkers (with less systemic capabilities). Finally, the second cluster had in the middle range level of emotions and sentiments, labeled middle thinkers. Briefly, this research was testing that the capabilities of individuals as systems thinkers can lead to revealing unique tweet-writing patterns and distinct emotional and sentimental scores associated with the level of systems thinking of COVID-19 experts.

Limitations of the study mainly were related to Twitter access since Twitter has restrictions related to the search of past tweets. Another limitation of the study is the validity of mapping systems thinking dimensions to emotional and sentimental measures of the selected Twitter users; further investigation with other samples of the population is needed to validate the current mapping. Future studies will be directed into investigating the relationship between Twitter users' systems thinking and their tweets for other samples of the population, such as politicians, artists, and other samples. Comparing the followers' network analysis of COVID-19 expert Twitter accounts with their tweets to better understand the impact of tweets on followers and also spread of knowledge and information within the Twitter network. The new research finding can be compared and combined with the current research results to analyze the validity and consistency between these studies.

The results of the current study have some practical implications. Using NLP and text mining analysis along with emotional and sentimental analysis helps to comprehend: systems thinking capability of influencers and celebrities, and their role in spreading the true news and knowledge to the community; how systems thinking is related to have a better expression of specific emotions in social media, which promotes the more efficient and effective transformation of information and knowledge to the community.

As the level of systems thinking skills of individuals can be enhanced, the social media activity of the individuals can be improved. The research shows there is a necessity to create a safe and healthy virtual environment for everybody, so everyone can express their opinions and beliefs in the direction of the community's values without disrespecting and undermining other' opinions and beliefs.

All supplementary files are reachable through:
https://drive.google.com/file/d/156Yz4kP7Bm7XYhGKMB6Aa9-4Rkxo-YyG/view?usp=sharing.

References

1. Aljedaani, W., et al.: Sentiment analysis on twitter data integrating textblob and deep learning models: the case of us airline industry. Knowl.-Based Syst. **255**, 109780 (2022)
2. Sunitha, D., Patra, R.K., Babu, N., Suresh, A., Gupta, S.C.: Twitter sentiment analysis using ensemble based deep learning model towards covid-19 in India and European countries. Pattern Recogn. Lett. **158**, 164–170 (2022)
3. Qorib, M., Oladunni, T., Denis, M., Ososanya, E., Cotae, P.: Covid-19 vaccine hesitancy: text mining, sentiment analysis and machine learning on covid-19 vaccination twitter dataset. Expert Syst. Appl. **212**, 118715 (2023)
4. Kharde, V., Sonawane, P., et al.: Sentiment analysis of twitter data: a survey of techniques. arXiv arXiv:1601.06971 (2016)
5. Sailunaz, K., Alhajj, R.: Emotion and sentiment analysis from twitter text. J. Comput. Sci. **36**, 101003 (2019)
6. Weber, C.T., Syed, S.: Interdisciplinary optimism? sentiment analysis of twitter data. Royal Soc. Open Sci. **6**(7), 190473 (2019)
7. Plunz, R.A., et al.: Twitter sentiment in New York city parks as measure of well-being. Landsc. Urban Plan. **189**, 235–246 (2019)
8. Schumaker, R.P., Jarmoszko, A.T., Labedz, C.S., Jr.: Predicting wins and spread in the premier league using a sentiment analysis of twitter. Decis. Supp. Syst. **88**, 76–84 (2016)
9. Arora, M., Kansal, V.: Character level embedding with deep convolutional neural network for text normalization of unstructured data for twitter sentiment analysis. Soc. Netw. Anal. Min. **9**(1), 1–14 (2019)
10. Wakade, S., Shekar, C., Liszka, K.J., Chan, C.-C.: Text mining for sentiment analysis of twitter data. In: Proceedings of the International Conference on Information and Knowledge Engineering (IKE). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing, p. 1 (2012)
11. Lazard, A.J., Wilcox, G.B., Tuttle, H.M., Glowacki, E.M., Pikowski, J.: Public reactions to e-cigarette regulations on twitter: a text mining analysis. Tob. Control **26**(e2), e112–e116 (2017)
12. Wylde, K.: Expert twitter accounts for coronavirus & covid-19 updates (2020). https://www.bustle.com/p/expert-twitter-accounts-for-coronavirus-covid-19-updates-22419348
13. Brown, A.: Coronavirus: the most essential people to follow on twitter during the covid-19 outbreak (2020). https://www.forbes.com/sites/abrambrown/2020/03/14/coronavirus-the-most-essential-people-on-twitter%CB%86tofollow-during-the-covid-19-outbreak/#1969611375f3
14. Moore, M.: The best twitter accounts to follow for reliable information on the coronavirus outbreak (2020). https://fortune.com/2020/03/14/coronavirus-updates-twitter-accounts-covid-19-news/
15. Jaradat, R.M.: Complex system governance requires systems thinking-how to find systems thinkers. Int. J. Syst. Eng. **6**(1–2), 53–70 (2015)
16. Donaldson, M.: Plutchik's wheel of emotions-2017 update (2017)

ADHD Prediction in Children Through Machine Learning Algorithms

Daniela Andrea Ruiz Lopez[1], Harun Pirim[2](✉), and David Grewell[3]

[1] University of the Andes, Bogotá, Colombia
[2] North Dakota State University, Fargo, ND, USA
harun.pirim@ndsu.edu
[3] Northern Illinois University, DeKalb, IL, USA

Abstract. Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder that affects approximately 5% of children worldwide. It is typically diagnosed based on the presence of inattentive and hyperactive symptoms. Our objective is to identify ADHD from a Machine Learning (ML) perspective, utilizing symptom information and features such as socioeconomic status, social behavior, academic competence, and quality of life. We conducted extensive experiments using the CAP dataset and various machine learning algorithms, including logistic regression, k-nearest neighbors, Support Vector Machines (SVMs), Random Forest, XGBoost, and an Artificial Neural Network (ANN). The ANN model demonstrated the highest accuracy, achieving an AUC metric of 0.99. As a result, we conclude that using ML algorithms to predict ADHD provides a better understanding of the etiological factors associated with the disorder and has the potential to form the basis for a more precise diagnostic approach. The code is available at: GitHub Repository.

Keywords: ADHD · Machine Learning · Logistic Regression · kNN · SVMs · Random Forest · Xgboost and Artificial Neural Network

1 Introduction

Attention-deficit/hyperactivity disorder (ADHD) is a common neurodevelopmental disorder with behavior of high distractibility and impaired executive functions [1]. The global prevalence is around 5%, but at least a further 5% of children have substantial difficulties that placed them just under the threshold to meet full diagnostic criteria for ADHD [1].

Capturing and analyzing electroencephalogram (EEG) and Magnetic Resonance Imaging (MRI) data to diagnose ADHD have increased [2]. However, it is usually diagnosed by the behavioral symptoms exposed on the Diagnostic Statistical Manual criteria [3]. It considers 18 ADHD symptoms divided on 9 Inattentive and 9 Hyperactive/Impulsive. According to them, ADHD can be diagnosed categorically (ADHD hyperactive/impulsive, ADHD inattentive, ADHD combined or non-ADHD) or dimensional, based on the total symptom count [4].

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 89–100, 2024.
https://doi.org/10.1007/978-3-031-56728-5_8

On the other hand, due to ADHD heterogeneity between individuals, computer-aided methods of diagnosing have emerged [2]. These approaches are based on supervised machine learning (ML) algorithms for classification, which learn how to map an input to its output from experience (example input-output pairs) [5]. Among the most used models stand out the logistic regression [6], k-Nearest-Neighbors (kNN) [7], Random decision forests [8], Support Vector Machine (SVM) [9] and Artificial Neural Network (ANN) [10].

To advance in the ADHD diagnosis through ML techniques, several studies have been developed to collect phenotypic data such as gender, age, IQ scores, diagnostic status, and medication status, others even include medical images such as EEG and MRIs from ADHD patients and non-ADHD subjects. The Children's Attention Project (CAP) [11] is one of the studies focused on the phenotypic data. It studied the mental health, social, academic and quality of life outcomes along 3 years for children between 6 and 8 years old with diagnostically-confirmed ADHD in contrast to non-ADHD controls [11].

According to the above, our aim is to distinguish ADHD and control groups using the relevant features reported in the CAP dataset [11] and the inter-connections between them through machine learning perspective to better understanding etiological factors associated with the disorder, increasing accuracy in predicting outcomes.

2 Methodology

2.1 Dataset Description

The CAP's dataset recorded symptoms and social, economic and academic information for 146 medication naive children with ADHD and 209 controls through parent interviews and teacher reports [12]. The ground-truth labels for CAP's dataset were obtained through a rigorous diagnostic process. First, any child reported by parents to have previously been diagnosed with ADHD was regarded as a positive screen; if not, there was an initial screening phase using the Conners' 3 ADHD index. Then, a face-to-face structured diagnostic interview was conducted using the NIMH Diagnostic Interview Schedule for Children IV (DISC-IV) to confirm the diagnostic status. Trained staff members with psychology degrees interviewed the parents of the participating children. Based on these results, the children were assigned ground-truth labels of either ADHD or control, forming the basis for classifying participants in the study [11].

The ADHD group comprised 61 Predominantly Inattentive (ADHD-I), 15 Predominantly Hyperactive/Impulsive (ADHD-H) and 70 Combined types (ADHD-CT). Children were classified as having an internalizing disorder if they met criteria for separation anxiety disorder, social phobia, generalized anxiety disorder, post-traumatic stress disorder, obsessive compulsive disorder, hypomania or manic episode, and an externalizing disorder if they met criteria for oppositional defiant disorder or conduct disorder [12].

Dataset is compound of 42 factors (columns) and 355 children (rows). Between the factors is the group to which each child belongs (0 if Control, 1 if ADHD); gender (0 if female, 1 if male), age, Socio-Economic Indexes for Areas (SEIFA), the code of the child, the presence of each of the 18 symptoms, the count of all of them, the count of the hyperactive symptoms, the count of the inattentive ones and the presence of externalizing and internalizing disorder. There is also information about the Clinical Evaluation of

Language Fundamentals (CELF), the Wide Range Achievement Test (WRAT) for math and reading, the social problems measure and the academic competence in the first and third age (2011 and 2014) of the study for each factor. Moreover, the sleep problems, the irritability and the Quality of Life in terms of emotion, family and time were also considered inside the dataset.

2.2 Dataset Preprocessing

For each of the factors mentioned previously, we estimated descriptive data corresponding to the Mean, Typical Error, Median, Mode, Standard Deviation, Sample Variance, Kurtosis, Skewness Coefficient, Range, Minimum, Maximum, Sum, and Count. Some of them are reported in the Table 2.

According to the statistics, most of the children interviewed are males, but the proportion between the groups is very similar (68%-32% in ADHD group and 63%-37% in control one). The mean age was 7.3 years old in both groups. Median and mode were also close to this value. SEIFA mean in children with ADHD is very close to that of the control group, but the median and mode are higher for control children. Children with ADHD have 13 symptoms on average (the inattentive symptoms are the most prevalent), while children in the control group just have 1 (in most cases, a hyperactive symptom). The externalizing disorder is more frequent than internalizing in both groups. It is present in 50% of the children with ADHD and 7% of the control children. Although CELF shows the same performance in the groups, WRATs about maths and reading, academic competence, and all the indicators of quality of life have higher statistics for the control group. On the contrary, irritability and social and sleep problems are more frequent in the ADHD group.

Since some of the data was missing, we made a data recovery using the mode for the categorical variables and an interpolation for continuous ones. Moreover, we considered a factor called *Desertion* that takes a value of 1 if a child had no values reported for data collected in 2014 or 0 if it was complete. A higher proportion of children in the control group dropped out of the study but considering that the difference in Desertion between groups is only 6%, it was not regarded as biased.

On the other hand, as feature values belong to different ranges, data normalization was implemented in both training and validation set. This data preprocessing technique is frequently used in machine learning to use the same scale for all the numerical values in the dataset, which avoids distorting the differences in the ranges of values and losing information. Furthermore, many commonly used algorithms require normalization to model the data correctly.

Correlations between factors are shown in Fig. 1. We noticed that age was highly correlated only to the CELF language baseline and the same exam after 3 years. SEIFA has low correlations with all factors. The symptoms have a high positive correlation between them and the social problems and irritability indicators. On the contrary, symptoms indicate a high negative correlation with the quality of life, such as with most academic indicators (maths, reading, and academic competence). In this way, the presence of ADHD symptoms seems to be related to lower grades, quality of life, and more behavioral problems.

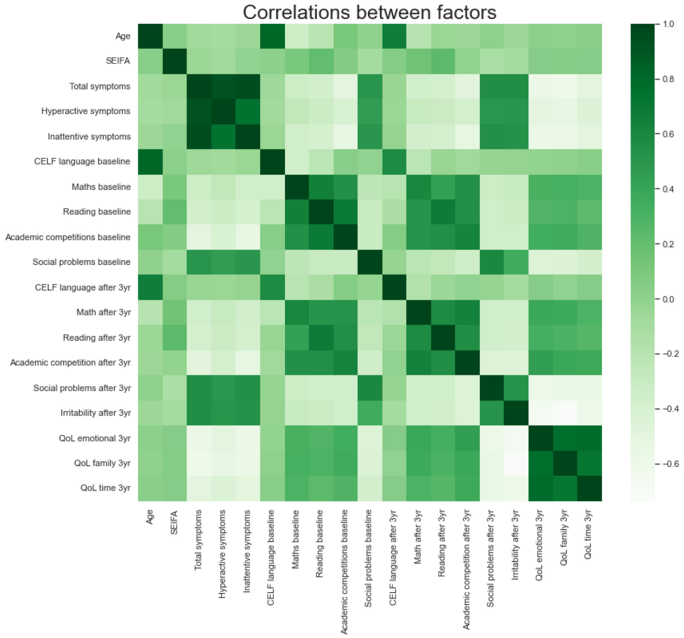


Fig. 1. Pearson correlation between numerical variables. The sign of the correlations indicates the proportionality (direct and inverse), so they are highlighted on a color scale ranging from light colors (highest negative correlations) to dark colors (highest positive correlations).

2.3 Evaluation Methodology

We initially divided the data into Train (70%), and test (30%) sets for evaluation. However, to avoid bias induced because of the data split, we used the 5-fold cross-validation method to evaluate. The metrics we worked with were accuracy (2.1), precision (2.2), recall (2.3) and Area Under the Curve (AUC) of the Receiver Operator Characteristic (ROC). The ROC curve plots the Recall (also called True Positive Rate (TPR)) vs. the False Positive Rate (FPR) (2.4).

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{2.1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2.2}$$

$$Recall/TPR = \frac{TP}{TP + FN} \tag{2.3}$$

$$FRP = \frac{TN}{FP + TN} \tag{2.4}$$

True Positives (TP) refer to the correct ADHD predictions, and True Negatives (TN) to the correct control predictions. In contrast, False Positives (FP) refer to the wrong

ADHD predictions, and False Negatives (FN) refer to the wrong control predictions. Then, Accuracy is the fraction of correct predictions to total predictions (either positive or negative class) [13]. Precision is the fraction of true positives to total predicted positives (Penalizes false positives) [13]. Recall or TPR is the fraction of true positives to total positives in the data (Penalizes false negatives) [13]. FPR is the fraction of true negatives to total negatives in the data [13]. Finally, AUC measures the ability of a classifier to distinguish between classes [13]. The higher the metrics, the better the model's performance.

2.4 Experiments

Logistic Regression. Logistic regression is similar to linear regression but with one critical addition. It analyses the independent variables to estimate a binary dependent variable instead of a continuous one. The logistic function compresses the output of the linear function to a range between 0 and 1. Then, it is an S-shaped function that gets closer to 1 as the input value increases above 0 and closer to 0 as the input value decreases far below 0 [14].

The hyperparameters tuned were the type of penalty for having too many variables. We tried with *None*, *L1*, and *L2* configurations. We also vary the regularization amount (C) to avoid overfitting between 0.1, 1, and 100. Finally, we changed the class weights from "None" to "balanced" to adjust weights proportionally to class frequencies.

K Nearest Neighbors (kNN). The kNN classifier finds the k-Nearest (k-most similar) instances in the training set to the query point and gets the labels of those training instances. Finally, the algorithm uses the mode of the nearby training labels to predict the label of the new instance [15].

The most important hyperparameter here is the number of neighbors to consider for predicting the class of the query point (K). Likewise, the weight function is crucial; if *uniform* is used, all points in each neighborhood are weighted equally, but if *distance*, points are weighted by the inverse of their distance. We tuned k with values between 1 to 15 for both weight functions.

Linear Support Vector Machine. Support Vector Machine is a linear model for classification and regression problems that creates a hyperplane to separate the data into classes. The closest points to the hyperplane from both classes are called support vectors, and the distance between the line and the support vectors is the margin. The goal is to maximize the margin to find the optimal hyperplane [9].

For this algorithm, we change the penalty (*L1* or *L2*) and the loss function (*Hinge* or *Squared hinge*) to measure how good are the train points fitted to the model. We also tried different values of the regularization hyperparameter C (0.1, 1, and 100) and tuned the class weights from unbalanced to balanced.

Kernelized Support Vector Machine. Linear support vector machines work well for simple classification problems, where the classes are linearly separable. But in many classification problems, the different classes are located in a way that a line or hyperplane cannot act as an effective classifier. Then, kernelized support vector machines can provide more complex models.

Here, the most important is to specify the kernel type used in the algorithm. The Radial Basis Function kernel (RBF) uses the squared Euclidean distance, the polynomial kernel represents the similarity of vectors in a feature space over polynomials, and the sigmoid kernel uses the sigmoid function. We set C and the class weight the same as in the linear version.

Decision Trees. Decision trees are easy to use and understand and are often an excellent exploratory method for getting a better idea of the influential features in a dataset. In a few words, decision trees learn a series of explicit if-then rules on feature values that result in a decision that predicts the target value.

The hyperparameters we changed are the function to measure the quality of a split (*giny*, *entropy* or *log_loss*), the maximum depth of the tree (integer or None to expand nodes until all leaves are pure), and same as in previous methods, class weight (balanced or not).

Random Forest. Random Forest combines several decision trees to improve generalizability. This way, different trees see different portions of the data and combine their results. Some errors are compensated with others, which leads to a prediction that generalizes better.

For this algorithm we set the number of trees in the forest (10, 100, and 1000), the function to measure the quality of a split (*giny*, *entropy* or *log_loss*), and the class weight (balanced or not).

Xgboost. Like Random Forest, Xgboost uses an ensemble of multiple trees to create more powerful prediction models for classification. However, this algorithm does not build and combine a forest in parallel; it builds a series of trees. Then, each tree attempts to correct the mistakes of the previous one in the series [16].

In this model, we considered a new hyperparameter called learning rate, which is the step size in updating the weights during training. We tuned this hyperparameter with values of 0.01, 0.1, and 1. Moreover, we varied the maximum depth of the trees (None, 3, and 10) and the number of estimators (2, 10, 100, and 200 trees).

Artificial Neural Network (ANN). The latest algorithm we tried was an Artificial Neural Network, specifically, a multi-layer perceptron (MLP). It is compound of three or more layers of nodes (input, hidden, and output). Each node is a neuron that employs a nonlinear activation function, except for the nodes in the input layer. It uses Backpropagation method for training.

The hyperparameters we changed are the activation function for the hidden layer (*relu*, *identity*, *logistic*, *tanh*), the solver for weight optimization (*adam*, *lbfgs* and *sgd*), the learning rate schedule for weight updates (*constant, invscaling, and adaptive*), and the initial learning rate used to control the step-size in updating the weights (0.01, 0.001, 0.0001).

3 Results

We tested different hyperparameter configurations for each model to choose the one with the highest AUC. If two or more experiments had the same value for this metric, we selected the one with the highest precision. All the models worked better with the original data normalized than with the original data without normalization. Then, table 1 shows the results of the normalized data set.

Table 1. Final metrics and models comparison

Model	Accuracy		Precision		Recall		AUC	
	Train	*Test*	*Train*	*Test*	*Train*	*Test*	*Train*	*Test*
Logistic Regression	0.987	0.983	0.99	0.986	0.978	0.973	1	0.999
kNN	0.994	0.986	0.995	0.986	0.991	0.98	1	0.999
Linear SVM	0.986	0.969	0.993	0.972	0.973	0.952	1	0.997
Kernelized SVM	0.983	0.977	0.962	0.954	0.998	0.993	1	0.999
Decision Trees	1	0.958	1	0.958	1	0.938	1	0.955
Random Forest	1	0.986	1	0.98	1	0.986	1	0.998
XgBoost	0.992	0.966	0.99	0.978	0.99	0.938	0.99	0.999
Neural Network	0.987	0.975	0.976	0.987	0.988	0.979	0.999	0.999

The best configuration for the Logistic Regression model used the L2 penalty and C with a value of 0.1. For kNN, the best k was 3 according to the k sensitivity test, and the k points were weighted uniformly regardless of their distance to the query point. The Linear Support Vector Machine worked better with L1 Penalty, Squared hinge loss function, and C equal to 0.1. Between the Kernelized SVMs, “RBF” showed a better AUC. As for Decision Trees, the criterion chosen was entropy with the maximum depth necessary to expand the nodes until all leaves are pure. In the best experiment of Random Forest, we used the default number of estimators (100) and the “gini” criterion. To XgBoost, the Learning Rate used was 0.01, the number of estimators was 100, and the maximum depth of the trees was 6. Finally, for the Neural Network, we selected “relu” as the activation function, “sgd” as the solver, and 0.001 for the initial learning rate (we decided it to be constant throughout the training).

On the other hand, we analyzed the features’ importance based on their relevance in the decision trees-based methods (Decision Trees, Random Forest, and XgBoost). Figure 2 shows that the most important symptoms are difficulty sustaining attention (“susatt”), not following instructions (“instruct”), not listening (“listen”), and avoiding mental effort (“avoid”) while being quiet (“quiet”) and often talking excessively (“talks”) had the lowest importance. In general, the inattention category shows higher average importance than the category of hyperactivity. Among the social-economic-academic features, we found the Academic Competence taken at the end of the study is the most relevant. In contrast, sleep problems and language tests were of low importance.

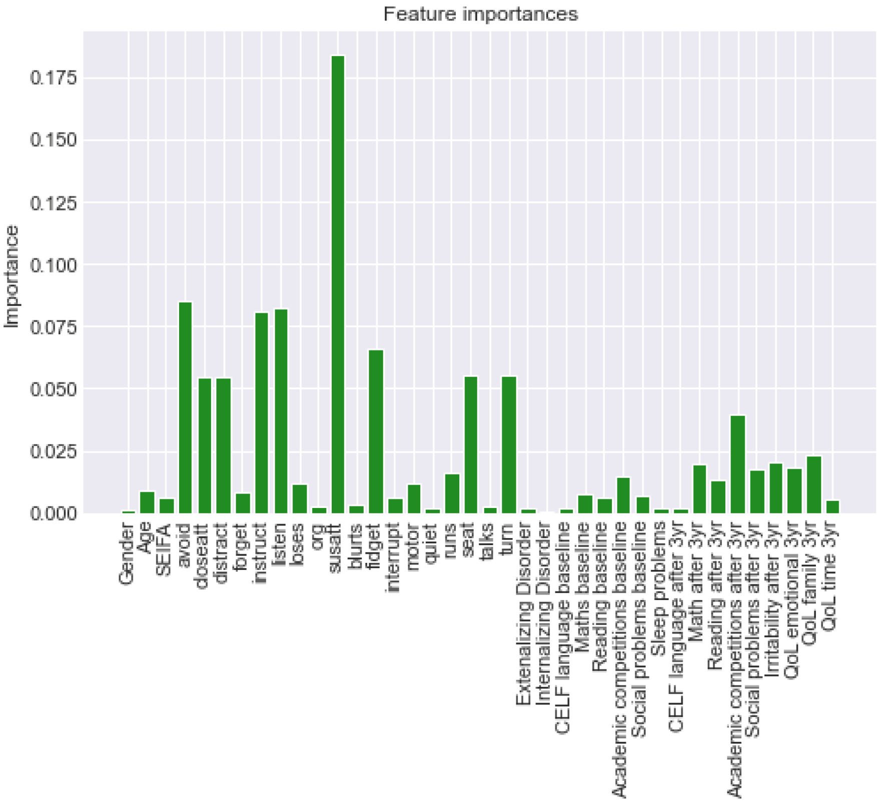


Fig. 2. Features importance in decision trees-based methods.

4 Analysis and Conclusions

As mentioned before, the original data showed worse metrics than after normalization. This shows the importance of scaling the data so that each feature has the same contribution. As this procedure can improve the data quality and the performance of the machine learning algorithms, it should be the first step of the data pre-processing.

On the other hand, it is important to mention why the AUC metric was preferred instead of accuracy. The AUC metric uses probabilities of predictions, which is essential to evaluate imbalanced data and make the evaluation more precise. Moreover, we chose Precision over Recall because it penalizes false positives, the error we want to avoid most because of the consequences of diagnosing a child with ADHD when it is invalid.

The compared models have both advantages and disadvantages. Logistic regression makes no assumptions about distributions of classes in feature space and can interpret model coefficients as indicators of feature importance. However, it assumes linearity between the dependent and independent variables; then, it has a good accuracy just for simple datasets that are linearly separable. kNN is intuitive and straightforward, does not have training steps, and has just one hyperparameter (k). In contrast, as the dataset grows, the algorithm's speed declines; features must have the same scale, and the algorithm is very susceptible to k and outliers. Linear SVM is simple and easy to train, scales well to large datasets, and works well with sparse data. Still, it does not generalize well for lower-dimensional data and assumes that data is linearly separable. Kernelized SVMs are more versatile and work well for low and high-dimensional data, but they need careful normalization of input data and hyperparameter tuning. The Decision Trees algorithm is easy to visualize and interpret, does not need feature normalization, and works well with datasets using a mixture of feature types. However, Decision Trees tend to overfit, and an ensemble of trees is usually needed for better generalization performance. Random Forest solves this problem but requires a higher computational cost and time, just like XgBoost, which is why it is not recommended for problems with high dimensional sparse features. By last, Artificial Neural Networks form the basis of state-of-the-art models. They can be assembled into advanced architectures that effectively capture complex features given enough data and computation. However, more complex models require more training time, data, and customization.

As noticeable in Table 1, all the methods report metrics above 0.9. Indeed, the most straightforward methods, such as logistic regression and kNN, had a performance comparable with the multilayer perceptron. These results may be because the dataset is well-behaved and the classes are well-separated, so the differences in the models' capabilities might be less significant. Also, the dataset does not have complex or non-linear patterns, so a simple model like logistic regression could perform as well as a more complex model like MLP. Then, it would be necessary to try the same model with a larger dataset to confirm the proposed model's efficiency and generalization. This new dataset should contain information about worldwide children with and without ADHD to reach a better abstraction. Also, considering more biological data, such as brain connectivity shown in Magnetic Resonance Images (MRIs), would make the classification model more robust.

Finally, it is crucial to analyze the relevance of features to diagnose ADHD according to Fig. 2. The higher the difference between groups for a feature, the higher its importance. For example, the feature with the highest importance in predicting ADHD is the disability of sustaining attention since few children in the control group present this symptom. Although the most important features belong to the symptoms category, not all are relevant. Even some symptoms have lower relevance than academic and social factors. In general, the socioeconomic-academic aspects measured 3 years later had higher importance than those measured at the beginning of the study. This behavior could indicate that the differences between both groups increase with the years.

Current ADHD diagnosis assumes every symptom has the same importance, then the combination of them is irrelevant, and the diagnosis can be made based only on the count of symptoms. However, mentioned above contradicts that hypothesis and shows the different relevance values of each feature. It also demonstrates that some social-economic and academic information correlates highly with ADHD, which is helpful in diagnosis.

In conclusion, it is possible to predict ADHD through Machine Learning techniques. Furthermore, ML allows a better understanding of the etiological factors associated with the disorder, which increases the precision of diagnosis. We demonstrated that symptoms do not have the same importance and are not the only relevant features to recognize the presence of ADHD. In this way, predicting ADHD with ML could allow us to identify that 5% of children worldwide are just under the threshold to be diagnosed because they do not fulfill the count of symptoms required, but they show the most important ones.

5 Annexes

Table 2. Descriptive data: Mean, median, mode and standard deviation for each factor.

Factors	Mean		Median		Mode		Standard Deviation	
	ADHD	*Control*	*ADHD*	*Control*	*ADHD*	*Control*	*ADHD*	*Control*
Gender	0.68	0.63	1.00	1.00	1.00	1.00	0.47	0.48
Age	7.27	7.33	7.25	7.34	7.44	7.59	0.44	0.38
SEIFA	1014.9	1015.6	1012.0	1022.0	997.0	1067.0	41.10	45.61
Total symptoms	12.65	1.84	13.00	1.00	12.00	0.00	2.97	2.26
Hyperactive symptoms	5.62	0.99	6.00	0.00	6.00	0.00	2.42	1.39
Inattentive symptoms	7.03	0.85	7.00	0.00	8.00	0.00	1.58	1.33
Externalizing disorder	0.51	0.08	1.00	0.00	1.00	0.00	0.50	0.27
Internalizing disorder	0.25	0.05	0.00	0.00	0.00	0.00	0.44	0.21
CELF language baseline	15.49	15.64	16.00	16.00	16.00	16.00	1.20	1.12
Math baseline	91.30	102.83	92.00	101.00	96.00	92.00	14.33	13.48
Reading baseline	98.10	111.93	97.50	111.00	105.0	108.00	17.34	13.59
Academic competitions baseline	86.17	103.56	87.00	105.00	64.00	107.00	14.18	11.87
Social problems baseline	2.99	1.09	3.00	1.00	2.00	0.00	2.17	1.34
Sleep problems	0.24	0.07	0.00	0.00	0.00	0.00	0.43	0.25

(*continued*)

Table 2. (*continued*)

Factors	Mean		Median		Mode		Standard Deviation	
	ADHD	*Control*	*ADHD*	*Control*	*ADHD*	*Control*	*ADHD*	*Control*
CELF language after 3yr	17.97	18.01	18.00	18.00	18.00	18.00	0.60	0.55
Math after 3yr	87.66	99.09	88.00	97.00	88.00	102.00	14.54	13.16
Reading after 3yr	95.81	106.41	94.00	106.00	89.00	106.00	14.36	13.52
Academic competitions after 3yr	88.87	103.54	91.00	104.00	91.00	121.00	13.84	12.26
Social problems 3yr	2.94	0.99	3.00	0.00	1.00	0.00	2.41	1.48
Irritability after 3yr	4.83	1.61	4.00	1.00	3.00	0.00	3.32	2.32
QoL emotional 3yr	54.40	85.34	62.50	87.50	75.00	100.00	30.01	18.24
QoL family 3yr	66.70	90.20	68.33	95.83	100.0	100.00	23.87	15.48
QoL time 3yr	74.14	96.86	83.33	100.00	100.0	100.00	29.61	10.76
Desertion	0.18	0.24	0.00	0.00	0.00	0.00	0.39	0.43

References

1. Sayal, K., Prasad, V., Daley, D., Ford, T., Coghill, D.: ADHD in children and young people: prevalence, care pathways, and service provision. Lancet Psychiat. **5**(2), 175–186 (2018)
2. Sethu, N., Vyas, R.: Overview of machine learning methods in ADHD prediction. In: Vyas, R. (ed.) Advances in Bioengineering, pp. 51–71. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-2063-1_3
3. Association, A.P.: Diagnostic and Statistical Manual of Mental Disorders: DSM-5-TR. G - Reference, Information and Interdisciplinary Subjects Series. American Psychiatric Association Publishing (2022)
4. Mueller, K.L., Tomblin, J.B.: Diagnosis of ADHD and its behavioral, neurologic and genetic roots. Top. Lang. Disord. **32**(3), 207 (2012)
5. Mahesh, B.: Machine learning algorithms-a review. Int. J. Sci. Res. (IJSR) **9**, 381–386 (2020)
6. Wright, R.E.: Logistic regression (1995)
7. Guo, G., Wang, H., Bell, D., Bi, Y., Greer, K.: Knn model-based approach in classification. In: Meersman, R., Tari, Z., Schmidt, D.C. (eds.) OTM 2003. LNCS, vol. 2888, pp. 986–996. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-39964-3_62
8. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)

9. Suthaharan, S.: Support vector machine. In: Machine learning models and algorithms for big data classification. ISIS, vol. 36, pp. 207–235. Springer, Boston, MA (2016). https://doi.org/10.1007/978-1-4899-7641-3_9
10. Zupan, J.: Introduction to artificial neural network (ann) methods: what they are and how to use them. Acta Chim. Slov. **41**, 327 (1994)
11. Sciberras, E., et al.: The children's attention project: a community-based longitudinal study of children with ADHD and non-ADHD controls. BMC Psychiat. **13**(1), 1–11 (2013)
12. Silk, T.J., et al.: A network analysis approach to ADHD symptoms: more than the sum of its parts. PLoS ONE **14**(1), e0211053 (2019)
13. Heeswijk. Precision and recall—a comprehensive guide with practical examples (2022)
14. Sharma, S.: Activation functions in neural networks - towards data science. Medium (2017)
15. Harrison, O.: Machine learning basics with the k-nearest neighbors' algorithm (2018)
16. Chen, T., Guestrin, C.: Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794 (2016)

Commonsense Validation and Explanation for Arabic Sentences

Farah Alshanik(✉), Ibrahim Al-Sharif, and Mohammad W. Abdullah

Department of Computer Science, Jordan University of Science and Technology, Irbid, Jordan
fmalshanik@just.edu.jo, {iaalsharif21,mwabdulah21}@cit.just.edu.jo

Abstract. Commonsense understanding poses a significant challenge, especially in complex languages like Arabic. However, recent advancements in deep learning have facilitated improvements in various language tasks, including the ability to distinguish commonsense in sentences. This research focuses on participating in the SemEval 2020 Task 4 (ComVE) competition by developing classification and text generation models tailored for the Arabic language. The competition comprises three subtasks: Subtask A involves choosing the sentence that makes sense between two given sentences, Subtask B requires selecting the most appropriate reason from multiple choices for a sentence that goes against common sense, and Subtask C entails generating an explanation and reason for a sentence violating common sense. Our models leverage a set of multilingual pre-trained transformer models and have achieved remarkable performance in the competition. In Subtask A, our accuracy reached 84.7%, surpassing the performance of other works in Arabic. Similarly, in Subtask B, our approach outperformed other multilingual approaches, achieving a score of 79.3% compared to the state-of-the-art BERT model's 61%. In Subtask C, our model generated explanations with a BLEU score of 24, which is considered acceptable in the domain of text generation, particularly in the context of Arabic.

Keywords: Natural Language Understanding (NLU) · Arabic Natural Language Processing · Commonsense Reasoning · AraGPT · AraBERT · BERT Multilingual

1 Introduction

Natural Language Processing (NLP) is a sub-field of artificial intelligence dedicated to enabling computers to comprehend human languages. In recent years, this field has experienced significant advancements in various NLP tasks, including machine translation, question-answering, machine reading, and commonsense explanation [21]. Commonsense explanation refers to the ability of NLP systems to identify and generate coherent sentences, while also resolving sentence ambiguities and determining the most likely reason [7].

Supported by organization Jordan University of Science and Technology.

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 101–112, 2024.
https://doi.org/10.1007/978-3-031-56728-5_9

Arabic, as the official language of 22 nations, is spoken by more than 400 million people, making it the fourth most widely used language on the internet [10]. However, Arabic poses unique challenges for Natural Language Processing (NLP) due to its intricate grammar, script, and dialectal variations. As a result, recent research efforts have focused on improving the performance of NLP models specifically designed for Arabic text. Despite the progress made, numerous challenges persist, particularly in the areas of speech recognition, sentiment analysis, and machine translation. The linguistic complexities inherent in Arabic emphasize the need for further exploration and research in the field of NLP. Despite the extensive research conducted in natural language processing across different languages, Arabic language processing has not received sufficient attention. This is especially evident in the domain of commonsense reasoning, where limited efforts have been made to develop models capable of understanding and reasoning in Arabic.

In SemEval-2020, Task 4 focused on Commonsense Validation and Explanation (ComVE) [16]. This task aimed to train models to possess reasoning and comprehension capabilities. Task 4 encompassed three subtasks. Subtask A involved selecting the sentence that made sense from a pair of sentences. Subtask B required choosing the sentence that provided the most plausible explanation as to why a given sentence contradicted common sense, from multiple options. Lastly, subtask C involved generating an explanation and rationale for why a particular sentence defied common sense. The objective of this paper is to provide insights into the current state of Arabic NLP in the domain of commonsense explanation and validation. It aims to highlight the importance of further research in this field. To accomplish this, the paper aims to enhance the results achieved in task A and B, surpassing the findings of previous studies. Additionally, the paper addresses the lack of work done on task C for the Arabic language by translating the datasets specific to this task and build a text generation model specifically designed for the given objective. The primary goal is to develop a model that can generate text relevant to the task at hand. By implementing this approach, the paper seeks to contribute to the advancement of the field and address the specific challenges associated with text generation in this context.

The paper comprises five sections, with Sect. 2 dedicated to the related work. Section 3 provides an overview of the methodology, including a brief description of the dataset and the three subtasks. Additionally, it discusses the dataset preprocessing techniques employed and provides an explanation of the models used in the study. Section 4 presents the experimental setup and highlights the obtained results. Finally, Sect. 5 concludes the paper, summarizing the key findings and proposing potential directions for future research.

2 Related Work

Jon et al. [11] actively participated in all tasks of SemEval 2020 Task 4. For tasks A and B, they employed the ALBERT model and showcased its impressive performance by testing it in the Czech language, demonstrating the model's capability to achieve exceptional results with robust machine translation. Notably, the ALBERT model attained an accuracy of 95.8% for task A and 93.1% for task B. In task C, they trained the sequence-to-sequence model BART and obtained a notable BLEU score of 22.39.

Roweida et al. [12] participated in subtask A of the Commonsense Validation and Explanation. This subtask required selecting the sentence that makes sense among two given sentences. The authors employed the pre-trained BERT model as a baseline for comparison with their own model. Their model utilized a voting ensemble approach. It is compared with four different state-of-the-art models: BERT, ALBERT, Roberta, and XLNet. The proposed method outperformed the baseline model, achieving an accuracy score of 96.1% compared to the baseline's 89.1%.

Ali et al. (2020) [9] trained five different pre-trained transformer-based models, employing varying sizes for each model, in the context of the Commonsense Validation and Explanation for the three subtasks. Remarkably, among the 10 models evaluated in subtask A, Roberta-large exhibited the highest accuracy, achieving an impressive test accuracy of 93%. Similarly, in subtask B, Roberta-large outperformed 9 other models, attaining a test accuracy of 92.3%. Moving on to subtask c, gpt2-medium garnered a BLEU score of 16.1187 and a human evaluation score of 1.94, showcasing its proficiency in generating responses.

Wang et al. [18] implemented a multi-task framework to address the subtasks a and b of SemEval-2020 task 4. The proposed system, based on the BERT architecture. The authors trained BERT and RoBERTa as their baseline models. Notably, in subtask A, Roberta exhibited the highest accuracy, achieving an impressive score of 86.2%. Subsequently, in subtask B, they incorporated the data from subtask A, resulting in an accuracy of 82.3%. To tackle subtask c, the authors leveraged the GPT model to generate explanations, and upon incorporating the respective subtask data, the model achieved a BLEU score of 12.94.

Zhao et al. (2020) [20] introduced the Knowledge-enhanced Graph Attention Network (KEGAT) architecture, which aims to enhance machine commonsense knowledge by combining structured knowledge from a knowledge base and unstructured text. Through the utilization of advanced data augmentation techniques و the model has improved its commonsense reasoning skills. The KEGAT model achieves state-of-the-art accuracy in the subtasks a and b of Commonsense Explanation (Multi-Choice). Impressively, the model attains accuracy scores of 96.60% and 94.68% for subtasks a and b, respectively, outperforming the baseline models.

Sirwe et al. (2020) [14] conducted an investigation into the commonsense reasoning task within SemEval-2020 task 4, which revolves around the development of a system capable of providing explanations and reasoning. In their study, they explored various state-of-the-art deep learning architectures suitable for the three distinct subtasks. Proposing an innovative approach inspired by question-answering tasks, they transformed the classification problem into a multiple-choice question task, resulting in a remarkable performance improvement of 96.06%. Additionally, for the second subtask, involving the selection of a reason why a statement does not make sense, they achieved a commendable score of 93.7%. Lastly, in the final subtask, aimed at generating a reason to a nonsensical statement, they harnessed the power of the generative model GPT-2, and achieved a BLEU score of 6.1732.

Emran et al. [1] trained a set of multilingual pre-trained transformer models for subtask A of the Commonsense Validation and Explanation task in Arabic, wherein the objective is to select the sentence that make sense. The authors employed the BERT, XLM-MLM, and XLM-RoBERTA models on a translated dataset from English to Arabic. Additionally, they used the BERT-Base Multilingual model as the baseline for comparison. Notably, the XLM-RoBERTa model surpassed the baseline, achieving an accuracy score of 81.2%, compared to the baseline's accuracy score of 72.2%. This signifies the superior performance of the XLM-RoBERTa model in the given task.

3 Methodology

The following section presents the methodology pipeline, which illustrates the steps followed in this research, as depicted in Fig. 1, Additionally, it covers the subtasks, dataset, and the models employed in this study.

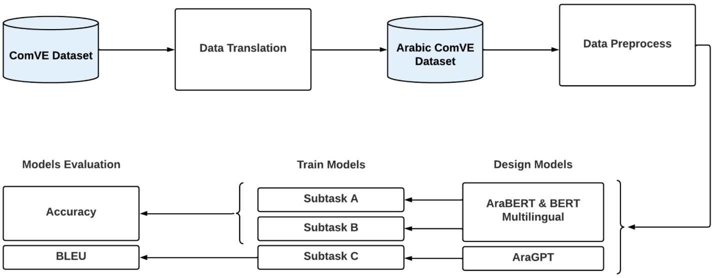


Fig. 1. Methodology Pipeline

3.1 SemEval-2020: Task 4

The Commonsense Validation and Explanation (ComVE) task evaluates a system's ability to determine which of two natural language statements makes sense to humans and which does not, while also providing explanations and reasons (SemEval Task 4 Com). SemEval-2020 Task 4 introduces this task, which comprises three subtasks: [16].

Subtask A: In this subtask, the model must select nonsensical sentence from a given pair of sentences.

Task: Which statement is against common sense?
Statement 1: She wear bangles in the hand.
Statement 2: She wear hat in the hand.
Answer: Statement 2

Subtask B: In this subtask, the model needs to choose the most appropriate explanation from three options for why an input sentence is against common sense.

Task: Choose the most fitting reason why this statement is against common sense.
Statement: She wear hat in the hand.
Reasons:
A: Hat usually black in colour and bangles are in different colours.
B: Hat usually wear on the head.
C: Hat will be big in size.
Answer: B

Subtask C: In this task, the model required to generate an explanation sentence why a given input sentence is against common sense.

Task: Generate the reason why this statement is against common sense.
Statement: She wear hat in the hand.
Referential Reasons:

– Hat usually wear on the head.
– A hat is not worn on hands.
– Hats are for your head.

3.2 Dataset

Task 4 of SemEval-2020 has its own dataset specifically collected and transformed for the competition [17]. The dataset is structured into separate files for each subtask, consisting of training, validation, and testing sets. An overview of the dataset can be found in Table 1 and Table 2.

In Subtask A, each record in the dataset contains two features: sent0 and sent1. Sent0 represents a text that goes against commonsense, while sent1 represents a commonsense statement. These pairs are designed to have a similar syntax structure, with only minor variations in the words used. The objective of the model is to correctly choose the option that contradicts common sense from the given pairs of sentences. Table 3 presents a sample for subtask A along with the corresponding answer provided by the model.

In Subtask B, the dataset consists of records with four features. The first feature corresponds to the sentence that goes against common sense, as identified in the previous subtask. The second, third, and fourth features represent different options that provide explanations for why the sentence contradicts common sense. Among these options, only one is correct, and the objective of the system is to accurately select the correct one. Table 4 presents a sample for subtask B along with the corresponding answer provided by the model.

In Subtask C, the dataset comprises records that correspond to sentences contradicting common sense, similar to the previous subtask. The objective remains the same: to identify the reason for each sentence's contradiction. However, in this subtask, the model is tasked with generating the reason from scratch. The generated text is then evaluated against three reference reasons explaining why the sentence goes against common sense. It is important to note that all of these reference reasons are correct. Table 5 presents a sample for subtask C along with the corresponding answer provided by the model.

Table 1. SemEval Task 4 Dataset Features & Subtasks

Subtask	Features	Use of features	Objective
A	2	Choose Option	Classification
B	4	Choose Option	Classification
C	4	Reference Answer	Text Generation

Table 2. Dataset Size

Subtask	No. records	Train	Validation	Test
A	11,997	10,000	997	1,000
B	11,997	10,000	997	1,000
C	11,997	10,000	997	1,000

Since we aimed to utilize this competition task in Arabic, it was necessary to obtain an Arabic version of the dataset. Fortunately, the paper titled "Is this sentence valid!" [15], presented translations for both subtask A and subtask B. As for subtask C, we utilized translation APIs, specifically Google API, to generate the Arabic version without encountering any difficulties[1]. Consequently, we now possess Arabic datasets for all subtasks. To the best of our knowledge, there is currently no publicly available dataset specifically tailored for task C in the field of commonsense explanation. Therefore, in this study, we aim to fill this gap by presenting a benchmark Arabic dataset that is specifically curated for the purpose of addressing the commonsense explanation problem, which involves generate a reson from scratch examining statements that contradict commonsense and determining why they do not make sense.

3.3 Preprocess

Text preprocessing is a crucial aspect of text mining as it significantly contributes to improving the quality of prediction algorithms and enhancing computational efficiency [3]. In our study, we leveraged the pre-processing capabilities provided by AraBERT, which utilizes a customized version of FarasaPy [6] designed specifically for Arabic text segmentation. Table 6 illustrates the text before and after undergoing the preprocessing step.

[1] https://github.com/ibrahim810/commonsense_ar_googleAPITranslate/.

Table 3. Subtask A Data Sample

label	Sentence 1	Sentence 0
0	سكب الحليب على حبوبه. He poured orange juice on his cereal.	سكب عصير البرتقال على حبوبه. He poured milk on his cereal.
1	أنا ألدغ بعوضة I sting a mosquito	البعوضة تلدغني A mosquito stings me
1	الزرافة شخص. A giraffe is a person.	ابنة الأخت شخص. A niece is a person

Table 4. Subtask B Data Sample

False Sentence	Options	Label
سكب عصير البرتقال على حبوبه.	A) عادة ما يكون عصير البرتقال برتقاليا لامعا. B) عصير البرتقال طعمه ليس جيدا مع الحبوب. C) عصير البرتقال لزج إذا سكبته على الطاولة.	B
أنا ألدغ بعوضة	A) الإنسان كائن ثديي. B) إنسان آكل C) الإنسان لا يمكن له أن يلسع البعوض	C
الزرافة هي شخص.	A) يمكن للزرافات شرب الماء من البحيرة. B) الزرافة ليست بشرا. C) عادة ما تأكل الزرافات أوراق الشجر.	B

3.4 Models

BERT Multilingual. BERT multilingual [8] is a widely utilized pretrained natural language processing model that has been trained on 104 languages. It fine-tuned using Wikipedia data, without the need for human labeling. Due to its extensive language coverage and effectiveness, it has become one of the most popular pretrained models in the field. BERT multilingual can be applied to various use case scenarios, including sequence classification, token classification, mask filling, and numerous other tasks. In our study, we employed the BERT multilingual model as a baseline for both subtask A and subtask B.

AraBERT. Arabic BERT [4] (AraBERT) was another model we employed for both subtask A and B. Similar to BERT multilingual, AraBERT adopts the same architecture as BERT but fine-tuned using Arabic datasets and resources. These include OSCAR, Arabic Wikipedia, The 1.5B words Arabic Corpus, The OSIAN Corpus [19], and Assafir news articles. With its specialization in the

Table 5. Subtask C Data Sample

False Sentence	References
سكب عصير البرتقال على حبوبه.	يسكب عصير البرتقال في كوب. عصير البرتقال طعمه ليس جيدا مع الحبوب. عصير البرتقال لا طعم له مع الحبوب.
أنا ألدغ بعوضة	الإنسان ليس لديه لسعات البعوض يلدغ الناس ، وليس العكس الإنسان لا يمكن له أن يلسع البعوض
الزرافة هي شخص.	الزرافة حيوان وليس بشري. الزرافة ليست بشرا. الزرافة حيوان.

Table 6. Data Sample After Preprocessing

Preprocessed	Original
يصب ال ماء في ال إبريق	يصب الماء في الإبريق
أستخدم ال إنترنت في ال راديو ال خاص بي	أستخدم الإنترنت في الراديو الخاص بي
أخذ ت ماري فراش ها في نزه ة	أخذت ماري فراشها في نزهة

Arabic language, AraBERT served as our primary model choice, leveraging its domain expertise to enhance performance in our study.

GPT-2. GPT-2 [13] stands out as one of the most renowned models in the domain of text generation. Since its creation by OpenAI in 2019, it has exhibited impressive performance across a range of NLP tasks, including text summarization, question answering, language translation, and text generation, among others. The model was fine-tuned using a custom dataset called WebText, which comprises posts and comments sourced from Reddit. To ensure the quality of the data, only posts and comments with a minimum of 3 karma (a Reddit score unit) were included. Although WebText is not publicly accessible, it encompasses the most frequently used domains within the dataset, showcasing the model's versatility and broad coverage.

AraGPT. Arabic GPT2 [5] is the model that has been utilized for subtask C. It is a custom GPT model specifically designed for the Arabic language. AraGPT is constructed using the GPT-2 architecture and fine-tuned using a diverse range of online resources, including the OSCAR corpus, Arabic Wikipedia dump, OSIAN Corpus, and others. The model is available in four different versions: base, medium, large, and mega. These versions vary in terms of training duration, the number of examples used for training, batch size, compute unit, number of layers, number of parameters, and size.

Metrics. Both subtask A and B are classification tasks, so we used the accuracy metric, which represents the percentage of correct predictions made by our model. The accuracy is computed as follows.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

The bilingual evaluation understudy (BLEU) is the metric we used to evaluate our work in subtask C. It is a unit used to assess the quality of machine-generated text. To evaluate the fluency, sufficiency, and acceptability of a translation, BLEU ratings are calculated using precision, recall, and other accuracy criteria to evaluate the fluency, adequacy, and acceptability of a translation.

$$BLEU = BP \cdot exp(\sum_{k=1}^{n} w_k log(p_k)) \tag{2}$$

$$BP = e^{min(1 - \frac{len(reference)}{len(prediction)}, 0)} \tag{3}$$

4 Experiments and Results

To conduct our experiments, we utilized the Colab Pro platform provided by Google, utilizing an A100 GPU and 40 GB of RAM. Various hyperparameters were employed to evaluate the testing dataset, including epochs, batch size, and learning rates (LR). The experiments were performed using the Adam optimizer. Notably, we observed that the optimal LR value was found to be 5e−5. For subtasks A and B, we conducted experiments with 10 epochs and a batch size of 16, while for subtask C, we increased the number of epochs to 40. These parameters were chosen based on our experimentation and evaluation process. To evaluate the performance of the pre-trained models, we utilized different metrics for each subtask. For subtasks A and B, we employed the accuracy metric. On the other hand, for subtask C, we utilized the BLEU metric to assess the quality of the generated text.

Table 7, presents the results of subtasks A, B, and C, using different models and metrics. For Subtask A and B, the models used are AraBERT and BERT Multilingual, and the metric used is accuracy. The results show that AraBERT performed better than BERT Multilingual with 84.7% accuracy and 79.3% accuracy respectively. For subtask C, the AraGPT model was utilized, and the evaluation was based on the BLEU metric. The obtained result was 24.821, indicating the quality of the generated text. In this research, our findings have surpassed the results reported in previous experiments conducted by [1,15] for subtask A. Additionally, we have achieved better results for subtask B compared to the previous work referenced [2].

Table 7. Results

Subtask	Model	Metric	Results
Subtask A	AraBERT BERT Multilingual	Accuracy	84.7% 65.0%
Subtask B	AraBERT BERT Multilingual	Accuracy	79.3% 61%
Subtask C	AraGPT	BLEU	24.821

5 Conclusion and Future Work

In this research, we developed classification and text generation models to participate in the custom Arabic version of the SemEval 2020 task 4 (ComVE) competition. In the first subtask of the competition, our approach achieved an accuracy of 84.7%, which is the highest score compared to other works conducted in Arabic. Furthermore, in the other subtasks, our work outperformed other multilingual approaches. For subtask b, our approach achieved a notable accuracy of 79.3%, surpassing the well-known state-of-the-art BERT model that only achieved 61%. In subtask c, we obtained a BLEU score of 24.821, which is considered acceptable in the field of text generation, particularly for the challenging Arabic language. It is worth mentioning that the majority of advancements in natural language processing primarily focus on the English language domain, making our achievements in Arabic even more significant. Arabic is known to present challenges in natural language processing due to the dominance of English in the field. Through this research, we have demonstrated the effectiveness of our classification and text generation models in addressing the challenges of commonsense understanding in Arabic. These results contribute to the advancement of natural language processing in Arabic and pave the way for further research in this field. In future, we aim to expand our research by conducting additional experiments to further explore the capabilities of different pre-trained models. By testing and comparing the performance of various models, we anticipate improving our results and enhancing the overall performance of the system. Additionally, we plan to explore alternative translation APIs for subtask C, aiming to optimize the translation process and potentially improve the quality of the generated text. These planned enhancements will contribute to a more comprehensive and robust research approach, enabling us to advance the field of study and achieve more impactful results.

References

1. Al-Bashabsheh, E., Al-Khazaleh, H., Elayan, O., Duwairi, R.: Commonsense validation for Arabic sentences using deep learning. In: 2021 22nd International Arab Conference on Information Technology (ACIT), pp. 1–7. IEEE (2021)
2. AL-Tawalbeh, S., AL-Smadi, M.: A benchmark Arabic dataset for commonsense explanation. arXiv preprint arXiv:2012.10251 (2020)

3. Alshanik, F., Apon, A., Herzog, A., Safro, I., Sybrandt, J.: Accelerating text mining using domain-specific stop word lists. In: 2020 IEEE International Conference on Big Data (Big Data), pp. 2639–2648. IEEE (2020)
4. Antoun, W., Baly, F., Hajj, H.: Arabert: transformer-based model for Arabic language understanding. In: LREC 2020 Workshop Language Resources and Evaluation Conference 11–16 May 2020, p. 9 (2020)
5. Antoun, W., Baly, F., Hajj, H.: AraGPT2:pPre-trained transformer for Arabic language generation. In: Proceedings of the Sixth Arabic Natural Language Processing Workshop, pp. 196–207. Association for Computational Linguistics, Kyiv, Ukraine (Virtual) (2021). https://www.aclweb.org/anthology/2021.wanlp-1.21
6. Darwish, K., Mubarak, H.: Farasa: a new fast and accurate Arabic word segmenter. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016). European Language Resources Association (ELRA) (2016)
7. Davis, E.: Logical formalizations of commonsense reasoning: a survey. J. Artif. Intell. Res. **59**, 651–723 (2017)
8. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. CoRR **abs/1810.04805** (2018). http://arxiv.org/abs/1810.04805
9. Fadel, A., Al-Ayyoub, M., Cambria, E.: Justers at semeval-2020 task 4: evaluating transformer models against commonsense validation and explanation. In: Proceedings of the Fourteenth Workshop on Semantic Evaluation, pp. 535–542 (2020)
10. Guellil, I., Saâdane, H., Azouaou, F., Gueni, B., Nouvel, D.: Arabic natural language processing: an overview. J. King Saud Univ. Comput. Inf. Sci. **33**(5), 497–507 (2021)
11. Jon, J., Fajčík, M., Dočekal, M., Smrž, P.: But-fit at semeval-2020 task 4: Multilingual commonsense. arXiv preprint arXiv:2008.07259 (2020)
12. Mohammed, R., Abdullah, M.: Teamjust at semeval-2020 task 4: Commonsense validation and explanation using ensembling techniques. In: Proceedings of the Fourteenth Workshop on Semantic Evaluation, pp. 594–600 (2020)
13. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I.: Language models are unsupervised multitask learners. OpenAI blog (2019)
14. Saeedi, S., Panahi, A., Saeedi, S., Fong, A.C.: CS-NLP team at SemEval-2020 Task 4: evaluation of state-of-the-art NLP deep learning architectures on commonsense reasoning task. arXiv preprint arXiv:2006.01205 (2020)
15. Tawalbeh, S., Al-Smadi, M.: Is this sentence valid? an Arabic dataset for commonsense validation. arXiv preprint arXiv:2008.10873 (2020)
16. Wang, C., Liang, S., Jin, Y., Wang, Y., Zhu, X., Zhang, Y.: SemEval-2020 task 4: commonsense validation and explanation. In: Proceedings of The 14th International Workshop on Semantic Evaluation. Association for Computational Linguistics (2020)
17. Wang, C., Liang, S., Zhang, Y., Li, X., Gao, T.: Does it make sense? and why? a pilot study for sense making and explanation. arXiv preprint arXiv:1906.00363 (2019)
18. Wang, H., et al.: Cuhk at semeval-2020 task 4: commonsense explanation, reasoning and prediction with multi-task learning. arXiv preprint arXiv:2006.09161 (2020)

19. Zeroual, I., Goldhahn, D., Eckart, T., Lakhouaja, A.: OSIAN: open source international Arabic news corpus - preparation and integration into the CLARIN-infrastructure. In: Proceedings of the Fourth Arabic Natural Language Processing Workshop, pp. 175–182. Association for Computational Linguistics, Florence, Italy (2019). https://doi.org/10.18653/v1/W19-4619, https://aclanthology.org/W19-4619
20. Zhao, Q., Tao, S., Zhou, J., Wang, L., Lin, X., He, L.: Ecnu-sensemaker at semeval-2020 task 4: Leveraging heterogeneous knowledge resources for commonsense validation and explanation. arXiv preprint arXiv:2007.14200 (2020)
21. Zhou, M., Duan, N., Liu, S., Shum, H.Y.: Progress in neural NLP: modeling, learning, and reasoning. Engineering **6**(3), 275–290 (2020)

Predicting Students Answers Using Data Science: An Experimental Study with Machine Learning

Malak Abdullah(✉), Naba Bani Yaseen, and Mohammad Makahleh

Department of Computer Science, Jordan University of Science and Technology,
Irbid, Jordan
mabdullah@just.edu.jo, {nmbaniyaseen21,maalmakahleh20}@cit.just.edu.jo

Abstract. In today's data-driven world, the abundance of information provides us with opportunities to explore the relationships between various data points, leading to progress in multiple domains. For instance, in the field of education, we can leverage students' past course performance and academic records to offer tailored guidance, allowing them to concentrate their efforts on specific areas for academic growth. By employing machine learning techniques, we can analyze data relations and predict future events based on historical data. In this study, we utilized machine learning techniques on the educational dataset from NeurIPS 2020. We aimed to improve the prediction of upcoming student performance by adding valuable features. To accomplish this, we explored several classification algorithms, including SVM, Naive Bayes, Logistic Regression, and Decision Tree. Additionally, we considered Ensemble methods such as Boosting, Bagging, and Voting. By assessing the optimal hyperparameter values for these algorithms, we aimed to optimize their performance. Our findings revealed that augmenting the dataset with more correlated features significantly improved prediction accuracy. Among the classifiers examined, Decision Tree, XG Boost, and Voting exhibited the best performance, achieving an accuracy rate of 74%.

Keywords: Education · Machine Learning · Classification

1 Introduction

Education, encompassing the study of teaching methods and educational institutions like schools, colleges, and universities, holds immense importance in our lives [1]. It serves as a catalyst for the development of our intellect and enriches our existence with value and excellence, while also fostering self-assurance and equipping individuals to participate in today's world [2]. Therefore, universal access to education for all children and youth should be a priority for every nation [3]. As technology and artificial intelligence (AI) continue to reshape various sectors, including education, the landscape of teaching and learning is evolving [4].

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 113–124, 2024.
https://doi.org/10.1007/978-3-031-56728-5_10

Machine Learning (ML) is the science of enabling computers to learn from data without explicit programming. ML systems leverage training datasets to learn, testing datasets to evaluate predictions, and performance metrics like accuracy, precision, and recall to assess models [5]. ML algorithms encompass a range of applications, including classification, regression, recognition, and clustering, allowing them to process vast amounts of data and predict categorical class labels based on training data. Notable classification algorithms include Naïve Bayes, Logistic regression, J48, KNN, Decision Tree, Neural Network (NN), and Support Vector Machine [6].

In the field of education, ML classification algorithms find numerous applications [7,8]. They help teachers identify struggling student clusters, predict student performance, and enable customized learning experiences based on individual data [9–11]. AI tools, by simulating student behaviors, contribute to educational theory and support teachers in guiding students effectively [10].

This research focuses on utilizing machine learning techniques to enhance the prediction of future student performance using the NeurIPS 2020 dataset. The study involves experiments and techniques aimed at improving performance prediction by incorporating valuable features, employing multiple ML algorithms, tuning hyperparameters, and applying Ensembling methods. The results emphasize the significance of input data quality, the impact of specific features like student smartness and question difficulty levels, and the role of hyperparameters in model performance. For example, the Decision Tree model achieves an accuracy of 74% under specific conditions, outperforming other algorithms. Among the Ensembling methods, the voting model attains the highest accuracy at 74.5%. The subsequent sections of this paper provide further details, including a literature review, methodology, results, and concluding remarks with future directions in Sect. 5.

2 Literature Review

Numerous researchers have explored using machine learning models to study educational data and its correlation with students' performance. Several studies have examined student performance in various educational settings, including higher education institutions and middle schools, focusing on predicting and classifying performance using different machine learning algorithms. Rashmi Agrawal et al. [12] aimed to assist educational organizations in evaluating student performance and determining the probability of success by applying machine learning algorithms. They employed a dataset of 163 instances with sixteen attributes from Keyboard 360, utilizing various techniques in the Weka tool. The Multilayer Perceptron classifier algorithm demonstrated the highest performance, providing valuable insights for decision-making in the student's best interest. Kamil Dimililer et al. [11] used different machine learning algorithms and eighteen trials to predict and classify student performance. According to the preliminary findings, student performance can be anticipated and improved if we prepare data before using machine learning techniques. After applying all algorithms, BP outperformed other categorization techniques and gave the highest accuracy.

In middle school, Harikumar et al. [13] investigated students' course performance based on their previous academic records in similar courses. They utilized Machine learning algorithms such as Naive Bayes, ID3, C4.5, and SVM to evaluate students' talents and interests as they might be linked to their performance. A data set of UCI machinery student performance was used. The metrics used to evaluate the models were accuracy and error rate. The results highlighted the superiority of the SVM algorithm, delivering the most accurate classification outcomes.

In a set of secondary schools, Ihsan A. Abu Amra et al. [14] aimed to develop a model for predicting students' performance using educational datasets. They applied two classification algorithms, KNN and Naive Bayes, to anticipate student performance and support educational institutions and ministries in enhancing academic outcomes. By comparing accuracy, recall, and precision, they determined that the Naive Bayes model yielded the highest accuracy due to the inherent relationships within the dataset. Some research has studied student performance by school location; In the schools located in the rural and the urban, Mohamed Shanavas et al. [15] used machine learning algorithms in the Weka tool to predict and classify student performance. Classification algorithms were used, and they were compared according to the execution time and the accuracy of the results. They found that the random forest gives better results than the other algorithms and gives the most accurate results.

3 Methodology

Figure 1 shows the steps that the proposed framework architecture perform it to complete the classification task. First, we used data pre-processing to the training and testing data set then we applied traditional classification models tuning, we also considered tuning the Ensembling models.

3.1 Dataset

For this project, we utilized the NeurIPS2020 Education Challenge dataset from tasks 1 & 2 [16]. This dataset comprises student responses to multiple-choice diagnostic questions spanning from September 2018 to May 2020. The data was collected from the Eedi website, a widely used online education platform in numerous schools worldwide. The platform primarily focuses on providing multiple-choice questions to students at primary and secondary levels, with each question presenting four options, of which one is correct. Notably, the platform has recently emphasized mathematics-related questions. Within the NeurIPS2020 Education competition, there are four tasks, with the first and second tasks sharing the same dataset, while the third and fourth tasks utilize a distinct dataset. In our project, we specifically concentrate on task one and exclusively analyze the datasets pertaining to this task.

1. **Primary Data:** main training data, consisting of records of answers given to questions by students. It can be found in the train_task_1_2.csv files. Table

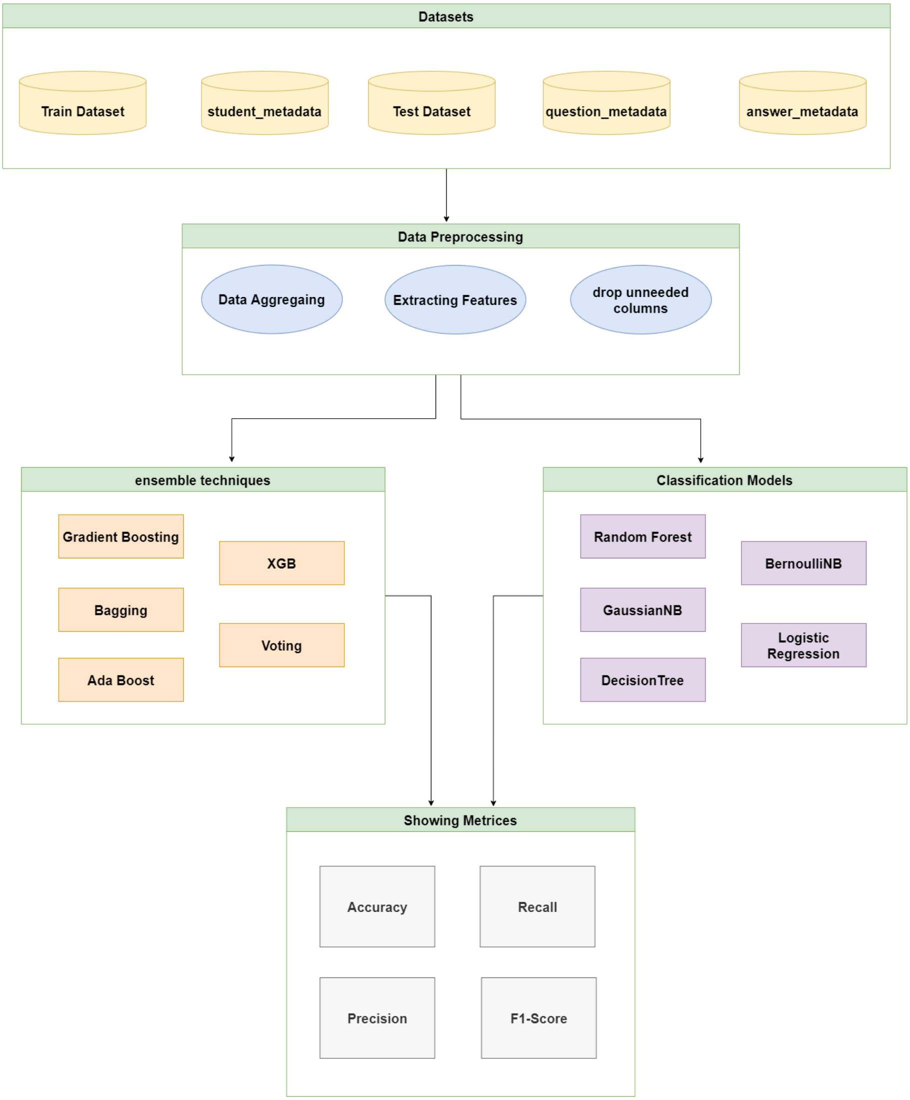


Fig. 1. Proposed framework architecture

Table 1. Columns Inside train_task_1_2 File

Number	Name	Data Type	Description
1	QuestionId	Numeric	ID of the question answered
2	UserId	Numeric	ID of the student who answered the question
3	AnswerId	Numeric	Unique identifier for the (QuestionId, UserId) pair, used to join with associated answer metadata
4	IsCorrect	Binary	Wither the answer was correct or not (1 is correct, 0 is incorrect)
5	CorrectAnswer	Numeric	The correct answer to the multiple-choice question (value in [1, 2, 3, 4])
6	AnswerValue	Numeric	Student's answer to the question (value in [1, 2, 3, 4])

1 shows the description for each column inside train_task_1_2.
For tasks 1 and 2, the individual answer records are randomly split into 80%/10%/10% training/public test/private test sets.

2. **Question Metadata:** Table 2 shows metadata provided about each question.

Table 2. Columns Inside Question Metadata

Number	Name	Data Type	Description
1	SubjectId	Numeric	ID Each subject covers an area in a list
2	QuestionId	Numeric	ID of the question

3. **Student Metadata:** Table 3 shows metadata provided about each students.

Table 3. Columns Inside Student Metadata

Number	Name	Data Type	Description
1	UserId	Numeric	An ID uniquely identifying the student
2	Gender	Numeric	The student's gender, when available. 0 is unspecified, 1 is female, 2 is male and 3 is other
3	DateOfBirth	String	The student's date of birth
4	PremiumPupil	Binary	Whether the student is eligible for free school meals or pupil premium due to being financially disadvantaged

4. **Answer Metadata:** Table 4 shows metadata provided about each individual answer record in the dataset.

3.2 Data Pre-processing

Preprocessing indicates all the transformations on the raw data in an understandable format before feeding it to the ML or DL algorithms. In machine

Table 4. Columns Inside Answer Metadata

Number	Name	Data Type	Description
1	AnswerId	Numeric	An ID uniquely identifying the answer, which can be joined to the primary dataset
2	DateAnswered	Numeric	Time and date that the question was answered, to the nearest minute
3	Confidence	String	Percentage confidence score given for the answer 0 means a random guess, 100 means total confidence
4	GroupId	Binary	The class (group of students) in which the student was assigned the question
5	QuizId	Binary	The assigned quiz which contains the question the student answered

learning, preprocessing of data is an essential step as we can derive useful data and high quality information that affect the efficiency of our model

- **Data Aggregation:** The training data file was read and merged with the file containing students' information based on the user ID. Subsequently, the resulting file was merged with the answers metadata file using the answer ID. Additionally, the training file was merged with the question file, extracting subject identifiers for each question in the training data. Furthermore, these files were combined with the testing dataset to consolidate the relevant information.
- **Extract Features:** To increase models efficiency we extracted new features from the current features and delete some of them.
  1. **Age:** Date of birth column exists is given for all students, so we extracted the age feature for each student using his date of birth. At first, we read the current date then we subtract the date of birth of each student separately from the current date to produce a column containing the ages of all students. For irrational values less than 10 or more than 90, or missing values with the average age which is 40.
  2. **Student smartness:** We calculated the intelligence of each student in the dataset by applying this equation for each student: student smartness column = student correct answers count/student answer count.
  3. **Question Easy Level:** We added a question related feature that measures the ratio that the question was solved correctly using this equation: Question Easy Level = question asked count/question solved correctly count.
  4. **Quiz Easy Level:** Similar to Question Easy Level, we calculated the degree of the current quiz questions was solved right with the equation: Quiz Easy Level = quiz appeared count/quiz solved correctly count
  5. **Subject Id:** We took the first three levels of the subjectId column because the indexes that follow the third index, most of their values are null, so they are not useful. After this process, we have three columns, we named them as subjectId_L1, subjectId_L2, and subjectId_L3.

- **Handle null Values:** We treated the missing values in these two columns: group id and premium pupil, and considered that the missing values are undefined values. We treated students with an empty group ID as not belonging to any group.
- **Unneeded columns drop:** We removed the columns: Correct Answer, Birth date, Subject Id, Answer Id, Confidence, and Scheme Of Work Id.
  Table 5 shows the features superset(final set of features) that will be used to compare algorithms.

Table 5. Class label instances count for training and testing datasets

Dateset	Ones count	Zeros count
Train Data	5,022,623	2,493,552
Test Data	628,413	312,223

- **Features affect:** To feed the algorithm with more meaningful data, we studied the effect of adding features on the model results in order to build superset features from all of the preprocessed features that lead to the best classification results. Table 6 shows the relation between adding more custom features to the dataset on the best classification accuracy we could get from running all classifiers including the ensembling methods classifiers.
  When we added the custom features (Subject_L1, Subject_L2, Subject_L3, Age) to the basic features, we could get 66% accuracy. Then we had all previous columns along with StudentSmarness which boosted the accuracy result up to 72%. So the student's ability to solve previous questions correctly is a very good indicator to his capability of solving upcoming questions correctly. Finally, we added QuestionEasyLevel feature to the dataset which resulted in increasing accuracy rate to 74%. We notice the strong relation between the correctly solving rate of a question with the ability of a student to solve it right. Table 7 shows the features superset(final set of features) that will be used to compare algorithms.

Table 6. The effect of adding features on Accuracy

Experiment	Columns	Best Accuracy
#1	Basic	0.64
#2	Subject_L1, Subject_L2, Subject_L3, Age	0.66
#3	Subject_L1, Subject_L2, Subject_L3, Age, StudentSmarness	0.72
#4	Subject_L1, Subject_L2, Subject_L3, Age , StudentSmarness, QuestionEasyLevel, QuizeEasyLevel	0.74.5

Table 7. final set of features

number	column name	Column data type
1	QuestionId	Numeric
2	UserId	Numeric
3	IsCorrect	Binary
4	Gender	Numeric
5	PremiumPupil	Binary
6	GroupId	Numeric
7	QuizId	Numeric
8	Age	Numeric
9	Subject_L1	Numeric
10	Subject_L2	Numeric
11	Subject_L3	Numeric
12	StudentSmartness	Numeric
13	QuestionEasyLevel	Numeric
14	QuizeEasyLevel	Numeric

4 Experiments and Comparison

In our study, we experimented with several popular machine learning algorithms for classification along with many combinations of the hyperparameters in order to tune the models during the experiments. Also we compared classification results using the Ensembling methods.

Evaluation Metric. For Evaluation metric in this study, we used four performance metrics to describe and compare the performance of above classifiers methods on the dataset: accuracy, precision, recall, and F1-score.

1. Accuracy: is a measure of a model's correctness, defined as the ratio of correct predictions to total predictions. The following equation is used to measure it [17]:
$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$
2. Precision: is defined as the ratio of correct positive predictions to total positive predictions, and it is used to calculate the positive predicted value. The following equation is used to measure it [18]:
$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$
3. Recall: is the ratio of the number of correct positive predictions to the total number of correctly predicted outcomes, it is also testing the model's sensitivity. The following equation is used to measure it [17]:

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

4. The F1-score: is defined as the harmonic mean of precision and recall, is also used to test the model's accuracy. The following equation is used to measure it [18]:

$$F1 - score = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

4.1 Models Tuning

Our experiments focused on getting and comparing the best classification results from each model by tuning model hyperparameters.

For each classifier, we repeated the experiment for all different values of the hyperparameters by using the Grid Search method to conclude the best value for each single hyperparameter with respect to the output results.

For the Naive Bayes model, we used the default hyperparameters provided by Sikit-Learn libray. In the Decision Tree classifier We studied the effect of criterion and max tree depth. The higher accuracy for the Decision Tree classifier was achieved with entropy criterion and 7 as the max tree depth.

In addition to the traditional classification models tuning, we also considered tuning the Ensembling models tuning by changing their hyperparameters. We run Random Forest Classifiers with all combinations of the number of estimators, max depth, and criterion parameters. Best accuracy rate observed on the number of estimators equals to 300, max depth is 7, and entropy as criterion. The study also covered the effect of Bagging Classifier hyperparameters. For the ADA boost model, the best values for its parameters were 100 for the number of estimators, and entropy for the criterion. Finally, changing the voting classifier estimators extremely impact the results of the models voting. During the experiment we tested many combinations of related and unrelated classifiers for voting. The best set of models were XG Boost, Gradient Boost, ADA Boost, and Decision Tree.

4.2 Models Comparison

In this study, we compared multiple classification algorithms and ensemble them to achieve the best accuracy outcomes. Figure 2 shows models classification accuracy comparison. It demonstrates the comparison of the highest accuracy result for each model using our features superset. We can observe that the Decision Tree achieved the highest accuracy rate among the models with 74%.

To find an ideal prediction model, we combine the models that we used in this research, voting ensemble method gave us an accuracy of 74.5%, while we achieved a lower accuracy of 74% when we used Gradient Boosting, XG Boosting, and ADA Boosting methods. Figure 3 shows Ensembling classification

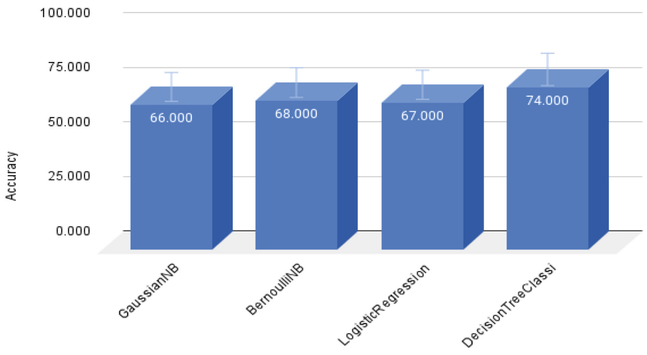


Fig. 2. Classification Models Accuracy

accuracy comparison. To evaluate the effectiveness of the models that we used, we made use of a combination of accuracy, precision, recall, and F1-score. Results are reported in Table 8 across all our experimental configurations. The results show that The Voting ensembling classifier shows the best metrics values around 74.5%. When comparing models based on the precision, we found that the best model is the one that has the highest precision score of 75% is Bernoulli NB.

Table 8. Models metrics results

Classifier	Accuarcy	Precision	Recall	F1
BernoulliNB	68	75	68	57
GaussianNB	66	57	66	54
LogisticRegression	67	45	67	54
DecisionTreeClassifier	74	74	72	73
RandomForestClassifier	73	72	73	72
GradientBoostingClassifier	74	74	74	74
BaggingClassifier	74	70	71	71
AdaBoostClassifier	74	74	75	74
XGBClassifier	74	74	75	74
VotingClassifier	74.5	74	75	74

In our dataset the class is imbalanced, there are too few examples of the minority class, the Zero class, to learn the decision boundary effectively. We used resampling technique by adding or removing examples from the training dataset to change the distribution of classes and to adjust the class distribution of a data set (i.e. the ratio between different classes). With the usage of resampling techniques, the number of ones and Zeros in this column became same as number

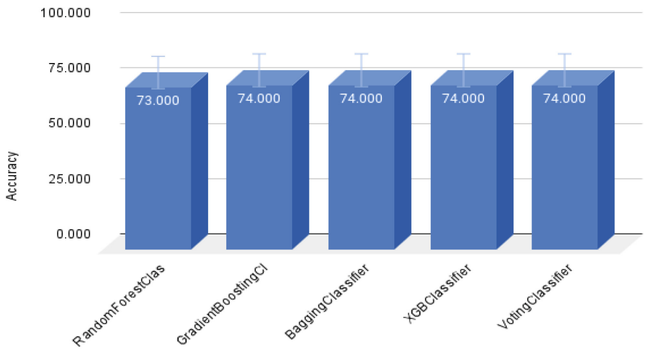


Fig. 3. Ensembling Classifiers Accuracy

on Ones. So the data is balanced, and the model is trained on the ones and zeros equally. However, the application of the resample techniques did not cause any positive effect on the accuracy but rather it to reduces it, as the accuracy became 53%.

5 Conclusion

Machine learning has significantly contributed to education by enabling various advancements, including predicting student behavior and identifying students needing assistance. In this paper, we applied machine learning techniques to the NeurIPS 2020 educational dataset, intending to enhance the prediction of students' future performance. Moreover, we explored the impact of different student features on the performance of classification algorithms, considering their hyperparameters. Furthermore, we compared various machine learning algorithms encompassing traditional and Ensemble classification methods. The results underscore the substantial influence of the data provided to the classification algorithms on their learning capabilities and overall performance. Our experiments revealed that specific features, such as the student's aptitude in solving previous questions, the difficulty level of questions based on past students' proficiency, and the ease of quizzes, played a significant role in enhancing the models' learning compared to other features like student age and gender.

Furthermore, we investigated the effect of various hyperparameters on the models' performance and identified the optimal values for each parameter during our study. Additionally, we evaluated the impact of different classification algorithms, including Decision Tree, Random Forest, Naive Bayes, and Logistics Regression. Among these algorithms, the Decision Tree algorithm demonstrated superior performance, achieving an accuracy of 74%. Moreover, we explored the utilization of Ensembling methods, which yielded promising results. All models exhibited acceptable accuracy and effectively performed the classification task. Notably, the voting model emerged as the most successful, attaining an accuracy of up to 74.5%.

References

1. Brighouse, H.: Education. Routledge (2012)
2. Orr, D.: What is education for. Context **27**(53), 52–58 (1991)
3. Kučak, D., Juričić, V., Dambić, G.: Machine learning in education-a survey of current research trends. In: Annals of DAAAM & Proceedings, vol. 29 (2018)
4. Holmes, W., Bialik, M., Fadel, C.: Artificial intelligence in education. In: Boston: Center for Curriculum Redesign, vol. 2019, pp. 1–35 (2019)
5. Géron, A.: Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media (2019)
6. De La Hoz, E.J., Fontalvo, T.: Methodology of machine learning for the classification and prediction of users in virtual education environments (2019)
7. Abdullah, M., Al-Ayyoub, M., AlRawashdeh, S., Shatnawi, F.: E-learningDJUST: E-learning dataset from Jordan university of science and technology toward investigating the impact of Covid-19 pandemic on education. Neural Comput. Appl. 1–15 (2021)
8. Abdullah, M., Al-Ayyoub, M., Shatnawi, F., Rawashdeh, S., Abbott, R.: Predicting students' academic performance using e-learning logs. IAES Int. J. Artif. Intell. **12**(2), 831 (2023)
9. Woolf, B.P., Lane, H.C., Chaudhri, V.K., Kolodner, J.L.: AI grand challenges for education. AI Mag. **34**(4), 66–84 (2013)
10. Woolf, B.P.: AI and education: celebrating 30 years of marriage. In: AIED Workshops. Citeseer (2015)
11. Sekeroglu, B., Dimililer, K., Tuncal, K.: Student performance prediction and classification using machine learning algorithms. In: Proceedings of the 2019 8th International Conference on Educational and Information Technology, pp. 7–11 (2019)
12. Jalota, C., Agrawal, R.: Analysis of educational data mining using classification. In: 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), pp. 243–247. IEEE (2019)
13. Darmayanti, I., Subarkah, P., Anunggilarso, L.R., Suhaman, J.: Prediksi potensi siswa putus sekolah akibat pandemi covid-19 menggunakan algoritme k-nearest neighbor. JST (Jurnal Sains dan Teknologi) **10**(2), 230–238 (2021)
14. Amra, I.A.A., Maghari, A.Y.: Students performance prediction using KNN and Naïve Bayesian. In: 2017 8th International Conference on Information Technology (ICIT), pp. 909–913. IEEE (2017)
15. Mythili, M., Shanavas, A.M.: An analysis of students' performance using classification algorithms. IOSR J. Comput. Eng. **16**(1) (2014)
16. Wang, Z., et al.: Diagnostic questions: the neurips 2020 education challenge. arXiv preprint arXiv:2007.12061 (2020)
17. Gunawardana, A., Shani, G.: A survey of accuracy evaluation metrics of recommendation tasks. J. Mach. Learn. Res. **10**(12) (2009)
18. Hossin, M., Sulaiman, M.N.: A review on evaluation metrics for data classification evaluations. Int. J. Data Min. Knowl. Manag. Process **5**(2), 1 (2015)

Arabic News Articles Classification Using Different Word Embeddings

M. Moneb Khaled[1], Muhammad Al-Barham[2], Osama Ahmad Alomari[2], and Ashraf Elnagar[1](✉)

[1] Department of Computer Science, University of Sharjah, Sharjah, United Arab Emirates
{U22104183,ashraf}@shrajah.ac.ae

[2] MLALP Research Group, University of Sharjah, Sharjah, United Arab Emirates
{malbarham,oalomari}@shrajah.ac.ae

Abstract. With the accelerated growth of the internet, vast repositories of unstructured textual data have emerged, necessitating automated categorization algorithms for organization and insight extraction. The Arabic language, however, poses particular challenges due to its inflected nature, large vocabulary, and varying forms. This study targets the development of robust automated classification systems for Arabic text, a language increasingly adopted online. In this paper, we propose a comparison of four prevalent pre-trained word embeddings: Word2Vec (represented by Aravec), GloVe, FastText, and BERT (represented by ARBERTv2), using the widely-adopted SANAD dataset of Arabic news articles. We provide a comprehensive comparison by applying a fixed deep learning architecture across all four word embeddings to ensure fairness. The motivation behind this comparison is to bridge the knowledge gap observed in the usage of popular word embeddings for Arabic news classification. Despite the state-of-the-art results from transformer models, a significant inclination towards older methodologies still persists. Hence, we aim to highlight the efficiencies of modern techniques. Results indicate that ARBERTv2 outperforms the other embeddings, achieving 95.81%, 98.68%, and 99.30% accuracy on the Akhbarona, Alkhaleej, and Alarabiya subsets of SANAD, respectively. Despite its large number of parameters, ARBERT's context-based word embeddings seem to offer superior performance. FastText stood out as the top performer among non-contextualized word embeddings due to its ability to capture morphological similarities and handle out-of-vocabulary words. Following closely behind was GloVe, and then came Aravec.

Keywords: Arabic news articles · Deep learning · Word embeddings · BERT

1 Introduction

The rapid growth of the Internet and the evolution of Web 2.0 have led to the emergence of vast repositories of online documents. With nearly 80% of the

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 125–136, 2024.
https://doi.org/10.1007/978-3-031-56728-5_11

data being textual and unstructured, putting to use its potential as a valuable source of information is crucial [10]. Therefore, there is a growing demand for automatic categorization algorithms to organize and extract insights from this unstructured information. Machine learning (ML) algorithms have been successfully implemented to manage this massive volume of data thereby improving the overall user experience by simplifying automated navigation, and making the search for specific information more feasible [1].

Among the fundamental tasks in Natural Language Processing (NLP), text classification plays a vital role. It involves categorizing text based on its content. This approach has been successfully applied in various areas [4] such as spam filtering, sentiment analysis, and language/dialect identification. Due to its capacity for detecting patterns, identifying relationships and providing faster results, ML-based data structuring is quite beneficial in the business domain. For example, marketers can leverage it to investigate and analyze keywords used by competitors, which assists strategic decision-making processes [3].

Despite the versatile application of text classification algorithms, particular challenges arise when dealing with the Arabic language. It is a highly inflected and derived language with different forms and a significantly larger vocabulary compared to English [16]. Moreover, Arabic is the mother tongue of over 422 million people [25], making it crucial to develop effective NLP solutions for this language. While research efforts have increased in Arabic computational linguistics, there is still room for further advancement. The Arabic language, as of 2019, is the fourth most popular language used online, accounting for 5.2% of all Internet users. Moreover, the growth rate for the number of Arabic online users has been the highest among all languages over the past nineteen years [10]. Furthermore, the inherent challenges of the Arabic language include its multiple forms (Modern Standard Arabic (MSA) and Dialectal Arabic), extensive vocabulary, distinctive alphabet set, irregular grammar, and unique sentence structure, which pose a challenging task for text classification systems [5]. Considering these statistics and challenges, the development of effective and robust automated classification systems for Arabic text is essential and needed.

With the increasing adoption of web scraping techniques, extensive datasets have started to arise, enabling Deep Learning (DL) algorithms to excel as they operate more efficiently with large amounts of data. DL models utilize deep neural networks to replicate the functions of the human brain, and they have consistently demonstrated superior performance when compared to classical ML methods in areas such as speech recognition, computer vision, and NLP.

Feature extraction is the process of converting raw data into a meaningful set of features, which can be utilized to train a model and it varies based on the type of the input [15]. Feature extraction in NLP has undergone significant evolution, revolutionizing the field and enabling more effective language processing tasks. In the early days of NLP, feature extraction relied heavily on handcrafted linguistic features, including n-grams, Bag of Words (BOW) representations, syntactic structures, and domain-specific features. Additionally, statistical techniques such as TF-IDF (Term Frequency-Inverse Document Frequency) were commonly used either alone or in combination with other features. TF-IDF, as a standalone, allowed capturing the importance of terms within a document or corpus based on their frequency and rarity. However, while these feature extraction methods were valuable, they had inherent limitations in fully capturing the complexity and context of languages [18]. On the other hand, with the advancement of DL, feature extraction in NLP has resulted in a paradigm shift. Instead of solely relying on TF-IDF and manual feature engineering, DL models can learn meaningful representations directly from raw text data. This has led to the development of word embeddings, such as Word2Vec, Global Vectors for Word Representation (GloVe), and FastText, which capture semantic relationships and contextual information of words in vector representations [19]. More recently, pre-trained language models, such as Generative Pre-trained Transformer (GPT), Bidirectional Encoder Representations from Transformers (BERT), and Transformer-based architectures, have further improved the feature extraction in NLP. These models are trained on tremendous volumes of text data and learn contextualized representations, capturing the fine-grained details of language usage. By leveraging transfer learning, these models can be fine-tuned for specific downstream tasks, yielding cutting-edge results across various NLP domains.

This paper proposes a thorough comparison between the four popular pre-trained word embeddings for DL in the Arabic language: Word2Vec, GloVe, FastText, and BERT. These word embeddings have been widely used and recognized in the field of Arabic NLP. We employed the Arabic news articles dataset, SANAD, to evaluate the word embeddings' performance. To the best of our knowledge, this work represents the first attempt to compare these word embeddings specifically in the Arabic language. Furthermore, it is the first of its kind to apply BERT to three subsets of the SANAD dataset. The comparison of these word embeddings will be conducted using a fixed DL architecture across the four methods. By evaluating their performance on the same dataset and utilizing a consistent architecture, we can provide a fair and comprehensive comparison.

The subsequent sections of this paper are organized as follows: In Sect. 2, we explore the related work. Section 3 provides an overview of the research background. Our proposed method is outlined in Sect. 4. Moving forward, Sect. 5 presents the results obtained from our experiments. Finally, in Sect. 6, we summarize the key findings and conclusions drawn from this study.

2 Related Work

Numerous researchers have dedicated their efforts to developing automatic text classification systems using a wide range of classical and cutting-edge methods. However, in this section, we will specifically focus on highlighting the research work conducted using DL on Arabic news articles classification.

Galal et al. [13] put forth a new method for Arabic text classification that combines Convolutional Neural Network (CNN) with a novel algorithm called GStem, designed to group Arabic words of similar roots based on word embedding distances and extra Arabic letters. The authors compiled a corpus of 6,000

Arabic news articles for testing their methodology. Skip-Gram Word2Vec was employed to vectorize the text. The authors conducted two experiments, one with CNN alone and another combining CNN with GStem, the latter showing a better performance with an accuracy of 92.42%. However, the method proposed was solely tested on one dataset. Hence, it is essential to validate the model's efficacy by evaluating it with additional datasets to ascertain its generalizability.

Mishal and Hamad [21] proposed a system for classifying Arabic text documents into distinct categories. The system, implemented with Apache Spark and NLP lib for text processing, required data preprocessing steps. The BOW method was employed for feature extraction, and for feature selection, the researchers identified Term Frequency attributes, using TF-IDF to evaluate each feature's significance within the document. They trained a CNN-based model using different sizes of the five-category SANAD dataset subset, despite there being two other subsets comprising seven categories. Their system demonstrated an accuracy of 96.94% on a 90/10 training/testing split.

Alhawarat and Aseeri [6] presented SATCDM (Superior Arabic Text Categorization Deep Model), a specialized approach for categorizing Arabic news documents. Their model, a combination of multi-kernel CNN architecture and skip-gram word embeddings, leveraged n-gram word embedding techniques to incorporate sub-word information. A pre-trained FastText model on Arabic was utilized to extract the word embeddings. Fifteen publicly available Arabic news document datasets were used for the study. The training process of the model incorporated stratified 5-fold cross-validation, where the data was divided into an 80% training set and a 20% validation set. Consequently, the model achieved an outstanding accuracy varying from 97.58% to 99.90%.

Sundus et al. [28] proposed a neural network model for Arabic text classification that utilized a feed-forward architecture. The first layer of the model was constructed using TF-IDF vectors. The researchers conducted experiments on two multi-class Arabic datasets and compared the outcomes to those obtained using the Logistic Regression (LR) algorithm. The first dataset, Khaleej-2004, contained 5,690 Arabic documents divided into four categories, while the second dataset held 1,445 Arabic documents spread over nine categories. The DL model outperformed the LR algorithm in terms of accuracy for both datasets, scoring 93.8% and 94.12%. However, the authors noted that comparing the DL model exclusively with LR was limited, suggesting that a more thorough evaluation could be achieved by comparing it with other ML algorithms.

Elnagar et al. [12] introduced two large datasets tailored for Arabic text classification, named NADiA and SANAD, focusing on multi-label and single-label categorizations, respectively. They evaluated nine DL models on these datasets. The effectiveness of Word2vec embeddings was also assessed. For single-label classification, the attention Gated Recurrent Unit (GRU) model proved superior performance, achieving an accuracy of 96.94% on SANAD. However, for multi-label tasks on NADiA, the CNN model exceeded others with an accuracy of 70.34% for up to 8 categories on the SkyNewsArabia dataset, and 88.68% for up to 10 categories on the Masrawy dataset using the HANGRU model.

Boukile et al. [7] proposed an approach that integrates the Term TF-IDF technique with a CNN model for Arabic text classification. They created a dataset of 319 million Arabic words using a web crawler, which was later separated into training and testing sets. The dataset consisted of 111,728 documents sourced from three Arabic online newspapers. Furthermore, they utilized the TF-IDF technique and a CNN model was subsequently used to classify different data sizes, and a small experiment was conducted to investigate the impact of filter sizes and feature maps. The CNN model's performance was benchmarked against classical ML algorithms. The results showcased the performance of the CNN model is higher with an accuracy rate of 92%.

From the previous work, we observe that not all the popular word embeddings are being used for Arabic news classification. Also, even though transformer-based models are cutting-edge currently, there seems to be a preferential lean toward older methodologies among researchers. Therefore, our paper aims to bridge this knowledge gap by providing a thorough comparative evaluation of four prevalent word embeddings: Word2Vec, GloVe, FastText, and BERT.

3 Research Background

In this section, we will provide a brief background on the various concepts and techniques that form the foundation of our research. The following subsections will explore the fundamental aspects of Word2Vec, GloVe, FastText, and BERT.

3.1 Word2Vec

Word2Vec is a popular model for learning word embeddings, which are dense vector representations of words. Since its introduction in the work of [20], this approach has become widely adopted in diverse tasks including text classification. The Word2Vec model utilizes two techniques, CBOW (Continuous Bag of Words) and the Skip-gram model. These are shallow neural networks that map words to a target variable, another word in this case. They both learn and adapt weights which serve as word vector representations. CBOW predicts the likelihood of a word based on its context, with no regard to the sequence of context words. Conversely, the Skip-gram model predicts the context for a specified word [24]. Word2Vec models are excellent at capturing semantic relationships between words and produce highly multidimensional word embeddings, which capture fine semantic relationships. However, Word2Vec does not account for word order within a sentence and assumes words are independent of each other (i.e., the meaning of a word can be derived from its nearby words). Therefore, it can not capture polysemy (i.e., words with multiple meanings). Also, it does not handle Out-of-Vocabulary (OOV) words [26].

3.2 GloVe

GloVe is another popular word embedding model developed by [22]. It employs an unsupervised learning approach to generate vector representations for individual words. Like Word2Vec, it is a method to generate word embeddings, but it differs by using statistical information from the entire corpus to learn these embeddings. GloVe combines the benefits of two major methods of deriving word vectors: local context window and matrix factorization methods. By training on the non-zero entries of a global word-word co-occurrence matrix, GloVe leverages statistical information to enhance its performance, which arranges the data in tabular form on how frequently words co-occur in a given corpus. This matrix is typically huge and is constructed from the entire corpus. The value stored in each cell of the global word-word co-occurrence matrix indicates the frequency of word i appearing in the context of word j. Once this matrix is generated, GloVe learns word vectors in such a way that the dot product of the vectors is equal to the logarithm of the probability of co-occurrence between the words. This approach allows GloVe embeddings to capture global relationships between words while retaining the ability to handle local context. However, Like Word2Vec, GloVe does not handle OOV words and cannot capture polysemy [26].

3.3 Fasttext

FastText, another word embedding method developed by [17], extends the capabilities of Word2Vec by incorporating subword information. The method breaks down each word into a collection of character n-grams, which are sequences of characters, ranging in size from 1 to the word's length. The inclusion of subword details, like the trigram representation for the word "banana" ["ba", "ban", "ana", "nan", "ana", "an"], allows FastText to infer vector representations for OOV words and typos from their subword components [23]. The word vector embedding for any given term is the sum of all these constituent n-grams. After the training phase of the neural network, word embeddings are available for all n-grams existing in the training dataset. This even includes the representation of rare words, as it is highly probable some of their n-grams also appear in other words. For instance, the prefixes and suffixes. This unique blend of features empowers FastText to perform exceptionally in text classification, where understanding word formation and composition is essential. Therefore, FastText addresses the challenges of capturing morphological similarities and managing OOV words, which Word2Vec and GloVe models were unable to resolve.

3.4 BERT

BERT, a cutting-edge language model introduced in [8], has demonstrated exceptional accuracy across a wide range of NLP tasks. Unlike the previous models, which generate static word embeddings, BERT generates contextual embeddings based on the Transformer and bidirectional architecture, which allows it to understand the context in which a word appears. The Transformer is a particular type of neural network architecture that BERT uses. It implements self-attention mechanisms to comprehend the context and relationships between words in a text sequence. Moreover, BERT's bidirectional approach utilizes the whole corpus in both directions, moving beyond the fixed window method employed by models like Word2Vec and GloVe. This technique captures a word's meaning based on its surrounding context, hence producing highly contextualized word embeddings [2]. Pre-training is a crucial step in the functioning of BERT. It is pre-trained on two tasks: Masked Language Model and Next Sentence Prediction. After pre-training, the model can be fine-tuned by adding a single output layer. This fine-tuning enables the creation of cutting-edge models for various tasks. However, there is a significant trade-off. BERT requires significant computational resources and time for training due to its deep architecture.

4 Methodology

In this section, we present a description of the utilized dataset, the techniques employed for data cleaning and preprocessing, the used word embedding methods, and an explanation of the suggested model.

4.1 Dataset

The Single-label Arabic news articles datasets (SANAD) [9], were collected using web scraping techniques from three popular news websites: akhbarona.com, alkhaleej.ae, and alarabiya.net. These datasets consist of three distinct sets and cover a range of categories including Religion, Tech, Finance, Culture, Sports, Politics, and Medical. Notably, the dataset from alarabiya.net does not contain any Religion or Culture categories. All the articles were written in MSA since they were gathered from news websites. The individual datasets with single-label tags were merged into a unified dataset called SANAD. After that, the merged dataset was split into training and testing sets. To maintain a balanced representation of categories, subsets from the SANAD dataset were used for both training and testing. As a part of the pre-processing phase, the collected articles were cleaned by removing punctuation marks and Latin alphabet characters. However, spelling errors were not addressed during this process. Table 1 illustrates the article count and categories in each dataset.

Table 1. A balanced subset of SANAD articles and categories count per dataset.

Source	Categories	Training	Testing	Total	Per Category
alarabiya.net	5	16,650	1,850	18,500	3,700
alkhaleej.ae	7	40,950	4,550	45,500	6,500
akhbarona.com	7	42,210	4,690	46,900	6,700

4.2 Pre-processing

The role of text pre-processing is to ensure dataset cleanliness and enhance the final results. The initial phase of our approach involves the exclusion of content that is not Arabic, a vital step when dealing with data retrieved from the internet. We also take steps to eradicate all diacritical marks, elongations (for instance, “مـــــــاء” is condensed to “ماء” (meaning “water”)), punctuation elements, and extra spaces. Prevalent methodology in Arabic computational linguistics involves normalizing certain Arabic characters, such as replacing the letter “ة” with “ه”, the characters “أْ”, “إ”, and “آ” with “ا”, and the letter “ي” with “ى”. Nevertheless, we decided not to implement this normalization step, as it can alter the contextual meaning of some words, such as “هنأْ” (meaning “congratulate”) and “هنا” (meaning “here”). A number of researchers utilize a stemming process, which condenses a word to its root form. We decided against this approach, given its potential to likewise impact the context of words.

4.3 Arabic Word Embeddings

Since all four word embeddings can be pre-trained on a large corpus of data, we looked for the most popular pre-trained word embeddings for the Arabic language. In the case of word2vec, the most commonly used is called Aravec [27]. This embedding offers both CBOW and skip-gram variants, trained on either Twitter or Wikipedia data with vector sizes of 100 or 300. We selected the CBOW variant with a 300-dimensional vector that had been trained on Wikipedia data since it is written in MSA, which aligns with our SANAD corpus. Moreover, CBOW has proven superior performance and faster training times when dealing with large datasets. Also, our choice of the 300-vector size was to maintain a consistent comparison with other word embeddings. For GloVe, we found a single pre-trained model available. This model, with a 256-dimensional vector, was trained on an extensive Arabic corpus that included books, Wikipedia, and Twitter data. Regarding FastText, its creators released two models: CBOW and skip-gram, both featuring a 300-dimensional vector. Again, we opted for the CBOW model, due to the reasons mentioned earlier. In addition, the CBOW model was trained on more data, having been published two years after the skip-gram model [14]. As for BERT, there are three prevalent models, MARBERT, ARBERT, and ARABERT. However, we excluded MARBERT as it was trained on both MSA and dialects. ARBERT and ARABERT were both trained solely on MSA. However, based on our experiments, we concluded that ARBERT was the superior choice and thus, selected it. ARBERT shares the same architecture as the original BERT and has 768 vector dimensions. We used ARBERTv2 [11] which is an updated version that is trained on more MSA data.

4.4 Proposed Model Architecture

In this research paper, we utilize a DL model that leverages a combination of layers to process text data. The first layer is an embedding layer that maps the input words to a continuous vector space of a specified dimension. It takes into account the whole vocabulary size and the maximum length of the text sequences for the four embeddings to ensure a fair comparison. Following the embedding layer, a one-dimensional convolutional layer is added with 256 filters and a kernel size of 5, and it uses the Rectified Linear Unit (ReLU) activation function. Next, a bidirectional GRU layer with 128 units is introduced. A dense layer with 256 units and a ReLU activation function follows. Finally, a dense layer is employed with a number of output labels and a sigmoid activation function. The model is compiled with the binary cross-entropy loss function and Adam optimizer with a learning rate of 0.001. To prevent overfitting and improve generalization, two callbacks are employed. The early stopping callback, which halts the training process if there is no improvement for a specific number of epochs, and the model checkpoint callback, which saves the model with the highest validation accuracy during training.

5 Experimental Results and Discussion

For our experiments, we selected accuracy as our measurement metric since our task is a classification problem. Accuracy is calculated as the ratio of correct predictions made by the model to the total number of predictions. Furthermore, we applied the four embeddings Aravec, FastText, GloVe, and ARBERTv2, with the specified model to the three subsets of SANAD. The experimental results showed remarkable differences in performance among the four embeddings when applied to the three Arabic news sources: Akhbarona, Alkhaleej, and Alarabiya.

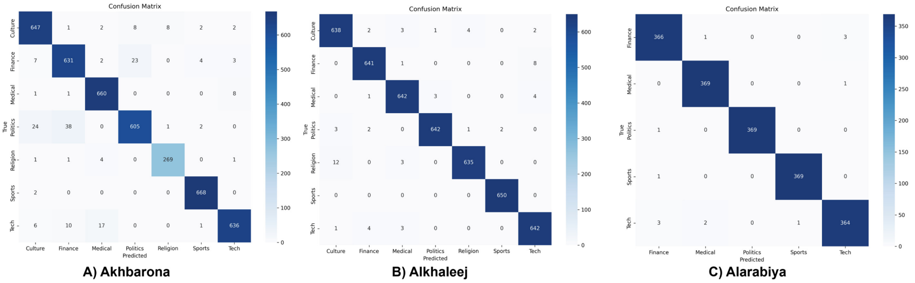
Table 2 illustrates the results of the four embeddings on the datasets. ARBERTv2 emerged as the best performer, achieving an accuracy of 95.81%, 98.68%, and 99.30% on Akhbarona, Alkhaleej, and Alarabiya respectively. This superior performance can be primarily attributed to ARBERTv2's context-aware embeddings. Unlike the other embeddings which generate a fixed representation for each word based on their training data, ARBERTv2 builds an embedding for the words predicated on their context. However, this comes at the cost of complexity, ARBERTv2 has a massive 162 million parameters, significantly more than the others. Among the non-contextualized embeddings, FastText proved to be the most efficient. It scored higher than both GloVe and Aravec, achieving accuracies of 94.27%, 97.65%, and 98.65% on Akhbarona, Alkhaleej, and Alarabiya. Its superior performance could be traced back to its resolution of the OOV issue and its capability to capture morphological similarities. GloVe, despite being trained on more data, did not surpass FastText. However, it outperformed Aravec with accuracies of 94.16%, 97.54%, and 98.43% on Akhbarona, Alkhaleej, and Alarabiya, respectively. Aravec, on the other hand, lagged behind. Being trained solely on Wikipedia data likely limited its performance, which is evident in its accuracy scores of 93.18%, 97.27%, and 98.32% on the previous news sources.

Following the generation of our results, we were curious to delve deeper into the predictions and understand where ARBERTv2 made mistakes. Therefore, We visualized the prediction outcomes by plotting confusion matrices for ARBERTv2 across the three datasets. As shown in Fig. 1, a noticeable overlap emerged within the Akhbarona dataset, specifically between the labels for 'Politics' and 'Finance', as well as between 'Politics' and 'Culture'. The model seemed to confuse these categories, leading to misclassifications. For the Alkhaleej dataset, the model's primary errors occurred between the 'Religion' and 'Culture'

Table 2. The accuracy of the four used word embeddings.

Word Embedding	Akhbarona	Alkhaleej	Alarabiya
Aravec	93.18	97.27	98.32
GloVe	94.16	97.54	98.43
FastText	94.27	97.65	98.65
ARBERTv2	**95.81**	**98.68**	**99.30**

labels. This suggests that the model had difficulty distinguishing between these two classes, potentially due to similarities in their linguistic features. On the other hand, the Alarabiya dataset displayed impressive results with ARBERTv2 since its labels had no overlaps.



A) Akhbarona

B) Alkhaleej

C) Alarabiya

Fig. 1. Confusion matrices for the best models (ARBERTv2).

When we wanted to benchmark our model's performance, we drew a comparison with the research conducted by Elnagar et al. [12]. Table 3 shows the results of both approaches. Their work was the only work that shared the same data split, and they have utilized word2vev trained on diverse sources such as News datasets (SANAD and NADiA), Wikipedia, and books. Despite their use of a well-trained Word2Vec embedding, our ARBERTv2 model consistently outperformed their system, with accuracy improvements of 1.81%, 1.82%, and 1.89% on the Akhbarona, Alkhaleej, and Alarabiya subsets, respectively. These results emphasize the robustness of our ARBERTv2 model and show that its advanced capabilities extend beyond the traditional Word2Vec framework, even when it is trained on extensive and diverse data sources.

Table 3. Comparison between state-of-the-art systems.

Word Embedding	Akhbarona	Alkhaleej	Alarabiya
Elnagar et al. [12] (Word2Vec)	94.00	96.86	97.41
Our Work (ARBERTv2)	**95.81**	**98.68**	**99.30**

6 Conclusion

In conclusion, this study reveals important insights into the performance of four distinct word embeddings, Aravec, FastText, GloVe, and ARBERTv2, when applied to the Arabic news article classification problem. Notably, ARBERTv2 proved superior due to its context-sensitive nature, delivering the best accuracy across all evaluated datasets. However, this improvement in accuracy comes with an increase in model complexity, illustrated by ARBERTv2's 162 million parameters, which may impact its applicability in resource-restricted situations. Among the non-contextualized embeddings, FastText outperforms both GloVe and Aravec. It demonstrates superiority in handling the OOV issue and capturing morphological similarities. GloVe's performance, although not as strong as FastText's, was comparable, considering it was trained on larger data sets. Aravec, on the other hand, showed the lowest performance.

These findings show the importance of considering the characteristics of each embedding along with the specific requirements of the task. Context-aware models like BERT may offer superior performance, but non-contextualized embeddings like FastText can still be very effective, particularly in scenarios where computational resources are limited. In the future, we will shift our focus to hyperparameters optimizations of DL models with minimal training parameters.

References

1. Ababneh, A.H.: Investigating the relevance of Arabic text classification datasets based on supervised learning. J. Electron. Sci. Technol. **20**(2), 100160 (2022)
2. Aftan, S., Shah, H.: A survey on BERT and its applications. In: 2023 20th Learning and Technology Conference (L&T), pp. 161–166. IEEE (2023)
3. Al Qadi, L., El Rifai, H., Obaid, S., Elnagar, A.: Arabic text classification of news articles using classical supervised classifiers. In: 2019 2nd International Conference on New Trends in Computing Sciences (ICTCS), pp. 1–6. IEEE (2019)
4. Alammary, A.S.: Bert models for Arabic text classification: a systematic review. Appl. Sci. **12**(11), 5720 (2022)
5. Alhaj, Y.A., et al.: A novel text classification technique using improved particle swarm optimization: a case study of Arabic language. Future Internet **14**(7), 194 (2022)
6. Alhawarat, M., Aseeri, A.O.: A superior Arabic text categorization deep model (SATCDM). IEEE Access **8**, 24653–24661 (2020)
7. Boukil, S., Biniz, M., El Adnani, F., Cherrat, L., El Moutaouakkil, A.E.: Arabic text classification using deep learning technics. Int. J. Grid Distrib. Comput. **11**(9), 103–114 (2018)
8. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
9. Einea, O., Elnagar, A., Al Debsi, R.: SANAD: single-label Arabic news articles dataset for automatic text categorization. Data Brief **25**, 104076 (2019)
10. El Rifai, H., Al Qadi, L., Elnagar, A.: Arabic text classification: the need for multi-labeling systems. Neural Comput. Appl. **34**(2), 1135–1159 (2022)

11. Elmadany, A., Nagoudi, E.M.B., Abdul-Mageed, M.: ORCA: a challenging benchmark for arabic language understanding. arXiv preprint arXiv:2212.10758 (2022)
12. Elnagar, A., Al-Debsi, R., Einea, O.: Arabic text classification using deep learning models. Inf. Process. Manag. **57**(1), 102121 (2020)
13. Galal, M., Madbouly, M.M., El-Zoghby, A.: Classifying Arabic text using deep learning. J. Theor. Appl. Inf. Technol. **97**(23), 3412–3422 (2019)
14. Grave, E., Bojanowski, P., Gupta, P., Joulin, A., Mikolov, T.: Learning word vectors for 157 languages. arXiv preprint arXiv:1802.06893 (2018)
15. Guyon, I., Elisseeff, A.: An introduction to feature extraction. Feature Extraction: Foundations and Applications, pp. 1–25 (2006)
16. Habash, N.Y.: Introduction to Arabic natural language processing. Synthesis Lectures Hum. Lang. Technol. **3**(1), 1–187 (2010)
17. Joulin, A., Grave, E., Bojanowski, P., Mikolov, T.: Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759 (2016)
18. Liu, Z., Lin, Y., Sun, M., Liu, Z., Lin, Y., Sun, M.: Representation learning and NLP. In: Representation Learning for Natural Language Processing, pp. 1–11 (2020)
19. Liu, Z., Lin, Y., Sun, M., Liu, Z., Lin, Y., Sun, M.: Word representation. In: Representation Learning for Natural Language Processing, pp. 13–41 (2020)
20. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)
21. Mishal, S.M., Hamad, M.M.: Text classification using convolutional neural networks (2022)
22. Pennington, J., Socher, R., Manning, C.D.: Glove: global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543 (2014)
23. Pratiwi, N.I., Budi, I., Alfina, I.: Hate speech detection on Indonesian Instagram comments using FastText approach. In: 2018 International Conference on Advanced Computer Science and Information Systems (ICACSIS), pp. 447–450. IEEE (2018)
24. Rong, X.: Word2vec parameter learning explained. arXiv preprint arXiv:1411.2738 (2014)
25. Salloum, S.A., Mhamdi, C., Al-Emran, M., Shaalan, K.: Analysis and classification of Arabic newspapers' Facebook pages using text mining techniques. Int. J. Inf. Technol. Lang. Stud. **1**(2), 8–17 (2017)
26. Singh, K.N., Dorendro, A., Devi, H.M., Mahanta, A.K.: Analysis of changing trends in textual data representation. In: Santosh, K.C., Gawali, B. (eds.) RTIP2R 2020. CCIS, vol. 1380, pp. 237–251. Springer, Singapore (2021). https://doi.org/10.1007/978-981-16-0507-9_21
27. Soliman, A.B., Eissa, K., El-Beltagy, S.R.: AraVec: a set of Arabic word embedding models for use in Arabic NLP. Procedia Comput. Sci. **117**, 256–265 (2017)
28. Sundus, K., Al-Haj, F., Hammo, B.: A deep learning approach for Arabic text classification. In: 2019 2nd International Conference on New Trends in Computing Sciences (ICTCS), pp. 1–7. IEEE (2019)

Tree Fruit Load Calculation with Image Processing Techniques

Merve Aral(✉), Nada Misk, and Gökhan Silahtaroğlu

Istanbul Medipol University, Istanbul, Turkey
merve.aral@std.medipol.edu.tr, {nada.misk,
gsilahtaroglu}@medipol.edu.tr

Abstract. Turkey holds a significant position in global olive production, with olives being a crucial component of its agricultural industry. The fruit load on trees directly correlates with olive tree yield, which in turn determines productivity. The Tabit Smart Agriculture R&D Center, located in the Koçarlı district of Aydın within Turkey's Aegean region, conducted a study using the YOLOv3 Convolutional Neural Network model to estimate olive tree loads. The primary aim of this research was to offer a more precise and objective perspective on olive harvesting, moving away from subjective assumptions based on predictions. Olive trees, playing a significant role in Turkey's agricultural output, are cultivated across various regions in the country. However, olive sales in Turkey still rely on approximations. To tackle this, image processing techniques were employed to introduce a more technological and practical approach to agricultural applications, particularly in estimating olive tree loads. Throughout the study, real-time datasets were generated by capturing images of olive trees at the Tabit Smart Agriculture R&D Center in Aydın. The focus was on accurately detecting and counting olives. After 6000 iterations, the obtained results were as follows: mAP 61%, Precision 70%, Recall 45%. The results of the study proved that the agricultural industry can actively shape the trajectory of future farming practices by adeptly embracing image processing techniques and deep learning models.

Keywords: YOLOV3 · Olive Trees · Image Processing · Deep Learning

1 Introduction

Agriculture holds great importance both in terms of the economy and trade in Turkey and around the world. The presence of four distinct seasons and Turkey's location in the temperate zone significantly impact the extensive agricultural lands and the diversity of crops grown. According to the Turkish Statistical Institute (TUIK) [1], 55.9% of Turkey's land is situated at an elevation of more than 1000 m. Areas with a slope of over 15% constitute 62% of the country's land. Class 1, 2, and 3 high-quality soils collectively account for 24.5% of the total, with 90% of these soils being designated as agricultural lands. Among the most cultivated fruits in Turkey, grapes, apples, and oranges, olive holds the third position. According to the data, Turkey boasts substantial agricultural potential in terms of product diversity and production. However, a lack

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 137–147, 2024.
https://doi.org/10.1007/978-3-031-56728-5_12

of adequate planning in the stages of production, coupled with challenges in adopting agricultural practices to technological advancements, has led to a situation where, despite increased production quantities, there is a failure to satisfy market demands and attain high-quality and dependable product outcomes. Furthermore, producers struggle to precisely determine the quality and quantity of their products post-harvest, relying solely on rough estimations. According to Yaşar [2], the fruit-growing sector in Turkey has demonstrated consistent production growth over the years. The fruits produced in Turkey also hold significant importance in international trade. Through the utilization of advanced techniques in fruit cultivation, classification, preservation, logistics, and marketing, Turkey has achieved comparable standards to those of many developed countries. As highlighted by Varjovi [3], the actual value and quantity of the yield obtained vary significantly due to shifting seasonal conditions and the variations in knowledge and capabilities among producers. According to Malaslı and Ağnı [4], in the conducted study, a new approach with a computer-based perspective is planned to be introduced to the traditional practice of selling unpicked fruit trees in bulk. This approach emphasizes the significance of the agriculture and fruit-growing sectors in our country.

Building upon this issue, a mathematical and computer-based approach will be implemented at the R&D center of Tabit Smart Agriculture Company, situated in the village of Koçarlı in Aydın, one of the provinces with significant olive production. This approach will employ image processing techniques and artificial neural networks to determine the fruit load of olive trees. By precisely evaluating the fruit load, both producers and buyers can make well-informed decisions regarding the optimal timing for selling products, capitalizing on market conditions and maximizing their benefits. This technology aims to address the existing challenges between buyers and sellers concerning fruit maturity, enabling a more efficient and profitable trade.

This study will focus on olives, which are the third most produced fruit in Turkey. The research will involve the detection and counting of olive fruits using computer-based methods by leveraging their shape, color, and areas. Turkey, ranking fourth in olive production, is continuously expanding its olive cultivation areas.

Table 1 presents information about the countries with the highest olive production, their four-year averages, production quantities (in thousand tons) across different years, and the corresponding percentages of global production. According to this table, Turkey holds the fourth position in terms of olive production.

Table 1. Olive production quantity in Turkey by country [14].

	2011	2012	2013	2014	4-year average	Olive production in the world
Spain	7.820	3.849	9.250	6.374	6.374	33.7
Italy	3.182	3.018	2.941	2.776	2.776	14.7
Greek	1.874	2.825	1.918	2.225	2.225	11.7
Türkiye	1.750	1.820	1.676	1.754	1.754	9.3
World	20.415	17.654	22.040	18.907	18.907	100

Olive has a wide distribution area in Turkey. Olive production is carried out in 41 out of 81 provinces and 270 out of 843 districts in the country. According to TUIK [5], table olive production in Turkey is predominantly centered in the Aydın province. As depicted in Fig. 1, our data analysis will be conducted in Aydın, which is the province with the highest olive production.

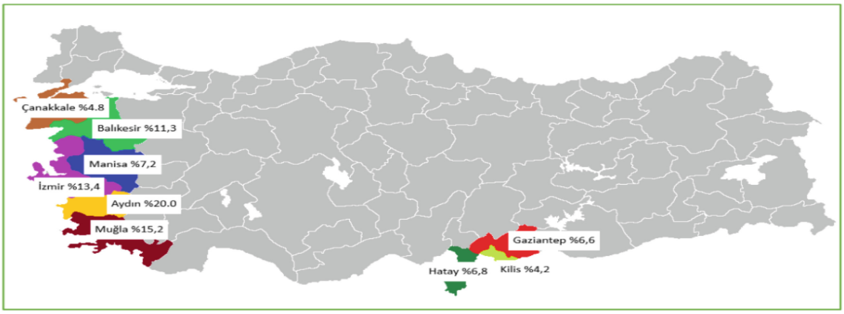


Fig. 1. Provinces with the highest olive fruit production [14]

As shown in Fig. 2, the Aydın province demonstrates highly productive olive trees in terms of yield per tree. Considering all of this data, a dataset comprising olive photos gathered from pertinent olive trees will be compiled at the Tabit Smart Agriculture R&D Center, situated in Aydın, the region with the most prolific olive production.

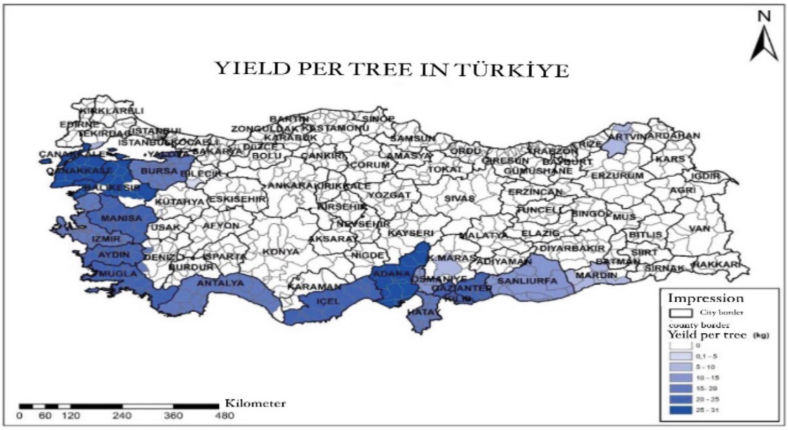


Fig. 2. Provinces with the highest harvest per tree [14]

2 Literature Review

According to Kurtulmuş and Vardar [6], it's feasible to estimate yield based on the information of peach quantity extracted from images and the median size of peach fruits. Determining the color of mature peaches from images presents a formidable challenge.

The color of unripe peaches is nearly indistinguishable from the hue of leaves. Moreover, the irregular lighting conditions in images captured under field sunlight hinder accurate fruit recognition. This study centers on identifying newly formed peach fruits on tree branches. An RGB-based segmentation method has been employed to differentiate peaches from the background. The recommended visual scanning technique entails horizontally and vertically scanning the image with a window. This technique aims to prevent the misclassification of similar peaches as a single fruit by introducing a step distance between them. In terms of comparing the performance of various classifiers based on the suggested initial visual scanning method, the Artificial Neural Network (ANN) algorithm demonstrated the highest success rate at 77%. When considering another suggested scanning method, the ANN algorithm again exhibited the highest success rate at 81%. The application encountered challenges due to consistent illumination conditions, resulting in the inability to detect most small peaches. It's promising to observe the success rate of this application carried out on images captured with a CCD camera.

In the study carried out by Linker [7] to determine the count of green apples in an orchard, various threshold values and radius intervals were assessed for region segmentation. The study confirmed the effectiveness of these parameters in counting apples accurately. The results indicated that a radius length of 36 and a threshold value of 0.25 were more successful in determining the apple count.

According to Pourreza et al. [8], they utilized image processing techniques, including the implementation of a threshold, on grain images of nine distinct wheat varieties cultivated in the region of Iran.

Mustafa et al. [9] conducted their developed method using a Matlab program. The defined features in the classification provide information about the ripeness of bananas. The authors demonstrate that the average ripeness of bananas can be calculated by determining the category to which the pixel values in the image belong. In this study, the authors employed the Canny edge detection method.

Yasar, G.H. [2], cycle is among the traditional image processing applications used in the agricultural sector. The cultivated fruits are usually sold in bulk and roughly presented for sale without being weighed. In this case, an engineering approach is aimed to include an application that balances both sides by aiming to eliminate distribution imbalance between buyers and sellers. Images in the RGB domain are converted to the HSV color space. In the second stage, the histogram of each color channel is separated into the HSV color space, resulting in distance. These computed histograms are represented on a graph that contains information about the color distribution of the image. At the third point, the pixel information obtained from the histogram is applied using the Otsu thresholding method. The green leaf area and the orange area are enhanced and added to the artificial neural network archives. In the fourth stage, noise removal and morphological operations are performed on the obtained binary image to perform image masking. Through this progressive process, fruit area, surroundings, average weight center, and quantity information are obtained from the image and added to a document for use with an artificial neural network approach. Finally, the estimated weight of the fruit is calculated based on the fruit characteristics and the time-dependent leaf area.

Galan et al. [15] asserted that the primary factors influencing crop yield and olive production include the number of flowers, particularly variable climatic factors, agronomic factors, and soil characteristics. The researchers emphasized that the flowering period, which has an impact on olive yield, can initially vary significantly depending on rainfall and temperature.

3 Materials and Methods

According to Fang et al. [11], YOLOv3 (You Only Look Once v3) is an object detection and classification algorithm. This algorithm aims to analyze an image or video in a single process to identify objects in the image and determine their classes. YOLOv3 operates using deep learning and convolutional neural networks. The relevant image or video is fed to a pre-trained YOLOv3 model.

The model divides the input image into multiple small parts and processes each part through the network. The network then classifies these segments and detects objects, predicting their position, size, and class to produce an output.In comparison to other object detection algorithms, YOLOv3 is faster as it detects all objects simultaneously in a single pass. As noted by Redmon and Faradi [12], it also excels in effectively detecting small objects and accurately identifying multiple instances of the same object within an image. YOLOv3 represents an enhanced version of the YOLO model.

3.1 YOLOV3 Algorithm

YOLOv3 Darknet53 serves as the foundational feature extraction backbone for the YOLOv3 model. Feature Maps: The acquired feature maps from Darknet53 are utilized for object detection, designed to capture objects at diverse scales. Lower-scale feature maps concentrate on detecting larger objects, while higher-scale feature maps target smaller ones.

According to Altınörs and Çelik [13], YOLOv3 Head Part: Further processing of the feature maps from Darknet53 occurs in the YOLOv3 head section. In this segment, the feature maps undergo additional processing to produce predictions for object classes and bounding boxes. These predictions encompass class labels, object positions, and confidence scores.

Output Results: The output of YOLOv3 Darknet53 encompasses class labels, positions, and confidence scores for the detected objects. Each object is assigned a class label, a bounding box, and a confidence score. The confidence score signifies the accuracy of object detection and is filtered based on a predefined threshold. By employing YOLOv3 Darknet53, the process of object detection is streamlined, enabling the extraction of information about objects within the images. Subsequent layers follow as presented in Table 2.

The removal of these layers allows Darknet-53 to focus solely on extracting meaningful features from the input data without performing additional operations like pooling or classification. This streamlined version of Darknet-53 can then be used effectively as a feature extraction backbone in the YOLOv3 model shown in Fig. 3.

Table 2. Darknet53 layers [11]

	Type	Filters	Size	Output
	Image			416x416
	Convolutional	32	3x3/1	416x416
	Convolutional	64	3x3/2	208x208
1x	Convolutional	32	1x1/1	208x208
	Convolutional	64	3x3/1	208x208
	Residual			208x208
	Convolutional	128	3x3/2	104x104
2x	Convolutional	64	1x1/1	104x104
	Convolutional	128	3x3/1	104x104
	Residual			104x104
	Convolutional	256	3x3/2	52x52
8x	Convolutional	128	1x1/1	52x52
	Convolutional	256	3x3/1	52x52
	Residual			52x52
	Convolutional	512	3x3/2	26x26
8x	Convolutional	256	1x1/1	26x26
	Convolutional	512	3x3/1	26x26
	Residual			26x26
	Convolutional	1024	3x3/2	13x13
4x	Convolutional	512	1x1/1	13x13
	Convolutional	1024	3x3/1	13x13
	Residual			13x13

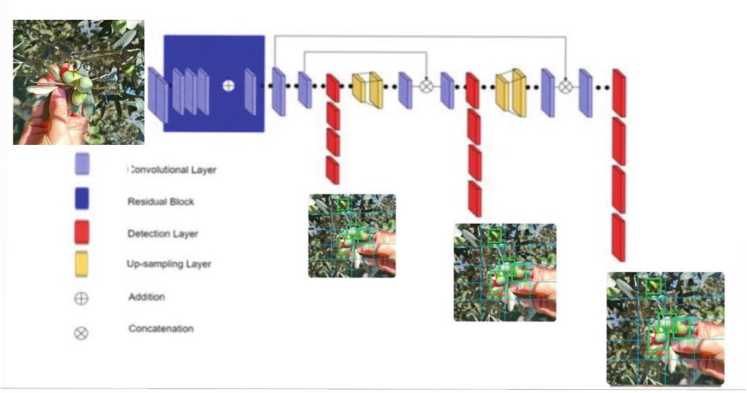


Fig. 3. YOLOv3 architecture

According to Tonguç and Balcı [14], YOLOv3 allows the model to identify suitable fitment boxes using k-means clustering on the training set, instead of manually selecting the fit boxes. The network estimates the width (tw) and height (th) of the box from the center points obtained from the clustering process. Also, the network estimates the center coordinates (tx, ty) of the settlement based on what is written by the sigmoid function. Thanks to these estimated offsets and coordinates YOLOv3 can accurately position and classify units in the image.

YOLOv3 uses predetermined insert boxes to detect which have different scales and shapes. These boxes are defined bounding boxes used to detect units in their various parts and aspect ratios. Rather than manually selecting the insertion boxes, YOLOv3 implements k-means clustering insertion to aggregate similar real constraints in the training dataset.

After the insertion definitions, YOLOv3 estimates the width and height of the bounding box as offsets of the center points obtained from the clustering process. This means estimating the bounding box based on the clusters identified during training the network. The network also estimates the center coordinates of the bounding box based on what is written by the sigmoid function.

It can accurately position and classify offsets and coordinates estimated using YOLOv3 in the image. The predicted offsets allow the nest placements to better fit the application in the image, while the predicted coordinates are the boxes, allowing to detect their centers with precision. Combined with classification predictions, YOLOv3 can effectively detect and classify multiple objects in real time.

Overall, it allows YOLOv3 to use k-means clustering for insert box selection and estimate offsets and coordinates, effectively and accurately identify and target object in images shown in Fig. 4.

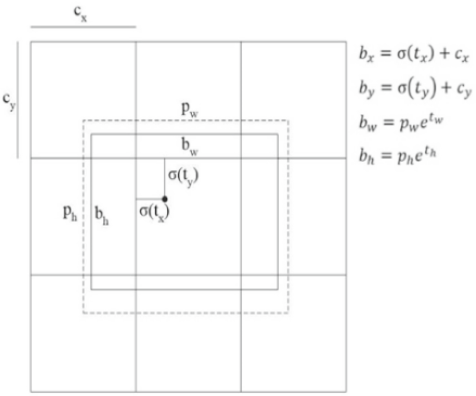


Fig. 4. Prediction of a box by an anchor [11]

σ(tx.) and σ(ty) represent the centroid. Locations relative to the grid cell, which are normalized using the sigmoid function to values. Between 0 and 1. The centroid of the predicted box (bx, by) is calculated by adding the absolute location of the top left corner of the grid cell (cx, cy.) to these σ(tx) and σ(ty.) values. The size of the predicted box (bw, bh) is determined by the values pw and ph, which indicate the predefined size of the anchor box. To ensure that the tw and th values are positive, etw and eth values are used. These statements explain how the centroid, size, and coordinates of the bounding box are calculated in YOLOv3.

YOLOv3 Non-max Suppression (NMS) is a method used in YOLOv3 to organize object detection results and eliminate duplicate object predictions. During object detection with YOLOv3, multiple bounding box predictions are generated for each object. However, in some cases, multiple predictions may be obtained for the same object, leading to complex results. Non-max Suppression is used to address this issue. Non-max Suppression is an algorithm that involves the following steps:

Sorting Object Predictions: Firstly, object predictions are sorted based on confidence scores above a certain threshold. This sorting provides a ranking of predictions based on their confidence levels. Selecting the Highest Scoring Prediction: Once the sorting is done, the prediction with the highest confidence score is selected and added to the result list. Checking Other Predictions: The remaining predictions are compared to the selected highest scoring prediction based on an overlap criterion. If the overlap ratio between two predictions exceeds a certain threshold, indicating that they represent the same object, the prediction with the lower confidence score is eliminated.

Repeating the Steps: The same process is repeated on the remaining predictions. The prediction with the highest confidence score is selected and compared to other predictions. Predictions with overlap ratios above the threshold are eliminated.

Through these steps, the most reliable object predictions are selected, and redundant predictions are removed. This results in cleaner and more organized object detection outcomes shown in Fig. 6.

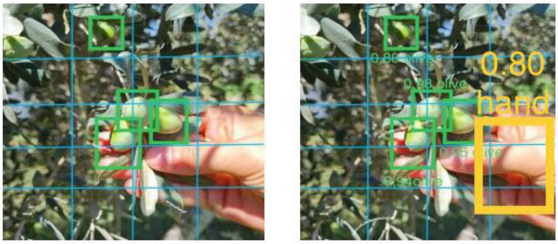


Fig. 6. Non-Maximal Suppression

3.2 Preparing New Image Dataset

The following steps were followed when creating a dataset for YOLOv3: Data Collection: A compilation of images featuring olive tree objects was gathered from the Tabit Smart Agriculture R&D center in the Koçarlı district of Aydın province. Special attention was given to ensuring that the images represented various perspectives, lighting conditions, and backgrounds. Image Labeling: The objects within the collected images were labeled. This process encompassed identifying the object class and establishing the coordinates of the bounding box. This information plays a crucial role in enabling the YOLOv3 model to accurately identify objects. The label "Olive" was allocated to the labeled objects, and multiple olives were labeled within the same photograph. Data Split: The generated dataset was divided into training, validation, and test sets. The training set comprises the majority of the images and is employed to train the YOLOv3 model. The validation set is utilized to assess the model's performance during training, typically encompassing distinct images from the training set. The test set is employed to evaluate the overall performance of the trained model and comprises new images. Data Augmentation: Data augmentation techniques were employed to expand the training dataset's size and enhance the model's capacity for generalization. This encompassed altering the images through operations like rotation, resizing, cropping, flipping, etc., to generate new instances. A total of 652 images were utilized for the training dataset, while 93 images were allocated for validation, and 47 images for testing, each possessing a size of 416 × 416 pixels.

4 Result and Discussion

In Fig. 7 below, we observe that after 6000 iterations, the YOLOv3 model achieves a mean average precision (mAP) rate of 61%, a precision rate of 70%, and a recall rate of 55%. These metrics, mAP, precision, and recall are standard measures for evaluating the YOLOv3 model's performance. They provide insights into its accuracy, precision, and recall characteristics, with higher values indicating better overall performance and success.

Figure 7 illustrates the graphical representations generated during the YOLOv3 model training process. Throughout this training, various metrics were recorded and visualized as graphs. These metrics encompass training loss, validation loss, accuracy, precision, recall, and other evaluation criteria. Figure 8 shows the result of training on an image of the prepared data set.

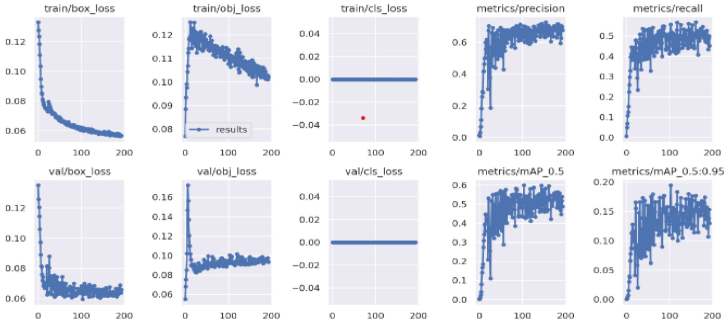


Fig. 7. Training graphs

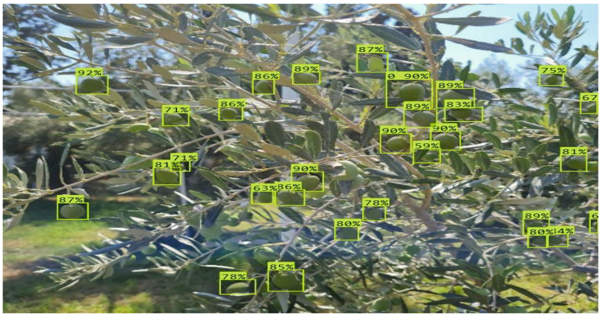


Fig. 8. Training result

5 Conclusion and Future Work

In this study, the Yolov3 model was applied to the olive fruit, which is extensively produced in Turkey, and based on the test results, it was observed that better results were obtained in less complex and close-up photographs. It was determined that the detection becomes more challenging depending on the reflection of sunlight and the distance. It was suggested to make changes in some HSV values to improve the detections. However, the model achieved high accuracy in general shots. The number of olives could easily be obtained from the positions of olives in the images. It was noted that the accuracy of olive count varies depending on the sharpness and distance of the photographs. It was expressed that the olive counts were easily obtained in general shots. Image processing techniques hold substantial potential to mold the future of the agricultural sector and bolster sustainable food production. When applied across diverse domains such as plant health, yield prediction, pest management, and quality assurance, these techniques can play a vital role in enhancing agricultural efficiency, promoting environmental friendliness, and fostering sustainability. As a result, the agricultural industry can actively shape the trajectory of future farming practices by embracing image processing techniques adeptly.

6 References

1. TUİK. https://data.tuik.gov.tr/Kategori/GetKategori?p=Tarim-111. Accessed 19 May 2023
2. Yaşar, G.H.: Calculation of tree fruit load using image processing. Master's thesis, Graduate School of Konya Technical University (2019)
3. Varjovi, M.H., Talu, F.M.: Automatic Harvest Estimation System for Apricot. Inönü University, Department of Computer Engineering, Malatya, Turkey (2016)
4. Malaslı, A., Ağnı, O.: The Place and Importance of Image Processing Techniques in Sustainable Agriculture (2018)
5. TUİK. https://data.tuik.gov.tr/Bulten/Index?p=Bitkisel-Uretim-Istatistikleri-2021-37249. Accessed 10 May 2023
6. Kurtulmuş, F., Vardar, A., Kavdır, L.: Detection of young peach fruits in color images taken under orchard conditions. Using texture and shape features. J. Agric. Mach. Sci. (2013)
7. Linker, C.R., Cohen, O., Naor, A.: Determination of the number of green apples in RGB images recorded in orchards. Comput. Electron. Agric. **81**, 45–57 (2012)

8. Pourreza, K.A., Pourreza, H., Fard, M.H.A., Sadrnia, H.: Identification of nine Iranian wheat seed varieties by textural analysis with image processing. Comput. Electron. Agric. **83**, 102–108 (2012)
9. Mustafa, N., et al.: Determination of size and ripeness of a banana. In: Proceedings of the Information Technology Conference, Kuala Lumpur, Malaysia, 26–28 August 2008 (2008)
10. Fang, W., Wang, L., Ren, P.: Tinier-YOLO: a real-time object detection method for constrained environments. IEEE Access 1 (2020)
11. Redmon, J., Farhadi, A.: YOLOv3: an incremental improvement arXiv:1804.02767 (2018)
12. Altınörs, A., Çelik, S.: YOLOv3 Derin Öğrenme Algoritması ile İHA Görüntülerinden Çevresel Atık Tespiti. Int. J. Innov. Eng. Appl. **7**(1), 76–85 (2023). https://doi.org/10.46460/ijiea.1195428
13. Tonguç, G., Balcı, B.A., Arslan, M.N.: Su Ürünleri Yetiştiriciliği İçin Balık Davranışlarının Bilgisayarlı Görüntü İşleme Yöntemleriyle İzlenmesi. J. Anatolian Environ. Animal Sci. **7**(4), 568–581 (2022). https://doi.org/10.35229/jaes.1197703
14. Ministry of agriculture and forestry. (https://www.tarimoman.gov.tr/BUGEM/kumelenme/Belgeler/Budama/Du%CC%88nyada%20ve%20Tu%CC%88rkiye%27de%20Zeytinc. Accessed 11 May 2023
15. Galan, C., Carinanos, P., Garcia-Mozo, H., Alcazar, P., Dominquez, E.: Model for forecasting Olea europaea L. airborne pollen in southwest Andalucia, Spain. Int. J. Biometeorol. **45**, 59–63 (2001)

Prediction and Analysis of Water Quality Using Machine Learning Techniques

Reshmy Krishnan[1](✉), A. Stephen Sagayaraj[2], S. Elango[2], R. Kaviya Nachiyar[2],
T. Indhuja[2], J. Kanishma[2], A. Mohamed Uvaise[2], and G. Kalaiarasi[3]

[1] Muscat College, Muscat, Sultanate of Oman
reshmy@muscatcollege.edu.om

[2] Department of ECE, Bannari Amman Institute of Technology, Sathyamangalam, Tamil Nadu, India

[3] Vignan's Foundation for Science, Guntur, Andra Pradesh, India

Abstract. Water in daily life is an imperative and needful resource to human beings and all other living organisms. Water contamination by adding pollutants leads to noxious effects and is the new way of introducing disease to people. Thus, machine learning plays a vital role in Predicting whether the water is in excellent or imperfect condition. The machine learning algorithm is used to predict the water quality, and statistical accuracy parameters are calculated for each machine learning algorithm. SVM, KNN, Decision Tree classifiers, XG Boost, Ada boost and Feed Forward Neural Networks are used to predict water conditions. Compared to Conventional Machine learning algorithms, Feed Forward Neural network predicts the water quality with 98% accuracy. Precision values and more statistical parameters are measured and compared by evaluating the accuracy values.

Keywords: Water Quality · Prediction · machine learning · feed-forward neural network

1 Introduction

Rivers, lakes, and streams are only examples of ambient water bodies with precise quality requirements. Additionally, water criteria for different uses and applications have their norms. For instance, irrigation water shouldn't be overly salty or include substances that can harm plants or the soil, damaging ecosystems. Depending on the industrial processes, water quality for industrial usage also requires distinct attributes [1]. Several investigations were carried out to establish the environmental aspects of the lake that highlighted physiological traits. A visual representation of the received on June 16, 2018, and amended on July 1, 2019. Using the Environmental Fluid Code, calculate the lake's temperature. The lake is thermally unratified, as evidenced by (EFDC), with an average depth of 3.21 m and a mean between 25 and 29 degrees Celsius [2]. When developing agricultural projects, determining the type of irrigation system, cropping pattern, and industrial water purification systems, evaluating water quality is a fundamental step [3]. Water is the most crucial resource because all forms of life must exist.

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 148–158, 2024.
https://doi.org/10.1007/978-3-031-56728-5_13

However, life itself constantly poses a threat to its purity. One of the most far-reaching communication media is water. As a result of rapid industrialization, the water quality is alarmingly declining. Poor water quality is one of the leading causes of the spread of deadly diseases. According to reports, 2.5 billion illnesses and 5 million fatalities have been attributed to water-borne diseases, which account for 80% of infections in developing nations [4]. The degree of water contamination can be precisely assessed by real-time control of upcoming changes in water quality. Also, precise estimates of water quality can serve as a foundation for policymakers and send information to the environmental management department to serve as an "early warning" [5]. The horrendous results of water contamination require a faster and less expensive other option [6]. The execution of artificial reasoning (simulated intelligence) prompts an adaptable numerical construction that can distinguish non-direct and complex connections between information and result information. There has been a significant debasement of the Johor Waterway Bowl in light of a few formative and human exercises.

Consequently, setting up a water quality expectation model for better water assets on the board is of fundamental significance and will act as an incredible asset [7]. The general populace needs access to safe, convenient water for drinking, domestic use, food production, and recreational pursuits. Better resource management and water availability may accelerate a country's economic growth. Everyone has the right to obtain sufficient water for personal and home use, and it is always available, safe, and competitively priced. Many people develop cancer, kidney failure, and other diseases yearly due to contaminated water [8]. Diarrhea, typhoid, gastroenteritis, cryptosporidium infections, etc., are some forms of hepatitis and giardiasis intestinal worms in Pakistan [6]. Models for machine learning that can effectively express ideas about the hydrological but necessitate a substantial amount of data. Even though in these procedures, results are produced without a close examination of the physical aspects of the water resource under investigation [9]. In most developed nations, treated wastewater discharge must be tested and monitored to ensure discharge permits are satisfied to safeguard the environment and human health. In the past, scientists had to gather and examine many wastewater samples to comprehend how the environmental effects of wastewater discharge components were caused [10]. Because of the growing population, there will be a severe water shortage sooner rather than later, and this will be primarily due to the amount of water used each day for domestic, industrial, and agricultural activities. We must start practising using various water sources, including water [11]. Humans and other living things depend entirely on clean water as a resource.

Consequently, there is tremendous societal and economic value in developing a water quality prediction model to forecast future water quality conditions [12]. Artificial intelligence (AI) has been used to predict water quality and produce accurate findings. The use of AI for modelling raw water quality for treatment purposes has not yet received much attention [13]. The other categories of learning algorithms include the decision tree and its numerous variations; regions are created by dividing the input space, and each area has its parameters [14].

The accuracy is measured by how many of the observed values it correctly predicted, where T Prefers to True Positive, TN Refers to True Negative, FP Refers to False Positive, and FN Refers to False Positive False Negative, is used to calculate accuracy. Ph, Temperature Dissolved Oxygen, Turbidity, chloramine, sulfate, conductivity, organic carbon, and trihalomethanes were the nine parameters used in this study. Many systems are available for water quality classification, but they lack accuracy, as a piece of computer-based intelligence has been generally applied in different fields, particularly in hydrology. As per Moubayed, there are four kinds of AI calculations: support learning, solo learning, semi-directed learning, and regulated learning. Moreover, a regulated learning calculation is known when an AI calculation gets the objective example and the component vector as a contribution to foster a model. The created model can be applied to decide the most recent standards and set results for the model [3]. By making knowledge summaries and discoveries and predicting system behaviour, they have improved the performance of water quality prediction [15].

The AI calculation has been generally utilized in water-related fields, for example, water assets, water the board, hydrology, environmental science, water quality, water level expectation, weather conditions anticipating, water release forecast, water quality gauging, etc. In any case, water quality expectations concentrated on in light of the AI calculation are restricted contrasted with other water-related applications given the restricted water quality information. The more significant part of the past water quality forecast investigations has anticipated month-to-month water quality, which is helpful data but insufficient from a pragmatic perspective [16]. The water of lousy quality can likewise be monetarily tested, considering that assets should be redirected to overhaul the water conveyance foundation whenever an issue emerges. For these reasons, the interest in further developed water treatment and quality control has been expanding to guarantee clean drinking water at reasonable rates. Deliberate investigations of crude water, removal frameworks, and authoritative checking issues are expected to determine these difficulties [17]. Thus, it is to have an automated system that can automatically classify water quality with less effort. This research aims to develop efficient models to predict values of water quality parameters based on their present values. The k-nearest neighbours (KNN) algorithm is a straightforward, supervised machine-learning technique that may be applied to classification and regression issues. Using the distance capability, the K-Nearest Neighbor (KNN) model, which is non-parametric, aims to store all applicable examples and predict the outcomes by calculating the average of the goals of the closest neighbours (k) (k worth is essential in determining the k-NN model execution).

Nonetheless, this study embraces five as k-esteem and Euclidean as a distance capability [18]. A support vector machine (SVM) is a machine learning algorithm that analyzes data for classification and regression analysis. A decision tree is a flowchart-like structure in which each internal node represents a test on a feature, and each leaf node represents a class [19].

SVM, a delegate factual learning calculation, can lay out a straightly isolated hyperplane for information order. SVM is highly hearty to overfitting [20]. Ada Boost, called Versatile Helping, is a strategy in AI utilized as a Group Technique. The most well-known calculation used with AdaBoost is choice trees with one level, which implies Choice trees with just one split. These trees are likewise called Choice Stumps (ADA). XGBoost, which represents Outrageous Inclination Helping, is a versatile, dispersed slope-supported choice tree (GBDT) AI library. It gives equal tree helping and is the leading AI library for relapse, characterization, and positioning issues. The data is acquired from the dataset and classified using different machine learning algorithms. According to a review of the literature, only a small number of authors have developed methods for forecasting the quality of water using experiments. However, manual processes require much time and effort [21]. The Proposed flow is shown in Fig. 1, discussed in the following chapters.

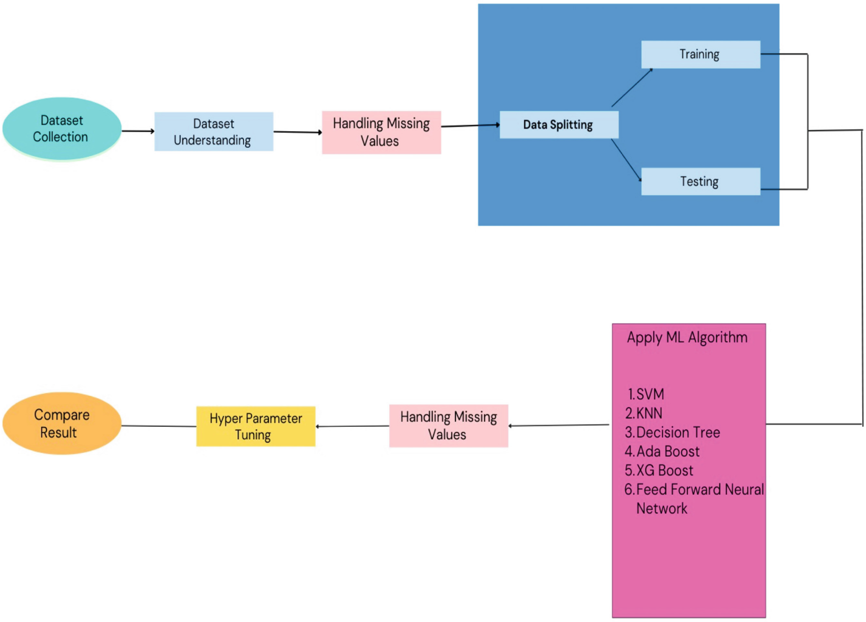


Fig. 1. Flowchart of Proposed Methodology

2 Materials and Methods

The Predefined water quality Dataset is considered for prediction, and the classification of water quality is done by using KNN, SVM, Decision Tree, XG boosting classifier in ADA boost classifier is made by measuring the statistical parameters.

2.1 Water Quality Dataset

The dataset used in this study is taken from the water quality drinking water portability. The dataset contains nine parameters: Ph., temperature, dissolved oxygen, turbidity, chloramine, sulfate, conductivity, organic carbon, and trihalomethanes. The Drinking water portability dataset has nine features and 9300 instances. Portability refers to whether drinking water is safe for the people or not. The Portability is with values ‘0’, which represents the water in Bad State and ‘1’ means the water in Good State, where a total of 5650-0 instances and 3650-1 instances.

2.2 Parameters of Water Quality Analysis

The water quality can be analyzed by implementing different parameters.

2.2.1 PH Value

PH is the main parameter to evaluate the acid-base balance of water. The maximum PH allowed range, according to WHO, is between 6.7 and 8.5. In line with WHO criteria, the current investigation range was 6.52 to 6.83.

2.2.2 Hardness

Hardness refers to the amount of dissolved minerals in the water, usually calcium and magnesium, which can affect its suitability for various purposes.

2.2.3 TDS - Total Dissolved Solids

TDS refers to the total amount of inorganic and organic substances dissolved in the water with high TDS value indicating that water is highly mineralized, 500 mg/l to 1000 mg/l prescribed for drinking.

2.2.4 Turbidity

Turbidity is a measure of the degree to which water loses its transparency due to the presence of suspended particles. These particles include clay, silt, organic matter, algae, and other microscopic organisms. Turbidity is an important water quality parameter, as high levels of turbidity can indicate the presence of contaminants, reduce the effectiveness of disinfection, and impact the health of aquatic ecosystems. The test is used to determine the quality of waste discharge about colloidal matter and measures the light-emitting capabilities of water. The Wondo Genet Campus’s mean turbidity value (0.98 NTU) is less than the WHO-recommended threshold of 5.00 NTU.

2.2.5 Chloramines

Chloramines are chemical compounds formed by the reaction of chlorine and ammonia. They are commonly used as disinfection in water treatment to kill harmful bacteria and viruses. It levels up to 4 mg per litre or is considered safe in drinking water.

2.2.6 Sulfate

Including groundwater and surface water, sulfates are frequently found in natural water sources. Gypsum, anhydrite, and other sulfate-containing minerals are often responsible for their introduction into water sources. Around 2,700 mg/L of sulfate are present in saltwater. Most freshwater sources have values between 3 and 30 mg/L, while certain regions have substantially higher levels (1000 mg/L).

2.2.7 Conductivity

Water conductivity is a crucial metric in many applications, including water quality testing, environmental monitoring, and industrial operations. Clean water acts more as an insulator than a conductor of electrical current. An increase improves the electrical conductivity of water in ion concentration. The amount of dissolved particles present typically determines the electrical conductivity of water. The ability of a solution to convey current through its ionic process is measured by electrical conductivity (EC). According to WHO guidelines, the EC value shouldn't be more than 400 S/cm.

2.2.8 Organic Carbon

Organic carbon in water refers to carbon-containing compounds from living organisms or their byproducts, such as decaying plant matter, sewage, or animal waste. Organic carbon in water can significantly impact water quality, ecosystem, and human health. The decomposing natural organic matter (NOM) and synthetic sources contribute to the source water's total organic carbon (TOC). The total amount of carbon (TOC) in organic compounds in pure water is a measurement of this. The US EPA estimates that treated drinking water has two mg/L of TOC and that source water, which is used for treatment, contains 4 mg/Lit.

2.2.9 Trihalomethanes

A class of compounds known as trihalomethanes (THMs) can develop when chlorine or other disinfectants treat drinking water containing organic waste. THM levels in drinking water vary depending on the amount of organic matter present, the amount of chlorine needed to treat the water, and the temperature of the treated water. THM concentrations in drinking water up to 80 ppm are regarded as safe.

3 The Proposed Methodology

The parameters of water quality are archived and analyzed using different machine-learning techniques

3.1 SVM

SVM is the most favoured supervised algorithm introduced by Vapnik and Chervonicks because of its fast response time and Higher accuracy. The SVM Algorithm also works in small samples as it works in both linear and nonlinear. It introduces the Hyperplane by maximal margin for easy mapping of inputs with feature space. The models exist within reach of the Hyperplane called the Support Vector. The Kernel Function helps the training samples to match with Higher dimensional space.

3.2 KNN

The KNN Algorithm works to classify the fields. The non-parametric areas help to determine the matched fields based on the neighbour values. It is used in applying

Pattern recognition and Performance analysis for classification. It is a slow learning process in which it estimates the function first and finishes the complete form of the type at the end.

The KNN Algorithm works in the following manner,

i. Calculation of Euclidean Distance d (r, RI) in which i = 1, 2, …… in the middle of points
ii. Arrange the n points in an ascending scale
iii. Examine the value k as positive and get hold of initial k Distance
iv. By using k Distance, evaluate the k points.
v. If $k > 0$, it matches the first category.
vi. If $k < 0$, the k values match with the following category

The value of k depends upon those in need by the deciders. The proposed idea projects the k value as two, classifying the water datasets as good or bad.

3.3 Decision Tree

The decision Tree algorithm is mainly used for regression and classification. It is also like the behaviour of humans in classification to understand the problem statement. The Decision nodes and Leaf nodes have a significant impact on the prediction. This implies the tree structure for an input and output described at each leaf. The proposed idea parameters use a decision tree to analyze the statistical parameters by decision rules.

i) The root portion of the tree must contain the features of the given Problem statement.
ii) The splitting is made in the training section as many divided sections.
iii) Search for similar values in the subset and repeat the steps to find the leaf in every part of the branch.

3.4 AdaBoost Classifier

AdaBoost Classifier is like the decision tree algorithm but with only one split. The more weight is given to wrongly classified algorithms. The more weight parameters are considered for the next iteration, which promotes the error reduction by more training.

3.5 XG Boost

Both algorithms originate from the Decision tree in which XG Boost mainly focused on weights assigned for all the variables individually. The prediction is made sequentially using the Decision Tree algorithm. The Loss function can be expressed mathematically by

$$\mathrm{L}^{(\mathrm{t})} = \sum\nolimits_{i=1}^{n} l(Yi, Yi^{(t-1)} + ft(Xi)) + \Omega(ft) \tag{1}$$

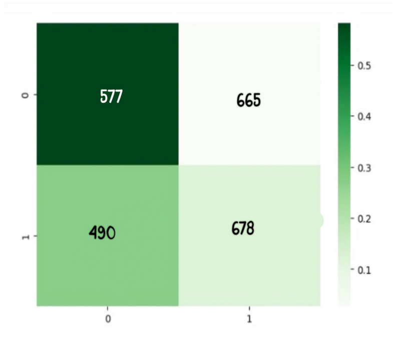
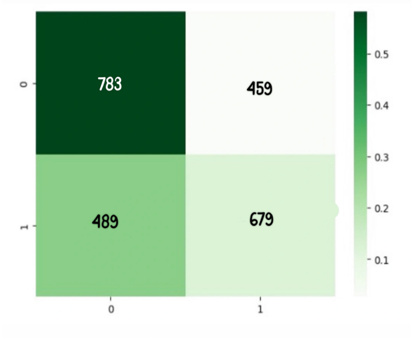
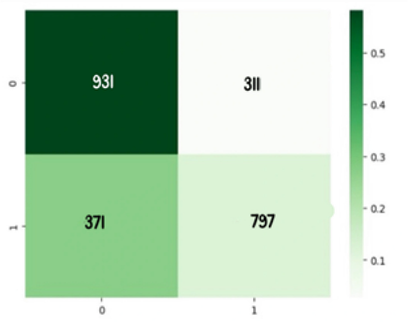
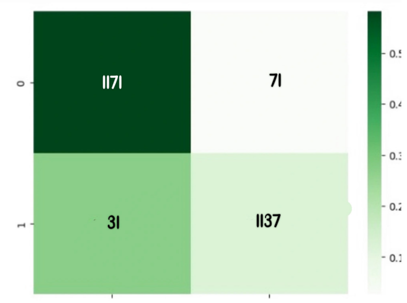
3.6 Feed Forward Neural Network

The feed-forward Neural network also classifies water quality as good or bad. The neural network is framed by adding more dense layers with different neurons in the category. The neural network ended up with the output in the range of [0, 1] by using sigmoid as an activation function. Finally, the network is created, allowing all the features as input, and it is processed by using a hidden layer and produces an output as a single value of 0 or 1.

4 Statistical Parameters of Classification

The Water quality datasets measure the statistical parameters, which contain 9300 instances and nine features. The Features extracted in the datasets have undergone classification like SVM, Decision Tree and KNN. The following chapter follows the Measurement of Statistical parameters for all the classifiers, resulting in greater accuracy in the Decision Tree Algorithm. The support vector machines predict the testing data sets with a lower accuracy and sensitivity of 54, Precision of 49, and F1 Score performance of about 49. The AdaBoost classifier provides greater accuracy than SVM, which increments the value of other statistical parameters. It provides an accuracy of 60% with an actual positive value of 783, a true negative value of 679, a false positive value of about 459, and a false negative value of about 489. The parameters of the XG Boost algorithm indulge the lesser false negative (misclassification) than SVM and KNN. The increment in true positive and true negative values proceeds for a further increase of accuracy percentage to 71. The decision Tree decides whether the water quality is good or bad in good terms. The accuracy and other statistical parameters are higher in value with higher true positive and true negative values. The neural networks join the cluster of correct classes by minimum distance with a greater accuracy of 9.77%. The true positive value is 1094; the true negative value is 716; the false positive is 21, and False Negative is 21 (Table 1).

Table 1. Confusion Matrix of the Classifiers

 Confusion Matrix of SVM	 Confusion Matrix of Ada Boost Classifier
 Confusion Matrix of XG Boost algorithm	 Confusion Matrix of decision tree

5 Comparison of Classifiers

The different classifiers are analyzed to predict the water quality as good or bad. TP, TN, FP, and FN predictions tend to measure the accuracy, sensitivity, specificity, precision and F1 Values. The Feed Forward Neural Network provides an accuracy of 98%, while KNN and Decision Tree follow a higher accuracy rate of 97%. The lower accuracy rate is 52 and 60 for SVM and AdaBoost, respectively. Feed Forward Neural Network provides a higher rate in all other hyperparameters (Table 2).

Table 2. Comparison of Classifiers

Parameters	SVM	KNN	Decision Tree	ADA Boost	XG Boost	Feed-Forward Neural Network
TP	577	1206	1171	783	931	1094
TN	678	1146	1137	679	797	716
FP	665	36	71	459	311	21
FN	490	22	31	489	371	21
Accuracy	52	97	97	60	71	98
Sensitivity	54	71	71	61	71	98
Specificity	50	96	96	59	71	97
Precision	49	97	97	63	71	98
F1	50	81	81	62	74	98

6 Conclusion

Water is one of the existing primary sources in which external pollutants add more contamination particles. There is a need to identify whether the water is in a safe condition by measuring various input features of water stored as a database with two classes as a target. The water quality datasets undergo different machine learning algorithms for the prediction. Out of six machine learning algorithms and feed-forward neural networks, the Feed-forward Neural Network produces the highest prediction accuracy, about 98%. The future work is about implementation and prediction using real-time data acquired from the sensors.

Acknowledgement. This publication is done with the funding support of Muscat College, Muscat, Sultanate of Oman.

References

1. Aldhyani, T.H., Al-Yaari, M., Alkahtani, H., Maashi, M.: Water quality prediction using artificial intelligence algorithms. Appl. Bionics Biomech. **2020** (2020)
2. Lerios, J.L., Villarica, M.V.: Pattern extraction of water quality prediction using machine learning algorithms of water reservoirs. Int. J. Mech. Eng. Robot. Res. **8**(6), 992–997 (2019)
3. Haghiabi, A.H., Nasrolahi, A.H., Parsaie, A.: Water quality prediction using machine learning methods. Water Qual. Res. J. **53**(1), 3–13 (2018)
4. Hayder, G., Kurniawan, I., Mustafa, H.M.: Implementing machine learning methods for monitoring and predicting water quality parameters. Biointerface Res. Appl. Chem **11**, 9285–9295 (2020)
5. Lu, H., Ma, X.: Hybrid decision tree-based machine learning models for short-term water quality prediction. Chemosphere **249**, 126169 (2020)
6. Ahmed, U., Mumtaz, R., Anwar, H., Shah, A.A., Irfan, R., García-Nieto, J.: Efficient water quality prediction using supervised machine learning. Water **11**(11), 2210 (2019). https://doi.org/10.3390/w11112210

7. Ahmed, A.N., et al.: Machine learning methods for better water quality prediction. J. Hydrol. **578**, 124084 (2019). https://doi.org/10.1016/j.jhydrol.2019.124084
8. Fu, Z.: Water quality prediction based on machine learning techniques. Dissertation. University of Nevada, Las Vegas (2020)
9. Shaikh, S.S., Shahapurkar, R.: Machine learning based quality prediction of greywater: a review. In: Information and Communication Technology for Competitive Strategies (ICTCS 2020), pp. 337–347 (2021)
10. Wu, J., Wang, Z.: A hybrid model for water quality prediction based on an artificial neural network, wavelet transform, and long short-term memory. Water **14**(4), 610 (2022)
11. Ortiz-Lopez, C., Bouchard, C., Rodriguez, M.: Machine learning models with potential application to predicting source water quality for treatment purposes: a critical review. Environ. Technol. Rev. **11**(1), 118–147 (2022)
12. Juna, A., et al.: Water quality prediction using KNN imputer and multilayer perceptron. Water **14**(17), 2592 (2022)
13. Ambade, B., Sethi, S.S., Giri, B., Biswas, J.K., Bauddh, K.: Characterization, behavior, and risk assessment of polycyclic aromatic hydrocarbons (PAHs) in the estuary sediments. Bull. Environ. Contam. Toxicol.Contam. Toxicol. **108**, 243–252 (2022)
14. PCRWR. National Water Quality Monitoring Programme, Fifth Monitoring Report (2005–2006); Pakistan Council of Research in Water Resources Islamabad: Islamabad, Pakistan (2007)
15. Najwa Mohd Rizal, N., et al.: Comparison between regression models, Support Vector Machine (SVM), and Artificial Neural Network (ANN) in river water quality prediction. Processes **10**(8) 1652 (2022)
16. Wu, H., et al.: Water quality prediction based on multi-task learning. Int. J. Environ. Res. Public Health **19**(15), 9699 (2022)
17. Hmoud Al-Adhaileh, M., Alsaade, F.W.: Modelling and prediction of water quality by using artificial intelligence. Sustainability **13**(8), 4259 (2021)
18. Burba, F., Ferraty, F., Vieu, P.: K-nearest neighbor method in functional non-parametric regression. J. Nonparametr. Stat. **21**, 453–469 (2009)
19. Geetha, A., Nasira, G.M.: Data mining for meteorological applications: decision trees for modeling rainfall prediction. In: Proceedings of the 2014 IEEE International Conference on Computational Intelligence and Computing Research, Coimbatore, India, 18–20 December 2014, pp. 1–4. IEEE, New York (2014)
20. Fan, Y., Lu, W.X., Miao, T.S., An, Y., Li, J., Luo, J.: Optimal design of groundwater pollution monitoring network based on the SVR surrogate model under uncertainty. Environ. Sci. Pollut. Res. Int.Pollut. Res. Int. **27**, 24090–24102 (2020)
21. Islam, N., Irshad, K.: Artificial ecosystem optimization with deep learning enabled water quality prediction and classification model. Chemosphere **309**, 136615 (2022)

Comparative Analysis of Feature Selection Techniques with Metaheuristic Grasshopper Optimization Algorithm

Qanita Bani Baker(✉) and Moayyad F. Alajlouni

Department of Computer Science, Jordan University of Science and Technology,
Irbid, Jordan
qmbanibaker@just.edu.jo

Abstract. Feature selection (FS) is a critical step used to identify the most relevant and informative features from a given dataset. Feature selection plays a crucial role in dimensionality reduction, improving the used model's performance, enhancing the model's interpretability, and reducing computational complexity. In this study, we conducted a comparative analysis of different feature selection approaches three of them are considered traditional approaches, namely Recursive Feature Elimination (RFE), Mutual Information, K-Best, and one is the heuristic approach which is the Grasshopper Optimization Algorithm (GOA). To evaluate the performance of these approaches, we applied them to classification and regression problems and we applied them in three distinct datasets: Zillow Home Value Prediction, Breast Cancer Wisconsin, and Adult Income. The results of this study indicate that the performance of feature selection methods varies depending on the considered dataset. We observed varying levels of effectiveness across the used datasets. For the Zillow dataset, the grasshopper optimization algorithm yielded the best performance, with a Mean Absolute Error (MAE) of 6.666. However, for the Breast Cancer dataset, the grasshopper optimization algorithm again emerged as the top-performing method, achieving an accuracy of 99.122%. While, for the Adult Income dataset, Mutual Information exhibited the best performance, achieving an accuracy of 86.150%. These findings highlight the importance of considering the complexities and characteristics of the dataset when choosing the feature selection method. Therefore, selecting an appropriate feature selection method may require applying different approaches for optimal feature selection and subsequent model performance.

Keywords: Feature Selection · Meta-Heuristic techniques · Recursive Feature Elimination (RFE) · Mutual Information · K-Best · Grasshopper Optimization Algorithm (GOA) · Regression · Classification

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 159–169, 2024.
https://doi.org/10.1007/978-3-031-56728-5_14

1 Introduction

In developing machine learning algorithms, Feature Selection (FS) plays a crucial role. Performance can be improved by understanding how features impact a model and their relevance. It is a significant task in machine learning. The main objective of feature selection is to reduce the dimensionality of the feature set while maintaining performance accuracy [3].

As the amount of digital data increases, feature selection has become a crucial component in many current applications, particularly in fields such as text mining, pattern recognition, image processing, computer vision, industrial applications, web pages, and bookmark classification, among others [10]. Recent studies have demonstrated the capacity of these algorithms to produce highly accurate results and enhance the attribute selection process. Many meta-heuristic algorithms, such as Simulated Annealing, Ant colony Algorithm, Artificial Bee Colony, Differential Evolution, Genetic Algorithm, Harmony Search, and more are used in feature selection challenges. Despite the wide use of these algorithms, research is ongoing to discover new algorithms that can address this issue. This is done to either find fewer selected features or to increase accuracy [26]. The main goal of optimization is to find variables that minimize or maximize the objective function based on global and local searches. Most of these algorithms are used to develop an approximation method to achieve the best solution to outperform the state-of-the-art objectives [19]. Meta-heuristic algorithms have gained popularity as effective problem solvers in recent years. Some of these algorithms were inspired by human behavior, such as social and political behavior, while others were heavily influenced by nature [7].

Meta-heuristics are a class of optimization techniques that aim to find approximate solutions to difficult problems by iteratively exploring the solution space [11]. They are particularly useful for problems with a large number of variables or when an exact solution is computationally infeasible, Some commonly used meta-heuristics include simulated annealing, genetic algorithms, and ant colony optimization [9]. The field of meta-heuristics is a rapidly evolving one, with new techniques and variations being proposed regularly [8]. Grasshopper optimization algorithm [20] is one of the Meta-heuristics techniques, known as GOA, and is a nature-inspired optimization algorithm based on the behavior of grasshoppers. GOA belongs to the family of swarm intelligence algorithms inspired by the behavior of groups of social organisms or animals to solve optimization problems. GOA is one of the main techniques utilized in this research and used in FS.

2 Related Work

A dimensionality reduction technique known as feature selection serves to increase the efficiency of learning algorithms while also streamlining learning and enhancing data presentation. We concentrate our discussion on earlier work on feature selection in this section. Theoretical and empirical investigations, which are required and adequate to manage feature selection, have benefited from significant contributions from numerous academics. Kanan et al. proposed an improved ant colony optimization strategy for feature selection in face recognition to achieve the lowest classification error. Their algorithm allows for the selection of a feature subset with the minimum feature length and the best classification performance [12]. Muni et al. introduced an approach for online feature selection based on genetic programming. Their proposed GP technique simultaneously constructs a classifier and selects a useful subset of features [13]. Abukhodair et al. used a metaheuristic optimization approach based on big data classification (MOBDC-MR) to efficiently classify large data by selecting the best features. The technique was experimentally validated using a benchmark dataset and evaluated using various metrics. The simulation results indicated that the MOBDC-MR technique had a promising advantage over other methods in terms of performance [2]. Rani et al. proposed an algorithm (the spider monkey optimization approach) for finding the best subset of attributes with the highest degree of classification accuracy. The data were chosen using a spider monkey optimization method, and the SVM was used to calculate the fitness for classification accuracy. The suggested algorithm performed better in classification accuracy than existing methods and selected fewer genes [18]. An improved feature selection algorithm (FACO) was designed by Peng et al, which was evaluated using the KDD CUP99 dataset in a simulation experiment on MATLAB 2014a. The results demonstrate that the FACO algorithm can increase the classification efficiency and accuracy of the classifiers, which has significant practical implications [16]. Spolaôr et al. proposed the ReliefF algorithm as a typical filtering technique. The algorithm screened out irrelevant genes using a threshold and identified the best gene subset. After dimension reduction, the data was classified and recognized using traditional classification procedures. This technique successfully removed redundant and unnecessary genes and improved classification effects with fewer distinctive genes [25].

Wang et al. used the RSFS algorithm for feature selection. To demonstrate its performance, they used two credit datasets from the UCI database. The experimental findings demonstrate that RSFS outperforms base classification methods in terms of reducing computational costs and enhancing classification accuracy [28]. Xi et al. focused on feature gene selection for cancer classification. They used an optimization algorithm to choose a subset of genes. The experimental findings demonstrate that the Binary Encoded Quantum-Behaved Particle Swarm Optimization algorithm (BQPSO)/SVM has a considerable advantage over the other two algorithms in terms of accuracy, robustness, and the number of feature genes selected [30]. Annavarapu et al. introduced the Multi-Objective Binary Particle Swarm Optimization (MOBPSO) algorithm for the analysis of data on cancer gene expression. They used a quick heuristic-based pre-processing technique to reduce some of the basic domain features in the initial feature set. The investigations used publicly available benchmark gene expression datasets for leukemia, colon cancer, and lymphoma. The outcomes of three benchmark cancer datasets show that the suggested strategy is workable and efficient [6]. Abualigah et al. suggested a new algorithm based on a hill-climbing technique to enhance text clustering. Four common text datasets with different features were used in the experiments. Interestingly, by creating a new subset of informational text features, the proposed hill climbing algorithm outperforms the other well-known techniques in terms of results [1].

Sivakumar et al. used an Artificial Bee Colony methodology to study, implement, and analyze a feature selection method for the classification of lung cancer image databases. The linear kernel-based support vector machine technique was utilized as the classifier for the chosen subset to assess the correctness of the chosen features. When compared to the unreduced feature set, the ABC with k-NN chose the fewest features with the maximum classification accuracy [23]. Yusta et al. proposed the use of GRASP, Tabu Search, and Memetic Algorithm as metaheuristic approaches to address the problem of feature selection. These methods were compared to a Genetic Algorithm (GA), which is a commonly used metaheuristic technique, as well as to other traditional feature selection techniques such as Sequential Forward Floating Selection (SFFS) and Sequential Backward Floating Selection (SBF). The results indicated that, on average, GRASP and Tabu Search produced superior outcomes compared to the other methods [32]. Sayed et al. introduced the Chaotic Salp Swarm Method (CSSA) algorithm as a promising approach for feature selection. Simulation data showed that CSSA can identify an optimal feature subset that maximizes classification accuracy while requiring the least number of features [22]. Qasim et al. proposed a novel chaotic binary black hole algorithm (CBBHA) and experiments on three chemical datasets showed that CBBHA outperforms the normal BBHA in terms of classification accuracy, number of chosen features, and computation time [17].

Sawhney et al. [21] evaluated the performance of the firefly algorithm, a novel evolutionary computing algorithm, in comparison to other feature selection techniques. The new approach was found to be effective in reducing the number of features, which reduces computational overhead and mitigates the dimensional curse [20]. While Nakamura et al. [15] introduced the binary bat algorithm which is a new feature selection approach inspired by nature. The approach was found to improve the effectiveness of the optimum path forest and outperform some popular swarm-based methods, according to the experimental works that tested the techniques on publicly available datasets. Nagpal et al. [14] investigated the use of the gravitational search algorithm (GSA) for feature selection in medical datasets. The results showed that GSA was successful in reducing the number of features, with an average reduction of 66%. A summary of the related work is shown in Table 1.

The goal of this study is to determine the effectiveness and robustness of the meta-heuristic feature selection algorithm called the Grasshopper Optimization Algorithm (GOA) [20] and to compare it to traditional feature selection techniques which are: Recursive Feature Elimination (RFE), Mutual Information, and K-Best. Three publicly available datasets are used to evaluate the used algorithms.

Table 1. Summary Of The Related Work

Ref	Proposed Technique(s)	Metaheuristic Category	Finding
[12]	An improved ant colony optimization strategy for feature selection in face recognition	Swarm Based	The suggested method outperforms FS methods based on GA and other methods.
[13]	FS using GA	Evolutionary based	The proposed method performed better for both two-class or multi-class problems
[2]	MOBDC-MR algorithm	Swarm Based	The technique had a promising advantage over the others in terms of performance
[17]	chaotic binary black hole algorithm (CBBHA)	Physics Based	The suggested algorithm compared to the normal BBHA lead to superior performance in terms of classification accuracy
[18]	spider monkey optimization	Swarm Based	SVM was used to calculate fitness for classification accuracy. The suggested algorithm performs better in classification accuracy.
[16]	FACO algorithm	Swarm Based	The findings show that the FACO method may improve classifier efficiency and accuracy.
[21]	Firefly algorithm	Bio-Inspired Based	The new method is effective in reducing the number of features
[25]	Relief algorithm	Swarm Based	This technique successfully removed redundant and unnecessary genes and improved classification effects with fewer distinctive genes
[28]	RSFS algorithm	warm Based	RSFS outperforms base classification methods in terms of reducing computational costs and enhancing classification accuracy
[30]	Encoded Quantum-Behaved PSO (BQPSO)/SVM	Swarm Based	The proposed algorithm shows an advantage over the other two algorithms in terms of accuracy, robustness, and the number of feature genes selected
[15]	Binary bat algorithm	Bio-Inspired Based	The proposed method boosts the efficiency and outperforms some swarm-based techniques, according to tests on publicly available datasets
[6]	Multi-Objective Binary Particle Swarm Optimization (MOBPSO) algorithm	Swarm Based	The investigations use publicly available benchmark gene expression datasets for leukemia, colon cancer, and lymphoma
[1]	The suggested new algorithm based on a -hill-climbing technique	Swarm Based	The proposed hill climbing algorithm outperforms the other well-known techniques in terms of results
[23]	Artificial Bee Colony methodology (ABC)	Evolutionary based	The ABC with k-NN chooses the fewest features with the maximum classification accuracy
[32]	GRASP, Tabu Search, and Memetic Algorithm metaheuristic approaches	Swarm Based	The findings indicate that, on average, GRASP and Tabu Search produce noticeably superior outcomes to the other approaches
[22]	Chaotic Salp Swarm Method (CSSA)	Swarm Based	The results show that CSSA is capable of finding a perfect feature subset that optimizes classification accuracy.
[14]	Gravitational search algorithm (GSA)	Physics Based	The GSA succeeded in lowering the number of features by an average of 66%

3 Materials and Methods

3.1 Methodology

In this study, we defined the used datasets with the complete set of available features. Then, feature selection algorithms are used to identify the most important features. We used several techniques, these are the Recursive Feature Elimination (RFE), Mutual Information, K-Best, and Grasshopper Optimization Algorithm (GOA). A subset of features is selected based on the results of the feature selection algorithm. The subset is chosen by ranking the features based on their importance. The performance of the feature subset is validated using a combination of performance metrics such as mean absolute error (MAE) for regression problem datasets and accuracy for classification problem datasets. The performance of the feature subset was compared to the performance of the complete set of features using the evaluation metrics. The goal is to determine which feature selection approach performs better and how it affects the performance of the model. The flow chart for the proposed method is shown in Fig. 1.

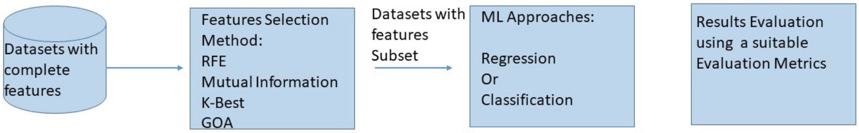

Datasets with complete features
Features Selection Method:
RFE
Mutual Information
K-Best
GOA
Datasets with features Subset
ML Approaches:
Regression
Or
Classification
Results Evaluation using a suitable Evaluation Metrics

Fig. 1. The proposed Method Flow

3.2 Feature Selection Methods

Recursive Feature Elimination (RFE) [31], Mutual Information [27], K-Best [4] and Grasshopper Optimization Algorithm (GOA) [20] are feature selection methods that are used in this work to reduce the dimensionality of the data and select a subset of the most informative features for a given task. RFE is a wrapper method that uses a supervised learning algorithm to rank the importance of features [31]. Mutual Information is a filter method that measures the dependence between features and the target variable [27]. K-Best is a filter approach that selects the k features with the highest score based on a specific scoring function [4]. The Grasshopper Optimization Algorithm (GOA) is a meta-heuristic optimization algorithm inspired by the foraging behavior of grasshoppers. GOA algorithm uses a population-based search to explore the feature space and identify the most informative features for a given task on a specific dataset. GOA is shown to be effective in various feature selection and optimization tasks, and it is known for its ability to handle high-dimensional and noisy data as presented in [5].

The first step in the Grasshopper Optimization algorithm is the initialization of the population of grasshoppers, each population represents a binary vector of feature selection. Due to its binary nature, features may either be chosen (1) or not (0). A population of 10 and 20 grasshoppers is used for the Adult Income and Breast Cancer Wisconsin datasets, respectively. However, this value is 12 for the Zillow Home Value Prediction dataset.

Various techniques are used to assess the algorithm's fitness depending on the datasets. The effectiveness of the feature subsets is evaluated for the first two datasets (Adult Income and Breast Cancer Wisconsin) using a Random Forest Classifier with a parameter called $n_estimators$ set to 10 and 100. A linear regression model is used for this purpose in the Zillow dataset. The mean squared error is used as the fitness score for the regression dataset while the accuracy of predictions serves as the fitness score for classification datasets. The fundamental GOA technique is updating each grasshopper's location according to its fitness score, repeating this process for a predetermined number of iterations (100 for the first two datasets and 50 for the Zillow dataset), and then choosing the features of the grasshopper with the greatest performance. Following optimization, a model is trained using the chosen features, and its performance is assessed. The Adult Income and Breast Cancer Wisconsin datasets obtained a high accuracy with the chosen features in the findings presented.

Initial feature engineering for the Zillow dataset entails removing columns with more than 30% missing values. After applying GOA, the Mean Absolute Error (MAE) is used to assess how well the Linear Regression model performed using both the original data and the optimized features. A subset of 9 characteristics for the Zillow Home Value Prediction dataset was chosen by GOA. The difference in the number of features chosen highlights the variety of datasets and the possible effects of data nature on feature selection procedures.

For the other feature selection techniques, for the K-Best we set k=10 in all used datasets. This method selects the top 'k' features based on statistical tests. For RFE (Recursive Feature Elimination), we also set the number of selected features called to 10 in all datasets where RFE method recursively removes the least important feature(s) and rebuilds the model. For the Mutual Information Regression, we evaluate the dependency between two variables, selecting those with higher dependency on the target. For all used data sets, we applied the default implementation to compute mutual information scores. The code and datasets related to this research are available upon request by contacting the researchers directly.

3.3 DataSets

In this project, we have used three datasets to evaluate the performance of the applied feature selection methods: the Zillow Home Value Prediction dataset, which is a dataset employed for a regression problem, the Breast Cancer Wisconsin dataset, which is a data set for a classification problem, and the Adult Income dataset, which is also for a classification problem. These datasets are publicly available and have been widely used in machine learning research. The

Zillow Home Value Prediction dataset includes features related to properties such as location, size, and age, as well as home value. The Breast Cancer Wisconsin dataset includes features related to breast cancer patients, such as the size and shape of the tumor, as well as the diagnosis. The Adult Income dataset includes features related to adult individuals such as age, education, occupation, and income.

3.4 Evaluation Metrics

For the Regression problem, we used Mean Absolute Error (MAE). MAE is a commonly used metric for evaluating the performance of regression models. It measures the average difference between the predicted values and the actual values of the target variable. The MAE is calculated by taking the absolute value of the difference between the predicted and actual values for each data point, and then averaging those values over the entire dataset [29], the mean absolute error formula is shown in Eq. 1.

$$MAE = (1/n) * \sum |y_i - \hat{y_i}| \tag{1}$$

For the classification problems, we used accuracy. The accuracy of a model is determined by how well it can find correlations and patterns between variables. the stronger a model's ability to generalize to 'unseen' data, the better predictions and insights it can provide. It's simple to figure out simply by dividing the number of right predictions by the total number of predictions [24], the accuracy formula is shown in Eq. 2.

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{2}$$

4 Result and Discussion

For the Zillow dataset, using the original features resulted in MAE of 6.673, while using RFE resulted in MAE of 6.672. Using Mutual Information resulted in MAE of 6.673 and using K-Best resulted in MAE of 6.676. The best performance was achieved using GOA with MAE of 6.666.

For the Breast Cancer dataset, using the original features resulted in an accuracy of 95.614, while using RFE resulted in an accuracy of 96.491. Using Mutual Information resulted in an accuracy of 96.491 and using K-Best resulted in an accuracy of 92.982. The best performance was achieved using GOA with an accuracy of 99.122.

For the Adult Income dataset, using the original features resulted in an accuracy of 84.972, while using RFE resulted in an accuracy of 85.167. Using Mutual Information resulted in an accuracy of 86.150 and using K-Best resulted in an accuracy of 83.898, and GOA resulted in an accuracy of 83.355, The best performance was achieved using Mutual Information with an accuracy of 86.150.

The results show that using feature selection methods, such as RFE, Mutual Information, K-Best, and GOA, can improve the performance of the models in comparison to using the original features. However, the performance improvement varies depending on the dataset and the evaluation metric used.

These results suggest that different feature selection methods may perform better or worse depending on the specific dataset and evaluation metric used. It is recommended to try multiple methods and compare their performance to determine the most appropriate one for a given task, Table 2 presents a summary of the results obtained.

Table 2. Summary of the results

Method	Dataset	Evaluation Metric	Value
Original Features	Zillow	MAE	6.673
RFE	Zillow	MAE	6.672
Mutual Information	Zillow	MAE	6.673
K-Best	Zillow	MAE	6.676
GOA	Zillow	MAE	**6.666**
Original Features	Breast Cancer	Accuracy	95.614
RFE	Breast Cancer	Accuracy	96.491
Mutual Information	Breast Cancer	Accuracy	96.491
K-Best	Breast Cancer	Accuracy	92.982
GOA	Breast Cancer	Accuracy	**99.122**
Original Features	Adult Income	Accuracy	84.972
RFE	Adult Income	Accuracy	85.167
Mutual Information	Adult Income	Accuracy	**86.150**
K-Best	Adult Income	Accuracy	83.898
GOA	Adult Income	Accuracy	83.355

5 Conclusion

In conclusion, this study has presented an analysis of the performance of several feature selection methods and applied them to three datasets: Zillow, Breast Cancer, and Adult Income. The results have shown that the performance of different methods varies depending on the dataset. For the Zillow dataset, the best performance was achieved by using the grasshopper optimization algorithm with an MAE of 6.666. For the Breast Cancer dataset, the best performance was achieved also using the grasshopper optimization algorithm with an accuracy of 99.122. For the Adult Income dataset, the best performance was achieved using Mutual Information with an accuracy of 86.150. These results demonstrate that different feature selection methods can have varying levels of effectiveness on different datasets, and it is important to consider the specific characteristics of the dataset when choosing a feature selection method. Overall, the results of this study provide insights into the strengths and limitations of different feature selection methods and can aid in the selection of an appropriate method for a given dataset experimentally. Further research can be done to optimize the performance of the feature selection methods used, such as experimenting with different parameter settings or trying other metaheuristic methods. Additionally, ensemble methods and testing the methods on other datasets can be used to generalize the results.

References

1. Abualigah, L.M., Khader, A.T., Al-Betar, M.A., Alyasseri, Z.A.A., Alomari, O.A., Hanandeh, E.S.: Feature selection with β-hill climbing search for text clustering application. In: 2017 Palestinian International Conference on Information and Communication Technology (PICICT), pp. 22–27. IEEE (2017)
2. Abukhodair, F., Alsaggaf, W., Jamal, A.T., Abdel-Khalek, S., Mansour, R.F.: An intelligent metaheuristic binary pigeon optimization-based feature selection and big data classification in a mapreduce environment. Mathematics **9**(20), 2627 (2021)
3. Agrawal, P., Abutarboush, H.F., Ganesh, T., Mohamed, A.W.: Metaheuristic algorithms on feature selection: a survey of one decade of research (2009–2019). IEEE Access **9**, 26766–26791 (2021)
4. Akman, D.V., et al.: K-best feature selection and ranking via stochastic approximation. Expert Syst. Appl. **213**, 118864 (2023)
5. Aljarah, I., Al-Zoubi, A.M., Faris, H., Hassonah, M.A., Mirjalili, S., Saadeh, H.: Simultaneous feature selection and support vector machine optimization using the grasshopper optimization algorithm. Cogn. Comput. **10**, 478–495 (2018)
6. Annavarapu, C., Dara, S., Banka, H.: Cancer microarray data feature selection using multi-objective binary particle swarm optimization algorithm. EXCLI J. **15**, 460–473 (2016)
7. Atashpaz-Gargari, E., Lucas, C.: Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition. In: 2007 IEEE Congress on Evolutionary Computation, pp. 4661–4667. IEEE (2007)
8. Baghel, M., Agrawal, S., Silakari, S.: Survey of metaheuristic algorithms for combinatorial optimization. Int. J. Comput. Appl. **58**(19) (2012)
9. Dorigo, M., Stützle, T.: The ant colony optimization metaheuristic: algorithms, applications, and advances. Handbook of Metaheuristics, pp. 250–285 (2003)
10. Forman, G., et al.: An extensive empirical study of feature selection metrics for text classification. J. Mach. Learn. Res. **3**(Mar), 1289–1305 (2003)
11. Gandomi, A.H., Yang, X.S., Talatahari, S., Alavi, A.H.: Metaheuristic algorithms in modeling and optimization. Metaheuristic Appl. Struct. Infrastruct. **1**, 1–24 (2013)
12. Kanan, H.R., Faez, K.: An improved feature selection method based on ant colony optimization (ACO) evaluated on face recognition system. Appl. Math. Comput. **205**(2), 716–725 (2008)
13. Muni, D.P., Pal, N.R., Das, J.: Genetic programming for simultaneous feature selection and classifier design. IEEE Trans. Syst. Man Cybern. Part B (Cybern.) **36**(1), 106–117 (2006)

14. Nagpal, S., Arora, S., Dey, S., et al.: Feature selection using gravitational search algorithm for biomedical data. Procedia Comput. Sci. **115**, 258–265 (2017)
15. Nakamura, R.Y.M., Pereira, L.A.M., Rodrigues, D., Costa, K.A.P., Papa, J.P., Yang, X.S.: Binary bat algorithm for feature selection. In: Swarm Intelligence and Bio-inspired Computation, pp. 225–237. Elsevier (2013)
16. Peng, H., Ying, C., Tan, S., Hu, B., Sun, Z.: An improved feature selection algorithm based on ant colony optimization. IEEE Access **6**, 69203–69209 (2018)
17. Qasim, O.S., Al-Thanoon, N.A., Algamal, Z.Y.: Feature selection based on chaotic binary black hole algorithm for data classification. Chemom. Intell. Lab. Syst. **204**, 104104 (2020)
18. Rani, R.R., Ramyachitra, D.: Microarray cancer gene feature selection using spider monkey optimization algorithm and cancer classification using SVM. Procedia Comput. Sci. **143**, 108–116 (2018)
19. Razmjooy, N., Khalilpour, M., Ramezani, M.: A new meta-heuristic optimization algorithm inspired by FIFA world cup competitions: theory and its application in PID designing for AVR system. J. Control Autom. Electr. Syst. **27**, 419–440 (2016)
20. Saremi, S., Mirjalili, S., Lewis, A.: Grasshopper optimisation algorithm: theory and application. Adv. Eng. Softw. **105**, 30–47 (2017)
21. Sawhney, R., Mathur, P., Shankar, R.: A firefly algorithm based wrapper-penalty feature selection method for cancer diagnosis. In: Gervasi, O., et al. (eds.) ICCSA 2018. LNCS, vol. 10960, pp. 438–449. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-95162-1_30
22. Sayed, G.I., Khoriba, G., Haggag, M.H.: A novel chaotic salp swarm algorithm for global optimization and feature selection. Appl. Intell. **48**, 3462–3481 (2018)
23. Sivakumar, S., Chandrasekar, C.: Feature selection using ABC forthe lung CT scan images. Int. J. Sci. Eng. Technol. **3**(6), 781–784 (2014)
24. Sokolova, M., Lapalme, G.: A systematic analysis of performance measures for classification tasks. Inf. Process. Manag. **45**(4), 427–437 (2009)
25. Spolaôr, N., Cherman, E.A., Monard, M.C., Lee, H.D.: Relieff for multi-label feature selection. In: 2013 Brazilian Conference on Intelligent Systems, pp. 6–11. IEEE (2013)
26. Talbi, E.G., Jourdan, L., Garcia-Nieto, J., Alba, E.: Comparison of population based metaheuristics for feature selection: application to microarray data classification. In: 2008 IEEE/ACS International Conference on Computer Systems and Applications, pp. 45–52. IEEE (2008)
27. Vergara, J.R., Estévez, P.A.: A review of feature selection methods based on mutual information. Neural Comput. Appl. **24**, 175–186 (2014)
28. Wang, J., Hedar, A.R., Wang, S., Ma, J.: Rough set and scatter search metaheuristic based feature selection for credit scoring. Expert Syst. Appl. **39**(6), 6123–6128 (2012)
29. Willmott, C.J., Matsuura, K.: Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Res. **30**(1), 79–82 (2005)
30. Xi, M., Sun, J., Liu, L., Fan, F., Wu, X., et al.: Cancer feature selection and classification using a binary quantum-behaved particle swarm optimization and support vector machine. Comput. Math. Methods Med. **2016** (2016)
31. Yan, K., Zhang, D.: Feature selection and analysis on correlated gas sensor data with recursive feature elimination. Sens. Actuators B Chem. **212**, 353–363 (2015)
32. Yusta, S.C.: Different metaheuristic strategies to solve the feature selection problem. Pattern Recogn. Lett. **30**(5), 525–534 (2009)

Supermarket Shopping with the Help of Deep Learning

Ioannis Symeonidis[1](✉), Panagiotis Chatzigeorgiou[1],
Christos Antonopoulos[2], Ignatios Fotiou[3], and Mary Panou[1]

[1] Hellenic Institute of Transport, Center for Research and Technology Hellas,
Thessaloniki, Greece
ioannis.sym@certh.gr

[2] Department of Electrical and Computer Engineering, University of Peloponnese,
Patra, Greece

[3] TOBEA SMLLC Old National Road, Patra, Greece

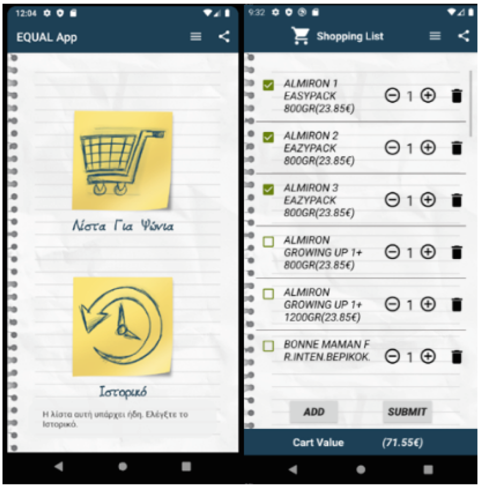
Abstract. This study presents the development of an innovative system designed to facilitate the customers, especially elderly and people with disabilities, in their shopping experience. The proposed solution employes deep learning for product identification and obstacle detection in combination with a smartphone app that serves as the user interface. The solution offers functionalities such as product selection, shortest path calculation, automatic shopping list fulfillment and obstacle detection. The presented solution is part of a greater system that also consists of a self-propelled cart with indoor localization capabilities and a supermarket cloud platform.

1 Introduction

Supermarket shopping can be a time-consuming and cumbersome experience for customers and even more for elderly and individuals with disabilities. However, daily shopping in a supermarket can also be an opportunity for socialization and encouragement for independent living. In EU-27 there are 85 million people with disabilities, aged 16 and over living in private household according to the EU-SILC (EU Statistics on Income and Living Conditions). Despite significant investments in outdoor infrastructure, less emphasis has been placed on the development of shared indoor infrastructure based on advances in information and communication technologies (ICT). To address this issue, we have developed an innovative system designed to facilitate shopping. The system is embedded in a self-propelled cart for people with disabilities. The system is comprised of four subsystems, the self-propelled cart with indoor localization capabilities, a cloud platform, a smartphone app and an embedded system attached to the shopping cart. The self-propelled cart and the cloud platform have been initially presented in [1]. In this paper the focus is on functionalities of the embedded system on the cart that are related to AI and the smartphone app that functions as the user interface of the system. The embedded system and the app support the list preparation, shortest route estimation, obstacle avoidance and automatic

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 170–175, 2024.
https://doi.org/10.1007/978-3-031-56728-5_15

fulfillment of the list. The mobile app was developed with the accessibility standards as defined by the official Android Developer page [6]. The app serves as the primary user interface for the shopping cart, allowing users to create and manage shopping lists, check the fulfillment of the list during shopping and find a shortest route based on the selected products of the shopping list (Fig. 1). The app communicates with the supermarket cloud infrastructure to retrieve product related information concerning availability, prices and location. The relevant AI functions of the system are the automatic list fulfillment with product object detection, the shortest route estimation and the obstacle detection. The above functions are performed on the embedded system and are visualized at the app.



Η λίστα αυτή υπάρχει ήδη. Ελέγξτε το Ιστορικό.

- ☑ ALMIRON 1 EASYPACK 800GR(23.85€) ⊖ 1 ⊕
- ☑ ALMIRON 2 EAZYPACK 800GR(23.85€) ⊖ 1 ⊕
- ☑ ALMIRON 3 EAZYPACK 800GR(23.85€) ⊖ 1 ⊕
- ☐ ALMIRON GROWING UP 1+ 800GR(23.85€) ⊖ 1 ⊕
- ☐ ALMIRON GROWING UP 1+ 1200GR(23.85€) ⊖ 1 ⊕
- ☐ BONNE MAMAN F R.INTEN.BEPIKOK. ⊖ 1 ⊕

Cart Value (71.55€)

Fig. 1. Smartphone app screenshots

2 Methods

2.1 Automatic List Fulfillment

The system employs deep learning for the object detection of the products during the automatic list fulfillment. Two cameras mounted on the shopping cart provide image input for a deep convolutional neural network (D-CNN) model. This camera-based object detection system is capable of identifying and categorizing items into predefined object classes that represent the products. As the products are identified by the model, the information is sent to the supermarket cloud platform and is then communicated to the app (Fig. 2). In order to create the object detection model an image dataset of 41 products was created. A total of 3,000 images were captured using the cameras mounted on the shopping cart. The images were taken from the two different camera positions on the cart and featured various combinations of products in the cart. This dataset was used to train the custom object detection model. During the preprocessing stage, the images were resized and padded to ensure they had appropriate dimensions for model retraining. The aspect ratios of the images were maintained to avoid distortion. Each product in the images was manually annotated using rectangular bounding boxes, with the corresponding class labels assigned to them. This task allows the model to learn both the location and the classification of the objects within the images. The MobileNet-SSD [7] was chosen as the network architecture for the object detection model. This architecture combines MobileNet for image feature extraction and the Single Shot Detector (SSD) for fast object localization. This architecture is primarily used in handheld and embedded devices for fast recognition. A pre-trained MobileNet-SSD model, initially trained on the COCO dataset [5] with 80 image classes, was modified to accommodate the custom dataset of 41 supermarket products. Transfer learning was employed to save training time by leveraging the pre-existing knowledge gained from the COCO dataset. The initial layers of the model, responsible for detecting low- and mid-level features such as edges and lines, were preserved. The last layer of the model was adjusted to accommodate the 41 product classes in the custom dataset. The PyTorch framework was used to retrain the MobileNet-SSD model on the custom dataset. The 3,000 images were divided into three sets: training, validation, and test sets. The model was trained using the training set and fine-tuned with the validation set to minimize the risk of overfitting. Finally, the test set was used to evaluate the model's performance on unseen data. During training, various hyperparameters such as learning rate, batch size, and the number of epochs were adjusted to optimize the model's performance. The accuracy metric was used to evaluate the model's ability to correctly detect and classify products in the images. Once the model was trained and evaluated, it was deployed on an NVIDIA Jetson Nano single board computer, by converting the model format from pytorch, to ONNX and finally to TensorRT engine. The NVIDIA Jetson Nano is a compact embedded system ideal for edge computing applications. The device received the image input from the two cameras mounted on the shopping cart and was able to perform real-time object detection at 15 frames per second.

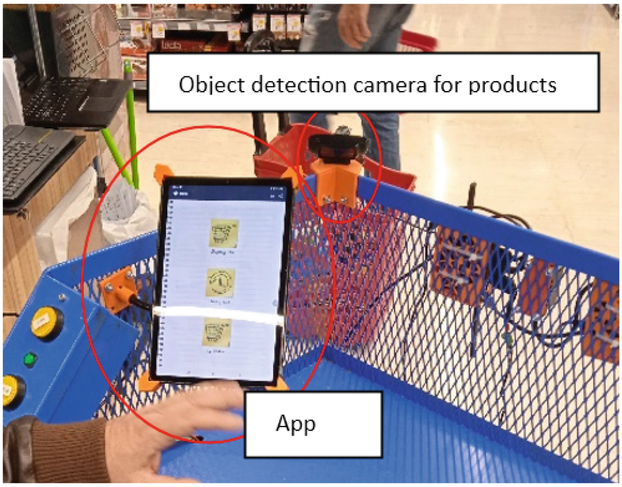


Fig. 2. Object detection camera for product identification

2.2 Shortest Root Estimation

The problem of routing in the supermarket, with target to find the shortest path for gathering all the products, can become very computationally demanding with the increase of the number of products. In the literature heuristic methods as genetic algorithms [2] are used instead of analytical ones as dynamic programming. In our solution we describe the problem as the Traveling Salesman Problem (TSP) [4]. The product collection process is modeled as a variant of the TSP, where products correspond to cities and the store's entry-exit corresponds to the first city. A network of nodes is created with fixed nodes at the beginning of the aisles and additional nodes at the locations of the products. The calculation of the shortest distances between products is performed using Dijkstra's algorithm [3]. Once the lengths of the shortest distances are calculated, we have a full correspondence with the TSP. To solve the TSP we use the simulated annealing method [8] (Fig. 3).

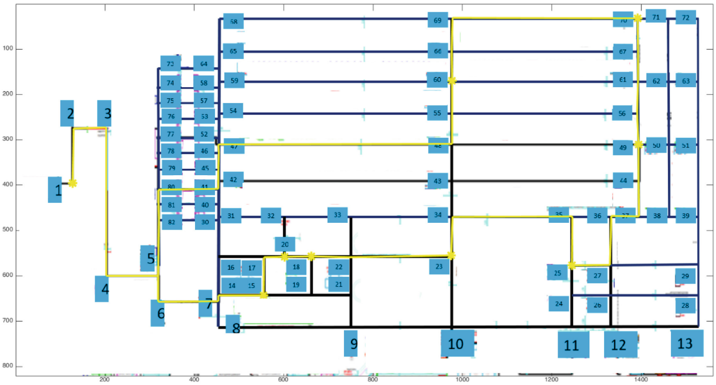


Fig. 3. The nodes of the supermarket corridors (cyan) and the rooting solution (yellow)

2.3 Obstacle Detection

To detect obstacles and prevent collisions with customers and other carts, our system incorporates the fusion of camera and LIDAR sensor technology. The sensors are mounted on the cart's front (Fig. 4). This integration enables the detection of obstacles and their distances from the cart, allowing for real-time path adjustments to avoid collisions. Sensor fusion is used to improve the accuracy and reliability of obstacle detection by combining data from multiple sensors. The LIDAR is suitable for measuring distance and creating a 3D point cloud of the environment, while the camera provides high-resolution images and additional information on object characteristics such as colour and texture. The camera is used to identify and detect the position of the obstacle and calculate the azimuth of the obstacle in the camera coordinate system. Human detection is performed using a ready-made deep learning neural network model, while the cart detection uses a new model developed with transfer learning on the MobilenetSSD architecture as it was described in the previous chapter. Once the obstacle is detected, the azimuth is used by the LIDAR to calculate the average distance from the 3D pointcloud. After the obstacle distance is calculated, it is sent to the cloud platform and it is then communicated to the app. The tablet mounted on the shopping cart displays the cart's trajectory and detected obstacles in real-time, providing warning of obstacles to the user and allowing for a safer shopping experience.

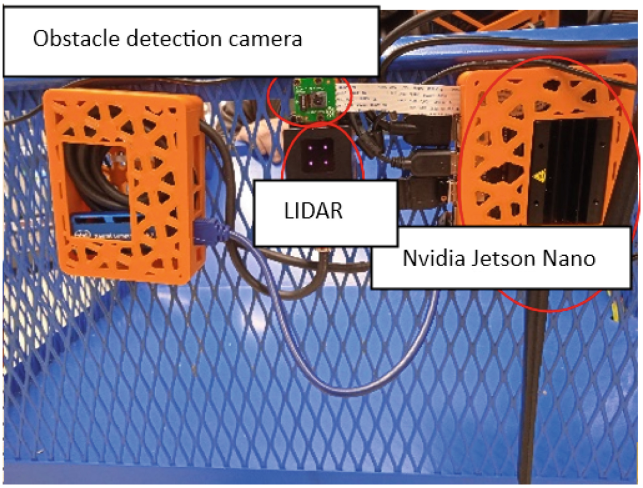


Fig. 4. Object detection camera for product identification

3 Results and Limitations

The system was tested in real-life supermarket scenarios to evaluate its accuracy and limitations. The performance of the system was analyzed based on the accuracy of product detection and the overall user experience. Challenges and limitations observed during field tests were documented to guide future improvements. The system achieved an accuracy of approximately 70 % on the product identification task. Limitations were observed in the detection of certain products, particularly those obscured by others in the cart. A solution to this problem is an increase to the training dataset with more images with multiple products in the cart in order to improve the model's detection capabilities. The proposed shortest root estimation and obstacle detection system for the supermarket cart was tested using the supermarket map and the products positions. The simulated annealing TSP algorithm successfully optimized the route for visiting products in the shopping list, significantly reducing the overall distance traveled. The camera and LIDAR sensor fusion technology effectively detected moving obstacles such as people and shopping carts. The tablet mounted on the shopping cart provided users with up-to-date information on the cart's trajectory and detected obstacles, offering valuable assistance to elderly or disabled shoppers who may have difficulties maneuvering the cart.

4 Conclusion

This study demonstrated the potential of deep learning in supermarket shopping with target to facilitate the customer experience. A dataset of images of supermarket products was created and used for training the object detection model for the automatic shopping list fulfilment. Additionally an heuristic algorithm was used for shortest route estimation by modeling the product collection process as a variant of the Traveling Salesman Problem. Furthermore, the integration of camera and LIDAR sensor technology for obstacle detection prevents collisions with customers and other carts. The proposed system holds significant potential for enhancing the supermarket shopping experience, particularly for elderly and disabled shoppers, by improving efficiency, ensuring safety, and offering valuable assistance. Additionally, extensive testing and validation in real-world supermarket settings will be conducted to further assess the system's performance and potential impact on the shopping experience.

Acknowledgements. The research was performed during the EQUAL project and it was partially funded from the Greek General Secretariat for Research and Innovation with contract number T1EDK-04183.

References

1. Antonopoulos, K., et al.: A distributed embedded systems IoT platform and associated services supporting shopping cart for disabled people. IEEE (2022)
2. Chen, X., Li, Y., Hu, T.: Solving the supermarket shopping route planning problem based on genetic algorithm. In: 2015 IEEE/ACIS 14th International Conference on Computer and Information Science (ICIS), pp. 529–533. IEEE (2015)
3. Dijkstra, E.W.: A note on two problems in connexion with graphs. In: Edsger Wybe Dijkstra: His Life, Work, and Legacy, pp. 287–290 (2022)
4. Jünger, M., Reinelt, G., Rinaldi, G.: The traveling salesman problem. Handb. Oper. Res. Manag. Sci. **7**, 225–330 (1995)
5. Lin, T.-Y., et al.: Microsoft COCO: common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8693, pp. 740–755. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10602-1_48
6. Principles for improving app accessibility (2022). Accessed 16 Dec 2022
7. Phan, H., He, Y., Savvides, M., Shen, Z., et al.: MobiNet: a mobile binary network for image classification. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3453–3462 (2020)
8. Stewart, W.R.: A computationally efficient heuristic for the traveling salesman problem. In: Proceedings of the 13th Annual Meeting of Southeastern TIMS, Myrtle Beach, SC, USA, pp. 75–83 (1977)

A Decision Support System for Detecting FIP Disease in Cats Based on Machine Learning Methods

Ozge Doguc[1](✉), Sevval Beyhan Bilgi[1], Seval Cagdas[1], and Nevin Yilmazturk[2]

[1] Istanbul Medipol University, Istanbul, Turkey
odoguc@medipol.edu.tr
[2] Pusula Software, Istanbul, Turkey

Abstract. Cats are close friends who live with us in all aspects of life. Many diseases endanger the quality of life of cats that live with us. One of the most dangerous is infectious peritonitis in cats, also known as FIP; which is a coronavirus that affects a cat's overall metabolism. There is no specific treatment for FIP and existing drugs are difficult to find and very expensive; therefore, early detection is very important. The most important thing for early detection is to know the body changes caused by the disease, i.e., symptoms, to take appropriate measures. By collecting and interpreting information such as the combination of symptoms, the age at which cats are most common, and the breeds most encountered, cat owners can take precautions even when they cannot be alert. Therefore, in this study, an early detection method for FIP disease in cats is introduced by making predictions using Naive Bayes algorithm. The dataset includes of 300 FIP symptoms used by Jones et al. [11], and from Ümraniye Vita Veterinary Clinic data were obtained from 150 cats who did not have FIP but went to the clinic for other diseases. This generated dataset is resampled using the Smote algorithm to enlarge the dataset. Then the Google Colab program is used to create a naive Bayesian model using the Python programming language. For this study a model is built using the Naive Bayes algorithm, and it is shown that the model can predict the FIP disease with 96% accuracy.

Keywords: Feline Diseases · FIP · Naive Bayes Algorithm · Machine Learning

1 Introduction

In many parts of the world there are organizations established to improve the lives of animals and continue to provide these services to this day. Some of these organizations are supported by the government, but most of the organizations that aim to improve the lives of animals and raise the standard of living are NGOs. In this case, our country has NGOs working to protect animals and meet basic needs such as shelter and food. In addition, our compatriots are also committed to helping these creatures. This is arguably the most important factor in improving an animal's quality of life. The fact that these

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 176–186, 2024.
https://doi.org/10.1007/978-3-031-56728-5_16

animals live in the same environment, eat the same containers, drink water, and defecate in the same areas means that outbreaks occur more often and spread faster. This also applies to those of you who live in our house. Disease spreads faster, especially in environments where many animals are cared for.

Cats, just like humans, can contract diseases, their standard of living can decrease, or they can lose their lives because of these diseases. Infectious peritonitis, also known as FIP in cats is a disease that affects all organs of the cat, causing limited movement, decreased appetite, fever due to infection and weakness due to fever. Symptoms in cats are important for diagnosing infectious peritonitis in cats. To determine this, it is important to understand the signs and interpret them correctly. There are two types of FIP, wet and dry. The wet PIF type is more difficult to detect and treat than the dry type. The reason is that the wet FIP is more aggressive and progressive, its symptoms are more observable in a laboratory setting. In addition to the data obtained in the FIP laboratory setting, there are observable symptoms, such as difficulty in walking, loss of appetite, weakness, fever, etc. Also, blood tests are needed to confirm the diagnosis. Recognizing the disease in cats through symptoms and taking the cat to the vet help detect the disease early. Veterinary experts have shown that most cats that survive FIP due to early detection. Since there is no single treatment, the treatment process is painful and long.

For this reason, identifying the disease at an early stage of detection through symptoms, and initiating treatment will help defeat this devastating disease. Nowadays, studies have been started using machine learning algorithms to detect many diseases in the medical field. The benefit of providing early detection using machine learning algorithms is that it helps to perform detection quickly without consuming a lot of resources to look at previous data. This study uses classification algorithms as decision support systems for FIP detection.

2 Literature Survey

Cats are animals that have been at the center of our lives for a very long time. According to Golab [8], cats have been living with humans for 4000 years. Domesticated since the time of Ancient Egypt, these lovely friends bring happiness to our homes and lives. In his article, Börkü [4] talked about the internal diseases that cats can experience throughout their lives and the diagnosis and treatment methods of these diseases. One of these diseases is feline infectious peritonitis, also known as FIP. According to Aytug [1], FIP is a type of infection caused by the coronavirus of cats. Although there is no definite information about how this disease is transmitted, it is thought that an infected cat can pass it on to other cats through defecation. FIP is the most difficult disease to diagnose in cats and this diagnosis is dependent on many factors such as the patient's age, gender, and tests come together. According to Öztürk [20], fatigue, lethargy, fluctuating fever, loss of appetite, weight loss, neurological symptoms, respiratory distress, etc. According to Kurban [14], who mentions symptoms like Öztürk [20], there are certain symptoms of FIP, such as swelling, observable changes in the eyes, due to fluid accumulation in the abdominal cavity or chest. According to the research conducted by Kahraman [12], the diagnosis of the FIP disease is not only made with clinical findings, but also that several different tests come together to make decisions. According to Kuruçay and Gümüşova [15], FIP treatment is carried out with antiviral drugs. However, they mentioned that there is no definite treatment method for this disease. For this reason, understanding the FIP disease and correlate its symptoms will be an important step in the early detection of the disease. Using data mining algorithms to derive decisions based on the symptoms of FIP will make the operations much easier.

Machine learning studies using data mining algorithms are expected to play an important role in disease prediction in the future [6, 22, 25–27]. Özlüer et al. [18] wanted to make the classification of diabetes disease using many machine learning algorithms in their study. The reason for using more than one algorithm in this study is to measure the performance differences between different algorithms. In their study, Özmen and Avcı [19] compared the performance of the classification algorithms using the data of heart diseases. As a result of the comparison, they said that data mining algorithms can make a prediction for early detection of diseases. Coşar and Deniz [7], in this study, aimed to detect heart diseases by using machine learning algorithms. Emphasizing that heart diseases cause 17 million deaths every year, Coşar and Deniz [7] emphasized in their study that early detection can help patients survive by early detection of a disease that can result in death. In this study, they tried many classification algorithms and chose the best performing algorithm. As a result of this study, they stated that the classification algorithm that gave the best performance was the Random Forest algorithm. According to Özlem and Güngör [17], collective classification algorithms are learning algorithms that produce more than one classification instead of a classifier and then classify new data with the votes taken from their predictions. Although classification algorithms are generally used in the detection of heart diseases, they are also preferred in the detection of other diseases [26]. The reason for this is the working principles. Baydan [2], in his doctoral study titled "Detection and classification of fractures on dog and cat tibia bones with deep learning", revealed that it is possible to detect animal diseases with artificial intelligence.

2.1 Cat Diseases

Cats, like humans, can get flu, fever, diarrhea, nausea, and many other diseases. Since they are social creatures, they can easily pass these diseases to each other. For this reason, it is very easy for them to catch the disease. Disease detection in cats is not as easy as in human diseases. There are multiple reasons for this. One of these reasons is that cats are aggressive and prone to violence when they enter different environments. The reason why they behave in an aggressive and violent way is the instinct to protect themselves against any threat that may or may occur. For this reason, since they do not stand still, it becomes difficult to examine them and the detection of diseases is indirectly difficult. Another reason is that they cannot tell us about the pain they are experiencing, in which parts they feel pain, and what kind of situation they are in. For this reason, veterinarians should carefully examine every detail, review every symptom and, if necessary, examine laboratory findings. Due to the inability to explain their own problems, the diagnosis of the disease may not always be correct. For this reason, the symptoms observed by the host at home are very important for veterinarians. In their study, Maden and Çuhadar [16] focused on the laboratory diagnosis of endocrine diseases of cats and dogs, and they analyzed how these diseases are detected, the results of laboratory tests and their interpretation. They are complex diseases that can be detected by evaluating the laboratory changes including hormone interactions, hematological, biochemical and urine tests, ultrasound, x-ray, and the findings obtained by the veterinarian together. Within the scope of the basic clinical findings and laboratory findings of the endocrine disease, co-occurring diseases and the symptoms that distinguish these diseases, namely the symptoms of the disease, are necessary for the clear detection of endocrine diseases and effective treatment management. In addition to these, endocrine diseases should always be followed without interruption to prevent the side effects of drugs and to evaluate how successful the treatment is. In this respect, some tests used in detection are also used for monitoring. Endocrine system diseases are a disease that can also be seen in humans. Problems in the endocrine system in humans occur because of the disruption of the functioning of the hormonal glands and the change in the hormonal balance. This situation is similar in cats and animals. In this study, Yılmaz et al. [17] evaluated data from the cats and dogs that visited the veterinary clinic between 1990–2000. According to this evaluation, it was determined that the most observed diseases in cats brought to the veterinary clinic were digestive system diseases, respiratory system diseases, endoparasitic diseases, urinary system diseases, infectious diseases, and skin diseases, respectively. Soylu et al. [23] conducted research on the use of coenzymes in cardiovascular diseases in cats and dogs in this project. Rather than detecting disease or comparing one disease with another, this research wanted to observe whether the use of coenzymes would be successful in the treatment of cardiovascular diseases. As a result of this research, it has been seen that the use of coenzymes is beneficial for cardiovascular diseases to an observable extent.

Apart from these diseases, cats can also get sick because they carry viruses. It is known that cats, like humans, carry coronavirus, and due to this disease, they can catch a disease that is difficult for both cats and their owners and can come to the point of death. This disease caused by the coronavirus is called feline infectious peritonitis, also known as FIP.

2.2 Feline Infectious Peritonitis (FIP)

There are many diseases that threaten the lives of cats. Due to these diseases, their living standards decrease, and they cannot behave in the way they want. Cat's infectious peritonitis, which we can define as one of the diseases whose symptoms should be known by cat owners and at the same time one of the most important, is FIP with its common name and its short name. FIP is a disease in cats caused by the infection of coronaviruses. Camkıranlar [5] described FIP disease in detail, which was first defined as "an important disease in cats" in 1963. It was first discovered by Wolfe and Griesemer and spread rapidly worldwide. FIP is caused by feline coronaviruses and is a chronic, progressive, and often fatal disease. There are two types of FIP: dry and wet forms, classified by Montali and Strandberg in 1972. Baydar et al. [3] studied the detection of FIP disease using observed symptoms and clinical data. The study found that FIP is more common in cats compared to infectious peritonitis due to its more widespread presence. Cats afflicted with FIP typically have ages between 6 months and 2 years, with the most obvious symptom being deterioration of the eye structure. Hartmann [10] studied the discovery of FIP by veterinarians, mutation progression, and transmission mode. He found that the spread of infectious peritonitis in cats in the same environment is faster and the rate of transmission is higher by defecation methods. Karayiğit et al. [13] described FIP disease in a Scottish Fold cat, who had symptoms such as swelling in the abdomen, decreased appetite, and depression. Despite all known treatment methods, the cat's condition worsened and ultimately died. According to Kuruçay and Gümüşova [15], veterinarians diagnose FIP primarily based on the age of the cat, where it came from, its breed, clinical findings, and the findings obtained because of the veterinarian's examination. They said that the symptoms of this disease are fever, visible swelling in the abdomen, and obvious deterioration of the anatomical structure of the eyes. They stated that antiviral drugs are used in the treatment of FIP and that these drugs are effective on patients, although not always.

FIP is a difficult disease to diagnose and treat due to its rapid progression and lack of widespread treatment. Early detection of FIP can increase the cat's survival rate, but it is costly and difficult to reach. Cat owners must be aware of symptoms and closely monitor their furry friends to avoid financial and moral burdens. Currently, no study has been conducted on early detection of FIP, but Şenol et al. [24] investigated trends in detecting epidemics using machine learning algorithms. This highlights the importance of early detection in addressing the challenges faced by cat owners in detecting diseases like FIP.

3 Material and Method

3.1 Naive Bayes Algorithm

Classification algorithms are popular data mining techniques that are widely used in the medical field, such as medical science and disease prediction. These algorithms group data based on similarities and brought together similar data sets. The Naïve Bayes algorithm, based on Bayes' conditional probability theorem, is a probabilistic classification method that uses independent assumptions between features. It is easy to model and performs better than other algorithms, making it more preferred in medical science.

Naive Bayes is a simple and powerful algorithm for predictive modeling. The model comprises two types of probabilities that can be calculated directly from the training data: (i) the probability of each class and (ii) the conditional probability for each class given each x value. Once calculated, the probability model can be used to make predictions for new data using Bayes theorem. When the data is real valued, it is common to assume a Gaussian distribution (bell curve) so that one can easily estimate these probabilities. Naive Bayes is called naive because it assumes that each input variable is independent. This is a strong assumption and unrealistic for real data; however, the technique is very effective on a large range of complex problems. The thought behind naive Bayes classification is to try to classify the data by maximizing $P(\mathrm{X}|\mathrm{C}i)P(\mathrm{C}i)$ using Bayes theorem of posterior probability (where X is the tuple in a dataset and "i" is an index of the class).

The working principle of the Naïve Bayes algorithm is a simple probability classifier that calculates a probability set by counting the frequency and combinations of values in each dataset. It assumes that all variables are independent, considering the value of the class variable. The conditional independence theorem allows for the calculation of conditional probability of each random parameter only for a given class label, rather than calculating conditional class probability for all combinations of random parameters.

The Naive Bayes classification method can be used in many areas, but the key is how it is classified. The probabilistic ratio between text and number data is crucial, and the classification algorithm's classification method is crucial in determining the classification accuracy. The mathematical representation of the Naive Bayes theorem is as follows:

$$P(Ci|X\ ) = P(X|Ci) * P(Ci)P(X)$$

3.2 Data Set

The data set created in this study was taken from two sources. One of the sources is the Ümraniye Vita Veterinary Clinic, which continues to care for its patients today, located in the Ümraniye district of Istanbul. From this veterinary clinic, 150 data were obtained on the symptoms of the disease in cats that did not have non-infectious peritonitis but were still sick in cats who visited the clinic. Another source is the symptoms of FIP disease in the article published by Jones, et al. [11]. From this data obtained, 300 uninfected peritonitis symptom data of cats were obtained. Due to the limited number of data obtained, all 450 data obtained were used. This data set consists of both the data of the sick cats with FIP disease detection and the cats who are sick but not diagnosed with FIP disease. These data are both visible signs of FIP disease and symptoms of diseases other than FIP disease. For classification, 16 features were examined. These features include fever, gait disturbance, weakness, swelling in the abdomen, weakness, refusal to eat, difficulty in breathing, decreased appetite, avoidance behavior, runny nose, sneezing,

Table 1. Sample data set.

Fever (1,0)	0: No 1:Yes
Walking Disorder (1,0)	0: No 1:Yes
Weakness (1,0)	0: No 1:Yes
Swelling In the Abdominal Area (1,0)	0: No 1:Yes
Rejecting The Meal (1,0)	0: No 1:Yes
Dyspnea (1,0)	0: No 1:Yes
Reduced Appetite (1,0)	0: No 1:Yes
Avoid Behavior (1,0)	0: No 1:Yes
Runny Nose (1,0)	0: No 1:Yes
Sneeze (1,0)	0: No 1:Yes
Shedding Of Feathers (1,0)	0: No 1:Yes
Red Skin (1,0)	0: No 1:Yes
Constipation (1,0)	0: No 1:Yes
Vomiting (1,0)	0: No 1:Yes
Diarrhea (1,0)	0: No 1:Yes

moult, and skin rash. With these data, it is planned to detect FIP disease with the Naive Bayes algorithm. The attributes and attribute values of the data are as given in Table 1:

The data set is very important for machine learning algorithms to work properly. For machine learning algorithms to work well, the distribution of the data set must be precise and free from bias. The equilibrium mentioned here is used in the sense that the number of labels fed with the dataset is close or equal to perform machine learning. An unbalanced data set is a situation where the number of observations in one cluster is greater or less than that of the other; and this was the case in the dataset collected for this study. At first, the algorithm over-learned because the dataset mostly consisted of cats with the FIP disease. For this reason, the Synthetic Minority Oversampling (SMOTE) method was used to correct this error and make the machine work properly.

With SMOTE, the goal is to develop a disease-free label synthetically, correcting the imbalanced dataset, and allowing it to perform more accurate machine learning. In short, the real purpose of the algorithm used is to enlarge and multiply the data smaller than the other label class to equal the amount of data larger than the other. In other words, it is a synthetic sampling method that uses interpolation to generate new samples. According to Harman [9], the SMOTE algorithm is based on the k-NN nearest neighbor algorithm and assume that an example of composite data can be interpolated between the original and one of the neighbors. Harman states that the working principle of the SMOTE algorithm is that each data sample in the data set, where one beacon class is lower than the other, will consider its neighborhood, randomly choosing one of its neighbor classes and generate composite data by interpolating the data between each sample and the selected nearest neighbor.

In the dataset used in the study, there were 300 datasets containing symptoms of feline infectious peritonitis (FIP) and 150 progressive diseases other than feline infectious peritonitis (FIP) but not must be infectious peritonitis in cats. (FIP). In other words, a total of 450 data was used. There is an imbalance in the data set because more cats have feline infectious peritonitis (FIP) than uninfected cats.

To balance the dataset, the Smote algorithm uses the data of 240 cats with infectious peritonitis (FIP) and 240 other progressive diseases with feline infectious peritonitis (FIP), i.e., diseases There was no feline infectious peritonitis (FIP). In other words, a balance sheet was created with a total of 480 data. In subsequent studies, this balance shifted with 215 cats with feline infectious peritonitis (FIP) and 430 data using progressive diseases other than feline infectious peritonitis (FIP), i.e., diseases without feline infectious peritonitis (FIP).

4 Findings and Discussion

There are some parameters to be obtained in machine learning algorithms. These parameters are considered when evaluating the learning performance of the machine. These parameters are precision, recall, f1- score and accuracy, which are defined as follows:

Precision. The percentage of samples that were positive in the data set used by samples that the classification model used predicted positive.

$$\text{Precision} = (\text{True Positive})/(\text{True Positive} + \text{False Positive})$$

Recall. A metric that measures how many of the true positive samples the classification model used can accurately detect. The Recall formula is as follows:

$$\text{Recall} = (\text{True Positive})/(\text{True Positive} + \text{False Negative})$$

F1-Score. It is a criterion that represents the balance of precision and recall parameters of the classification model used. F-1 score is calculated with the following formula.

$$\text{F1 - Score} = 2 * (\text{Precision} * \text{Recall})/(\text{Precision} + \text{Recall})$$

Support. It refers to the number of samples in the dataset of each class in the used classification model. In other words, support is a numerical value that represents the sample number of samples in each class.

Accuracy. A performance scale that measures how accurately a model can make predictions. It is generally used in classification problems and is calculated as the ratio of correctly guessed samples to the total number of samples.

$$\text{Accuracy} = (\text{Number of Correctly Estimated Samples}) \\ /(\text{Total Number of Samples})$$

In this model, the data set is divided into 80% training and 20% testing sets. The reason is that the data set we used was not enough at first, an imbalance occurred, and then we tried to eliminate this imbalance by generating aggregated data. The algorithm is over-fitted because more cats develop infectious peritonitis (FIP) from cats with the 'Sick' label than from cats with the 'Not Sick' label with no infectious peritonitis (FIP) but have other disease markers, and therefore no steady state. When overfitting, the created model learns the details of the data set and the noise caused by the redundancy, thus having difficulty in capturing the exact patterns. The result of this output is as follows (Table 2):

Table 2. Result Obtained Before Using SMOTE Algorithm

	precision	recall	F1-score	support
NOT SICK	1.00	1.00	1.00	31
SICK	1.00	1.00	1.00	59
accuracy			1.00	90
macro avg	1.00	1.00	1.00	90
weighted avg	1.00	1.00	1.00	90
Model Accuracy Rate	**1.00**			

As can be observed in this result, it is seen that the number of samples that are not sick is 31, but the number of samples that are sick is 59, and the result is 1.00, that is, 100% due to this imbalance. The result being 100% shows us that there is an overfitting in this model. The smote algorithm was used to correct this excessive learning.

Another finding was the production of synthetic data with the smote algorithm. To prevent over-learning, the accuracy of the model was found to be 0.96, because of the smote algorithm, which used data replication. The output of the model is as follows (Table 3):

Table 3. Result Obtained After Using SMOTE Algorithm

	precision	recall	F1-score	support
NOT SICK	1.00	0.93	0.96	240
SICK	0.93	1.00	0.97	240
accuracy			0.96	480
macro avg	0.97	0.96	0.96	480
weighted avg	0.97	0.96	0.96	480
Model Accuracy Rate	**0.96**			

After generating synthetic data with the Smote algorithm, the number of samples that were not sick and the number of samples that were sick were balanced with 240 samples. After this equilibrium condition, 96% accuracy was obtained.

5 Conclusion

Cats, like humans, get sick and their health is threatened by many different viruses. One such disease is infectious peritonitis in cats, known as FIP in short. Feline infectious peritonitis (FIP) is a disease caused by mutated coronaviruses found in cats. This disease attacks a cat's immune system, affecting their gait, appetite, breathing, in short, all their vital activities. There are two types of this disease. These two categories are divided into wet and dry FIP. Both species have distinct characteristics. We consider these characteristics to be symptoms of illness. Although there is no known single treatment today for infectious peritonitis in cats, the vaccines used to treat FIP are very expensive. Since the vaccine is not a single dose, the course of treatment is highly dependent on our finances. Unfortunately, because the treatment is so expensive, not everyone can afford it. Considering all this, we can say that FIP is a deadly disease in cats. Veterinarians express the importance of early diagnosis to prevent this disease and save our beautiful friend's life. Since cats cannot express changes and problems in their bodies, cat owners need to notice changes and act immediately. This step is about noticing the change and taking your adorable friend to the vet. When diagnosing FIP, your veterinarian will first look at the symptoms and then diagnose whether the blood test results support those symptoms. They also started treatment after feline infectious peritonitis was diagnosed. As mentioned earlier, the treatment period will tire the cat and its owner, and if this treatment does not produce satisfactory results, the animal will be very painful to lose. This study is intended to help cat owners diagnose FIP early.

This study provides early diagnosis of feline infectious peritonitis (FIP) using the Naïve Bayes algorithm. Another goal of this study is to raise awareness of feline infectious peritonitis (FIP). Machine learning algorithms are used in the diagnosis of many diseases such as heart disease and diabetes. However, this technology is not used in the diagnosis of animal diseases. It can be said that this study is the first in this regard. In the study, Naive Bayes algorithm, one of the machine learning classification algorithms, is used. Naive Bayes algorithm is based on Bayes theorem developed by Naive Bayes. Classification algorithms are preferred as the model used in this study because they are often used in disease diagnosis and Naive Bayes algorithm is one of the algorithms giving the best results because of its simple operating principle. simple. The Naive Bayes algorithm model is written in the Python programming language. The reason to use Python programming language is that it contains many different libraries and prototyping is very easy with these libraries. After the model was created, the dataset was tested in that model, as described earlier and it is resampled using Smote.

Experiments were made with the data set, where 80% of the data set is reserved for training and 20% for testing. In this way, the training of the model was kept at the highest level. The Smote method is used for synthetic data generation, and as a result, the model successful predicted the FIP cases with 96% accuracy. The results show promising classification of FIP disease using the Naive Bayes algorithm.

References

1. Aytuğ, N.: Kedi enfeksiyonları 1: Zorlayan tanı; kedilerin enfeksiyöz peritonitisi (2009)
2. Baydan, B.: Derin öğrenme ile köpek ve kedi tibia kemikleri üzerindeki kırıkların tespiti ve sınıflandırması (2021)
3. Baydar, E., Eröksüz, Y., Timurkan, M.Ö., Eröksüz, H.: Feline infectious peritonitis with distinct ocular involvement in a cat in Turkey. Kafkas Üniversitesi Veteriner Fakültesi Dergisi **20**(6), 961–965 (2014)
4. Börkü, M.K.D.H.: Kedi-köpek enfeksiyöz hastalıkları (iç hastalıkları)
5. Camkıranlar, M.: KEDİ ENFEKSİYÖZ PERİTONİTİSİ DOĞAL ENFEKSİYONLARINDA PATOLOJİK ve İMMUNOHİSTOKİMYASAL İNCELEMELER (Master's thesis, Aydın Adnan Menderes Üniversitesi Sağlık Bilimleri Enstitüsü) (2019)
6. Candan, H., Durmuş, A., Harman, G.: Genetik Algoritma ve Sınıflandırıcı Yöntemler ile Kanser Tahmini. Veri Bilimi **2**(1), 30–34 (2019)
7. Coşar, M., Deniz, E.: Makine Öğrenimi Algoritmaları Kullanarak Kalp Hastalıklarının Tespit Edilmesi. Avrupa Bilim ve Teknoloji Dergisi **28**, 1112–1116 (2021)
8. Golab, M.: Kedilerin kökeni: kediler nereden geldiler? Neden insanlarla yaşamayı seçtiler?
9. Harman, G.: Destek Vektör Makineleri ve Naive Bayes Sınıflandırma Algoritmalarını Kullanarak Diabetes Mellitus Tahmini. Avrupa Bilim ve Teknoloji Dergisi (32), 7–13 (2021)
10. Hartmann, K.: Feline infectious peritonitis. Vet. Clin. Small Anim. Pract. **35**(1), 39–79 (2005)
11. Jones, S., Novicoff, W., Nadeau, J., Evans, S.: Unlicensed GS-441524-like antiviral therapy can be effective for at-home treatment of feline infectious peritonitis. Animals (Basel) **11**(8), 2257 (2021). https://doi.org/10.3390/ani11082257. PMID: 34438720; PMCID: PMC8388366
12. Kahraman, M.A.: Feline infeksiyöz peritonitli kedilerde adenozin deaminaz ve C reaktif proteinin diagnoztik önemi (Master's thesis, Sağlık Bilimleri Enstitüsü)

13. Karayiğit, M.Ö., Aydoğdu, U., Başbuğ, O., Dörtbudak, B., Karataş, Ö., Tuzcu, M.: Scottish Fold Irkı Bir Kedide Feline İnfeksiyöz Peritonitis Olgusu. Cumhuriyet Üniversitesi Sağlık Bilimleri Enstitüsü Dergisi **1**(1), 46–50 (2016)
14. Kurban, M.Y.: Kedilerin enfeksiyöz peritonitisinin patolojik, sitolojik ve immunohistokimyasal yöntemlerle araştırılması (Master's thesis, Balıkesir Üniversitesi Sağlık Bilimleri Enstitüsü) (2021)
15. Kuruçay, H.N., Gümüşova, S.: Feline İnfeksiyöz Peritonitise Genel Bakış ve Antiviral Yaklaşımlar. Turkish Veterinary Journal **3**(1), 4–12 (2021)
16. Maden, M., Çuhadar, F.: Kedi ve Köpeklerde Endokrin Hastalıkların Laboratuvar Tanısı. Turkiye Klinikleri J Vet Sci **4**(3), 35–60 (2013)
17. Özlem, A.K.A.R., Güngör, O.: Rastgele orman algoritması kullanılarak çok bantlı görüntülerin sınıflandırılması. Jeodezi ve Jeoinformasyon Dergisi **106**, 139–146 (2012)
18. Özlüer Başer, B., Yangın, M., Sarıdaş, E.S.: Makine Öğrenmesi Teknikleriyle Diyabet Hastalığının Sınıflandırılması. Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi **25**(1), 112–120 (2021). https://doi.org/10.19113/sdufenbed.842460
19. Özmen, Ö., Khdr, A., Avcı, E.: Sınıflandırıcıların Kalp Hastalığı Verileri Üzerine Performans Karşılaştırması. Fırat Üniversitesi Mühendislik Bilimleri Dergisi , Fırat Üniversitesi Mühendislik Bilimleri Dergisi 153–159 (2018)
20. Öztürk, D.: Feline enfeksiyöz peritonit şüpheli kedilerde tümör nekroz faktör, asetilkolinesteraz ve diğer biyokimyasal parametrelerin değerlendirilmesi (Master's thesis, Balıkesir Üniversitesi Sağlık Bilimleri Enstitüsü) (2022)
21. Potur, E.A., Erginel, N.: Kalp Yetmezliği Hastalarının Sağ Kalımlarının Sınıflandırma Algoritmaları ile Tahmin Edilmesi. Avrupa Bilim ve Teknoloji Dergisi **24**, 112–118 (2021)
22. Saritas, M.M., Yasar, A.: Performance analysis of ANN and Naive Bayes classification algorithm for data classification. Int. J. Intell. Syst. Appl. Eng. **7**(2), 88–91 (2019)
23. Soylu, H.: Kedi ve Köpeklerde Kardiyovasküler Hastalıklarda Koenzim Q10 Kullanımı. Cumhuriyet Üniversitesi Sağlık Bilimleri Enstitüsü Dergisi **4**(1), 40–44 (2019)
24. Şenol, A., Canbay, Y., Kaya, M.: Makine Öğrenmesi Yaklaşımlarını Kullanarak Salgınları Erken Evrede Tespit Etme Alanındaki Eğilimler. Bilişim Teknolojileri Dergisi **14**(4) (2021)
25. Taşçi, M.E., Şamli, R.: Veri madenciliği ile kalp hastalığı teşhisi. Avrupa Bilim ve Teknoloji Dergisi 88–95 (2020)
26. Vembandasamy, K., Sasipriya, R., Deepa, E.: Heart diseases detection using Naive Bayes algorithm. Int. J. Innov. Sci. Eng. Technol. **2**(9), 441–444 (2015)
27. Yilmaz, Z., Kennerman, E., Şentürk, S., Temizel, M., Aytuğ, N.: Uludağ üniversitesi veteriner fakültesi iç hastalıkları küçük hayvan kliniğine getirilen kedi ve köpeklerin değerlendirilmesi (1990-2000). Uludag Univ. J. Fac. Vet. Med. **21**, 23–31 (2002)

A Numerical Simulation for the Ankle Foot Orthosis Using the Finite Element Technique with the Aid of an Experimental Program

Maryam I. Abduljaleel[1], Muhsin J. Jweeg[2](✉), and Ahmed K. Hassan[3]

[1] Department of Mechanical Engineering, Faculty of Engineering, University of Kerbala, Karbala, Iraq
m09152086@s.uokerbala.edu.iq

[2] College of Technical Engineering, Al-Farahidi University, Baghdad, Iraq
muhsin.jweeg@uoalfarahidi.edu.iq

[3] Prosthetics and Orthotics Engineering Department, College of Engineering, University of Kerbala, Karbala, Iraq
dr.ahmed_kh74@uokerbala.edu.iq

Abstract. Ankle-foot orthosis (AFO) is a device that supports the ankle and foot part of the body when there is a muscle weakness or a nerve damage, Ankle-foot orthoses are prescribed to individuals with minimal spinal cord injury and excellent trunk muscle control (AFOs). In this paper, two types of composite laminates were used in the experimental program. The first sequence of layers arranged as follows, (2Perlon + 1Carbon fiber + 2Perlon + 1Kevlar + 2Perlon) called Sequence1. And the other sequence is (2Perlon + 3Carbon fiber + 2Perlon + 3Kevlar + 2Perlon) called Sequence2. In the numerical investigation, the performance of the AFO materials is evaluated using mechanical qualities such as fatigue and tensile testing. This study uses FEM as a numerical technique to demonstrate the impact of fatigue performance on a structural element with the assistance of ANSYS Workbench 14 software. It is used to predict how total deformation, maximum stress, fatigue life, and safety factor will behave. The experimental results have shown that the ultimate tensile stress was 67 MPa and 80 MPa for the first and second type of layers respectively. The patient's height (176 cm), weight (78 kg), and approximate age of 39 and he was suffered from drop foot. By using FEM (ANSYS) (Von-Mises), the equivalent stress and safety factor of the fatigue have been calculated for the provided AFO model. The resulting ANSYS findings are shown that the profiles of the fatigue safety factors for the composite material (sequence1) AFO equal to 2.86211, for the composite material (sequence2) AFO equal to 3.68318, and for the Polypropylene AFO equal to 1.90683. The difference between the yield stresses of composite material and the highest stresses is produced by the orthosis which indicates the feasibility of the notion that composite materials can support the patient's weight and serve as an alternative to the materials currently is employed to make the AFO. Where the highest stresses of the PP and composite material (sequence1) AFO are equal to 18.033 MPa, and for composite material (sequence 2) AFO is equal to 17.583 MPa and yield strength for composite material (sequence 1) (50 MPa), yield strength for composite material (sequence 2) (63 MPa) compared to polypropylene's yield stress of 24.3 MPa.

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 187–197, 2024.
https://doi.org/10.1007/978-3-031-56728-5_17

Keywords: AFO · polypropylene · composite material · ANSYS · drop foot

1 Introduction

The human foot plays a very important role in human movements such as standing, running, walking, jumping, etc. The human foot in general made up of three sections: the forefoot, the middle of the foot, and the hind foot. External biomechanical devices called ankle foot orthoses (AFOs) are used to stabilize joints, enhance gait, and restore physical function in lower limbs. The orthoses are manufactured either from a single material or several materials depending on the design and quality that planned to grant the orthoses strength, required elasticity, stiffness, durability, and wear resistance, acceptance by the customer, corrosion resistance, and cost are factors must take into regard when designing the orthoses [2]. The use of a single case design was used to assess how a dynamic AFO affected post-stroke hemiplegic ambulation. Using a dynamic AFO and walking on flat ground, one patient with stroke-related hemiplegia underwent gait analysis. The dynamic AFO improved the individual's overall gait, including temporal-spatial features and gait velocity, likely because the subject expended less energy during walking [3]. A review was conducted to clarify the best way to gauge the rigidity of an ankle-foot orthosis (AFO). This knowledge is crucial to ensuring proper orthotic intervention for individuals with abnormal gait. The two primary methods for examining AFO rigidity are (1) bench-testing analysis, in which an AFO is fastened to or connected to a measuring equipment, (2) functional analyses, which involve taking measures when a person is moving while wearing an AFO [3].

It was shown that ankle-foot orthoses can ameliorate impairments in the lower limb neuromuscular motor system that regulate gait (AFOs). Present AFO technologies involve active devices that provide power for the foot movement by combining a number of technologies, passive devices with articulated and fixed joints and semi active devices that adjust damping at the joint [4]. The foot drop was shown to be a common problem seen by medical rehabilitation professionals in clinical settings. Lower motor neurons, often as a result of issues with the lumbar spine's roots or with peripheral nerves, are to blame for foot drop. However, while examining such a patient, it is important to consider other uncommon differential diagnoses of foot drop. It is generally possible to distinguish between such causes with the use of a thorough neurological assessment and ancillary tests like magnetic resonance imaging and an electro diagnosis [5]. The effect of the position of the bending axis on a person's ability to walk when wearing a dynamic passive ankle orthosis and foot was presented in detail. Orthoses for the ankle and foot that are designed to be dynamically passive are frequently used to increase the task of injured ankle muscles. This research aims to quantify how different bending axes affect how well people walk. "The Passive Dynamic Ankle Foot Orthotics" have been produced using additive manufacturing for thirteen participants who have unilateral ankle muscle weakness in both the eccentric and central axes of flexion. While participants strolled normally on flat ground, kinematic histogram data was collected along three axes of curvature [6].

The different types of orthoses were presented to explain the use and importance of each type of orthoses. And explain that for a range of acute and chronic foot and ankle problems, foot orthoses and shoe modifications are a crucial part of no operative therapy procedures, the biomechanics, normal operation, and consequences of diseases affecting the foot and/or ankle should all be known to the orthopaedic foot and ankle surgeon [7]. The importance of using biopolymers and natural fibers were studied as alternatives for the future due to the increasing environmental awareness regarding carbon emissions and the depletion of petroleum resources. The researcher specialized in studying kenaf fibres and their different types and applications, as they are considered one of the most favourable natural fibers due to its suitable physical and mechanical properties [8].

Recently a research program suggested using three different designs of passive ankle foot orthosis to help a patient who has had ankle surgery to operate a vehicle, particularly during emergency braking. The researcher used a polyprophelene (PP) and carbon fiber for analyse the most suitability material for this special AFO. The researcher concluded that polyprophelene (PP) will shatter since the maximum stress is far higher than its yield strength, making it unsuitable for this particular AFO's role in absorbing impact during emergency braking. Carbon fiber, however, performs the reverse [9–13]. Three distinct styles of Ankle Foot Orthosis (AFO) were chosen among the products of his hands (adjustable hinge AFO, modified static AFO and modified night splint AFO). For this, he made use of: (polypropylene, steel, foam and Acrylonitrile butadiene styrene). To learn the outcomes of the (Von-Mises) stress and fatigue safety factor analysis, He performed a series of tests, including the F-Socket, Dynamic gait analysis, the gait cycle, and study of the AFO's models. In particular, he wanted to increase the ankle foot orthosis's resistance to fatigue while also enhancing its mechanical qualities. It was determined by the study's author that the tweaked static AFO When compared to the modified night splint AFO and the adjustable hinge AFO, it far outshone both in terms of aesthetics and wear ability [14–18]. Extensive research covers the modelling of the prostheses using different types of composite materials which can be used in manufacturing of the ortheses with different types of additives [19–24].

In this work, the finite element technique is used with the aid of an experimental data to create a novel lamination of AFO model. Finite element analysis is a numerical technique for making an approximate solution for variables in a problem that is challenging to solve analytically. Today, a wide range of engineering and scientific sectors employ the finite element method (FEM) extensively. Because of its ability to handle complicated non-linear material characteristics and geometrical boundaries, the approach is acknowledged as one of the most potent numerical techniques. The stress and deformation contours will be presented and the factor of safety is calculated accordingly.

2 The Numerical Analysis

Finite element analysis is a numerical technique for making an approximate solution for variables in a problem that is challenging to solve analytically. Today, a wide range of engineering and scientific sectors employ the finite element method (FEM) extensively. Because of its ability to handle complicated non-linear material characteristics and geometrical boundaries, the approach is acknowledged as one of the most potent numerical techniques [25–28].

The primary advantage of numerical solutions is the speedy analysis possible with modern computers that have vast amounts of memory and processing power**.** This study uses FEM as a numerical technique to demonstrate the impact of fatigue performance on a structural element with the assistance of ANSYS Workbench 14 software. It is used to predict how total deformation, maximum stress, fatigue life, and safety factor will behave [29–31].

Following are the three distinct stages of the general analysis carried out by ANSYS:

- Use the geometry to build a model.
- To find the solution, use the boundary condition load.
- Evaluating the results.

3 Experimental Work

The materials of the AFO needed for the current work are as follows (Fig. 1): Carbon, Kevlar, kenaf fibers, perlon, polyvinyalcohol (PVA), polypropylene and hardener.

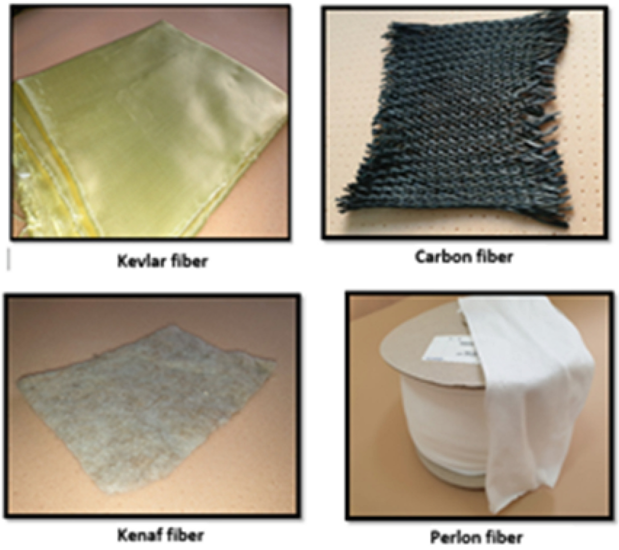


Fig. 1. Four types of reinforcement fibers

3.1 Specimens Preparation for Mechanical Properties

To determine the mechanical characteristics of the composite material, there should be some samples created from it. The best method for preparing these samples is the vacuum method is a preparation technique for composite specimens that prevents cavities or defects to ensure that the sample is successful and it is as follows:

1. A positive mold is made out of plaster that measures (40 × 20 × 10) cm and place a steel beam within it, as shown in Fig. 2.a.

2. The positive mold is made into the vacuum device, as shown in Fig. 2.b.
3. Using PVA to cover the positive mold, as shown in Fig. 2.c.
4. Arranging the layers within the specified sequence (perlon, carbon fiber and kevlar fiber), as shown in Fig. 2.d.e.f.
5. Using another PVA to cover the positive mold.
6. Adding the hardener to the lamination
7. Lamination is poured from the top of the opening PVA on the Materials while the vacuum device is operating, as shown in Fig. 2.g.
8. After pouring the Lamination, the opening PVA is closed directly to prevent air from entering
9. Waiting 60 min for the mold to harden
10. The composite material specimen should be a rectangular shape, as shown in Fig. 2.h. After that, for the purpose of conducting the tests (tensile and fatigue tests), the mold is cut by using a CNC machine and according to ASTM standard.

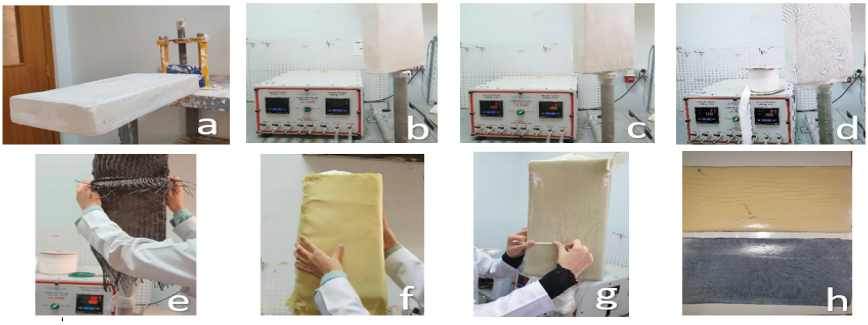


Fig. 2. Steps of preparing composite material specimens

Two new AFO models were made based on the dimensions of the drop foot in the patient in this study:

1. Ankle-Foot-Orthosis (AFO) from composite materials (2Perlon + 1Carbon fiber + 2Perlon + 1Kevlar + 2Perlon) for the right foot, as shown in Fig. 3
2. Plastic ankle-foot orthosis (polypropylene) for the right foot, as shown in Fig. 4.

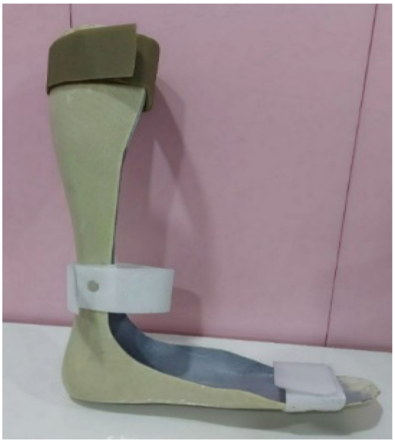


Fig. 3. Ankle-Foot-Orthosis (AFO) from composite materials

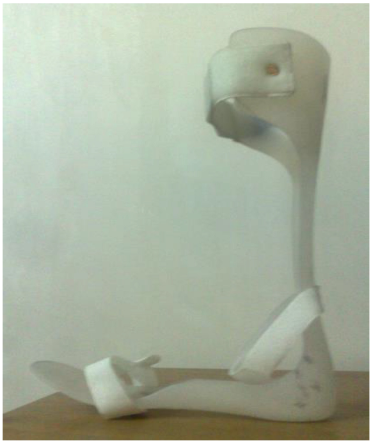


Fig. 4. Plastic ankle-foot orthosis (polypropylene)

4 Results and Discussion

In order to know the outcomes of the study of (Von-Mises) stress and the safety factor of fatigue. Six different Sequences were tested and only two Sequences (Sequences1 and Sequences2) were used in thesis based on the results of the examinations of the samples. The Sequences1 and the Sequences2 were analysed in the ANSYS program. The AFO is made from the Sequence1 because its cost is lower and its results extracted in the ANSYS program are close to the results of the Sequences2.

4.1 The Boundary Condition for Analysis of the Orthosis

In a typical analysis using ANSYS, there are three different steps: creating the geometry as a model in step one, applying the boundary conditions load in step two to get the solution, and reviewing the results in step three. To analyse the model of the orthosis and determine its Von Miss Stress and deformation, the boundary condition must be used. The boundary condition for the AFO for the right leg includes applying the observed interface pressure values at various areas for the Polypropylene, the composite material (sequence 1) and the composite material (sequence 2) as demonstrated in the Figs. 5 and 6.

To complete the overall findings of this model, the mechanical characteristics of each set of composite material properties are added to the ANSYS data.

4.2 Von Mises Analysis

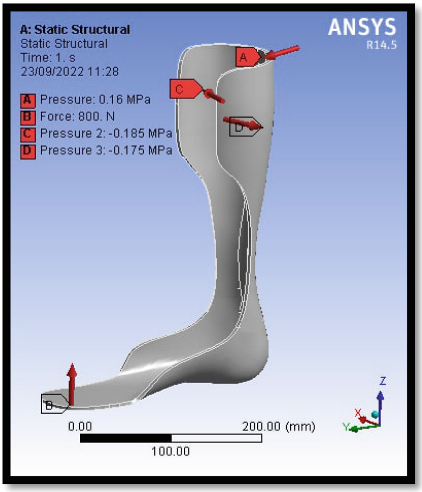


Fig. 5. The boundary conditions applied on the right AFO for the Polypropylene material and the composite material (sequence 1).

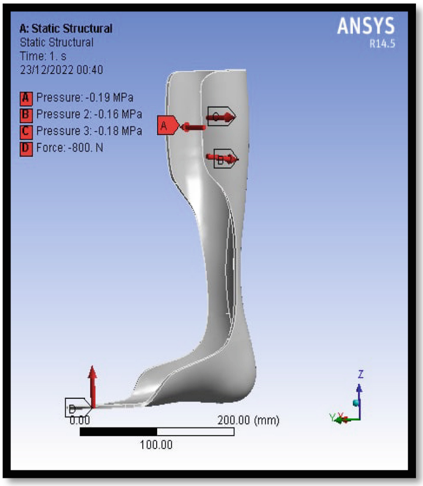


Fig. 6. The boundary conditions applied on the left AFO for the composite material (sequence 2)

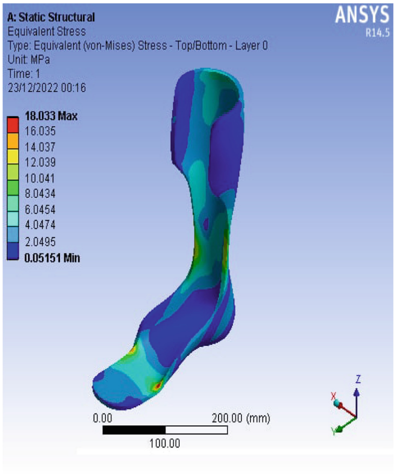


Fig. 7. Analysis of Von Mises stress of composite material (sequence1) AFO.

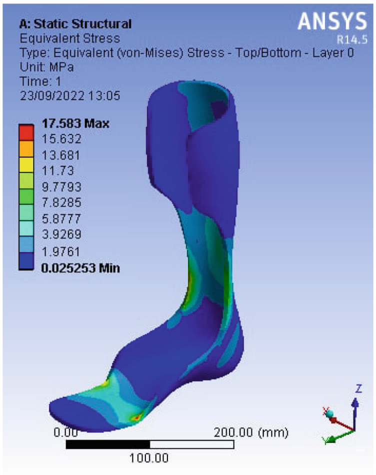


Fig. 8. Analysis of Von Mises stress of the composite material (sequence2) AFO.

The numerical analysis is carried out to determine the stresses generated in the components of the AFO as a result of the pressure created at the interface between the orthosis and the patient's muscles and body weight during walking. The results showed that the maximum value of stress generated in the PP and composite material (sequence1)

AFO are equal to 18.033 MPa as shown in Fig. 7, and for composite material (sequence 2) AFO is equal to 17.583 MPa as shown in Fig. 8. It should be noted that the difference between the yield stresses of composite material and the highest stresses produced by the orthosis indicates the feasibility of the notion that composite materials can support the patient's weight and serve as an alternative to the materials currently employed to make the AFO. Additionally, the static stress study suggests that composite material is the optimum material for making orthoses since there is a significant difference between its maximum stress and yield strength for composite material (sequence 1) (50 MPa), yield strength for composite material (sequence 2) (63 MPa) compared to polypropylene's yield stress of 24.3 MPa.

4.3 The Numerical Analysis of Deformation

The deformation study revealed the magnitude and location of the AFO's overall deformation. The maximum deformation value of the PP AFO equal to 4.1247 mm as shown in Fig. 9, for the composite material (sequence1) AFO equal to 3.7943 mm as shown in Fig. 10, and for the composite material (sequence2) AFO equal to 3.4823 mm as shown in Fig. 11. It should be noted that the AFO's deformation values are acceptable when made of composite materials since the orthosis must deform within the range of the aforementioned values when the patient's skin is under pressure from the interface.

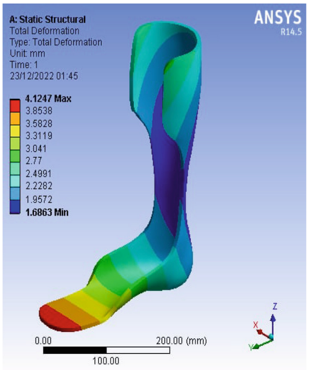


Fig. 9. Deformation analysis of the PP AFO.

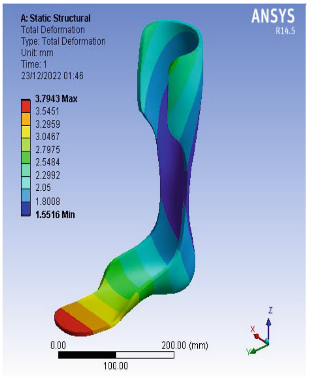


Fig. 10. Deformation analysis of the composite material (sequence1) AFO.

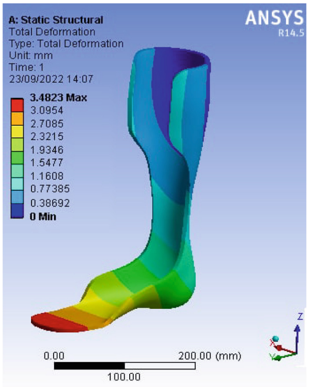


Fig. 11. Deformation analysis of the composite material (sequence2) AFO.

4.4 Safety Factor

FEM software was used to analyse the AFO model and calculate the safety factor for fatigue.

The safety factor for the suggested composite material has been approved for design; however, it should be noted that the value of the safety factor changes from area to region based on the distribution of the produced stresses and the endurance stress for each class of composite materials.

Each colour represents a particular gradient of safety factor values.

The values of safety factor for PP and the composite material (sequence1) AFO is equal to (1.90683, 2.86211), respectively as shown in Figs. 12 and 13, and for the composite material (sequence2) AFO equal to 3.68318 as shown in Fig. 14**.**

If the safety factor is equal to or greater than (1.25), the fatigue safety factor will be safe in design.

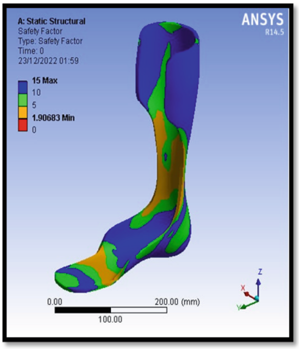


Fig. 12. Safety factor for PP

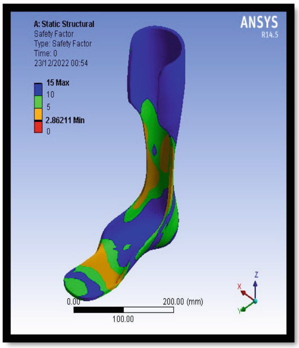


Fig. 13. Safety factor for (sequence 1)

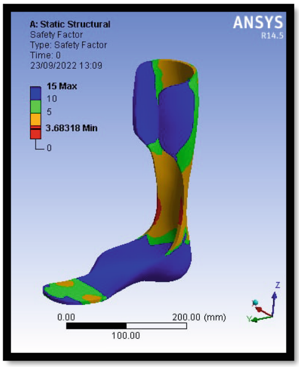


Fig. 14. Safety factor for (sequence 2)

5 Conclusions

1. The composite material ankle-foot orthosis showed great performance in the safety factor for fatigue and equivalent Von-Mises stress, which resulted in the longer life design.
2. The fatigue safety factor for layers made of (2Perlon + 1Carbon fiber + 2Perlon + 1Kevlar + 2Perlon) equals to (2.86211), for layers made of (2Perlon + 3Carbon fiber + 2Perlon + 3Kevlar + 2Perlon) equals to (3.68318), and for layers made of plastic material (Polypropylene) equal to (1.90683), which are safe in design because it is greater than (1.25).
3. The difference between the yield stresses of composite material and the highest stresses is produced by the orthosis which indicates the feasibility of the notion that composite materials can support the patient's weight and serve as an alternative to the materials currently is employed to make the AFO. Where the highest stresses of the PP and composite material (sequence1) AFO are equal to 18.033 MPa, and for composite material (sequence 2) AFO is equal to 17.583 MPa and yield strength for composite material (sequence 1) (50 MPa), yield strength for composite material (sequence 2) (63 MPa) compared to polypropylene's yield stress of 24.3 MPa

References

1. Yamamoto, S., Ebina, M., Iwask, M.: Comparative study of mechanical characteristics of plastic AFOs. Prosthet Orthotic, pp. 59–64 (1993)
2. Kenneth, A.A., Jicheng, X.T., Elizabeth, H.W., William, K.D., G'eza, F.K.: Technologies for powered ankle-foot orthotic systems: possibilities and challenges. IEEE/ASME Trans. Mechatron. (2011)
3. Philippe, D.P.F.: Development of a two-dimensional biomechanical multibody model for the analysis of the human gait with an ankle-foot orthosis. Master's Thesis in Biomedical Engineering Biomaterials, Biomechanics and Rehabilitation University of Minho (2012)
4. Krishnaprasad, S.A.: Arnold Chiari Malformation with Holocord Syringomyelia Presenting as Unilateral Foot Drop: A Case Report, Government medical college, Kozhikode, Kerala (2012)
5. Jumaa, S.C.: Vibration analysis and measurement in knee ankle foot orthosis for both metal and plastic KAFO type. In: ASME International Mechanical Engineering Congress and Exposition, American Society of Mechanical Engineers (2013)
6. Ranz, E.C.: The influence of passive-dynamic anklefoot orthosis bending axis location on gait performance in individuals with lower-limb impairments. Clinical, pp. 13–21 (2016)
7. Osama, E., Tyler, S., Adam, F., Daniel, F., Keith, W.: Uses of braces and orthotics for conservative management of foot and ankle disorders. American Orthopaedic Foot & Ankle Sociaty (2018)
8. Dulina, T., et al..: Kenaf fiber composites: a review on synthetic and biodegradable polymer matrix. Jurnal Kejuruteraan (2019)
9. Daud, R., Daud, M., Ayu, H.M., Shah, A.: Design & analysis of ankle foot orthosis for assisting car driver after ankle surgery. Int. J. Eng. Trends Technol. (IJETT) (2020)
10. Ali, S.K.: Enhancement of Mechanical Properties and Fatigue life of the Ankle Foot Orthosis (AFO), A Thesis Submitted to the Mechanical Engineering Department/University of Technology in a Partial Fulfillment of the Requirements for the Degree of Master of Science in Mechanical Engineering (2021)
11. Rolf, M.N., Jorunn, L.H.: Estimation of gait cycle characteristics by trunk accelerometry. J. Biomech. 121–126 (2004)
12. Jweeg, M.J., Ameen, S.H.: Experimental and theoretical investigations of dorsiflexion angle and life of an ankle-Foot-Orthosis made from (Perlon-carbon fibre-acrylic) and polypropylene materials. In: 10th IMEKO TC15 Youth Symposium on Experimental Solid Mechanics (2011)
13. Jweeg, M.J., Resan, K.K., Ismail, M.T.: Study of creep-fatigue interaction in a prosthetic socket below knee. In: ASME International Mechanical Engineering Congress and Exposition (2012)
14. Jweeg, M.J., Alhumandy, A.A., Hamzah, H.A.: Material characterization and stress analysis of openings in syme's prosthetics. Int. J. Mech. Mechatron. Eng. IJMME-IJENS **17**(04) (2017)
15. Jweeg, M.J., Hammoudi, Z.S., Alwan, B.A.: Optimised analysis, design, and fabrication of trans-tibial prosthetic sockets. In: IOP Conference Series: Materials Science and Engineering, 2nd International Conference on Engineering Sciences, vol. 433 (2018)
16. Jweeg, M.J., Ahumdany, A.A., Mohammed Jawad, A.F.: Dynamic stresses and deformations investigation of the below knee prosthesis using CT-Scan modeling. Int. J. Mech. Mechatron. Engineering IJMME-IJENS **19**(01) (2019)
17. Jweeg, M.J., Hammood, A.S., Al-Waily, M.: Experimental and theoretical studies of mechanical properties for reinforcement fiber types of composite materials. Int. J. Mech. Mechatron. Eng. IJMME-IJENS **12**(04) (2012)
18. Abbas, S.M., Takhakh, A.M., Al-Shammari, M.A., Al-Waily, M.: Manufacturing and analysis of ankle disarticulation prosthetic socket (SYMES). Int. J. Mech. Eng. Technol. (IJMET) **09**(07), 560–569 (2018)

19. Jweeg, M.J., Al-Waily, M., Muhammad, A.K., Resan, K.K.: Effects of temperature on the characterisation of a new design for a non-articulated prosthetic foot. In: IOP Conference Series: Materials Science and Engineering, vol. 433, 2nd International Conference on Engineering Sciences, Kerbala, Iraq, 26–27 March 2018
20. Al-Waily, M., Hussein, E.Q., Al-Roubaiee, N.A.A.: Numerical modeling for mechanical characteristics study of different materials artificial hip joint with inclination and gait cycle angle effect. J. Mech. Eng. Res. Dev. (JMERD) **42**(04), 79–93 (2019)
21. Hussein, S.G., Al-Shammari, M.A., Takhakh, A.M., Al-Waily, M.: Effect of heat treatment on mechanical and vibration properties for 6061 and 2024 aluminum alloys. J. Mech. Eng. Res. Dev. **43**(01), 48–66 (2020)
22. Abbas, E.N., Jweeg, M.J., Al-Waily, M.: Fatigue characterization of laminated composites used in prosthetic sockets manufacturing. J. Mech. Eng. Res. Dev. **43**(5), 384–399 (2020)
23. Abbas, E.N., Al-Waily, M., Hammza, T.M., Jweeg, M.J.: An investigation to the effects of impact strength on laminated notched composites used in prosthetic sockets manufacturing. In: IOP Conference Series: Materials Science and Engineering. 2nd International Scientific Conference of Al-Ayen University, vol. 928 (2020)
24. Al-Waily, M., Tolephih, M.H., Jweeg, M.J.: Fatigue characterization for composite materials used in artificial socket prostheses with the adding of nanoparticles. In: IOP Conference Series: Materials Science and Engineering, 2nd International Scientific Conference of Al-Ayen University, vol. 928 (2020)
25. Kadhim, A.A., Abbod, E.A., Muhammad, A.K., Resan, K.K., Al-Waily, M.: Manufacturing and analyzing of a new prosthetic shank with adapters by 3D printer. J. Mech. Eng. Res. Dev. **44**(3), 383–391 (2021)
26. Jebur, Q.H., Jweeg, M.J., Al-Waily, M., Ahmad, H.Y., Resan, K.K.: Hyperelastic models for the description and simulation of rubber subjected to large tensile loading. Arch. Mater. Sci. Eng. **108**(2), 75–85 (2021)
27. Al-Waily, M., Jweeg, M.J., Jebur, Q.H., Resan, K.K.: Creep characterization of various prosthetic and orthotics composite materials with nanoparticles using an experimental program and an artificial neural network, Materials Today: Proceedings (2021)
28. Haider, S.M.J., Takhakh, A.M., Al-Waily, M.: A review study on measurement and evaluation of prosthesis testing platform during gait cycle within sagittal plane. In: 14th International Conference on Developments in eSystems Engineering, IEEE Xplore (2021)
29. Haider, S.M.J., Takhakh, A.M., Al-Waily, M., Saadi, Y.: Simulation of gait cycle in sagittal plane for above-knee prosthesis. In: 3rd International Scientific Conference of Alkafeel University, AIP Conference Proceedings, vol. 2386 (2022)
30. Jweeg, M.J., Alazawi, D.A., Jebur, Q.H., Al-Waily, M., Yasin, N.J.: Hyperelastic modelling of rubber with multi-walled carbon nanotubes subjected to tensile loading. Arch. Mater. Sci. Eng. **114**(2), 69–85 (2022)
31. Haider, S.M.J., Takhakh, A.M., Al-Waily, M.: Designing a 3D virtual test platform for evaluating prosthetic knee joint performance during the walking cycle. Open Eng. **12**, 590–604 (2022)

Numerical and Experimental Simulations of Damage Identification in Carbon/Kevlar Hybrid Fiber-Reinforced Polymer Plates Using the Free Vibration Measurements

Dhia A. Alazawi[1], Muhsin J. Jweeg[2](✉), and Mohammed J. Abbas[1]

[1] Department of Mechanical Engineering, Faculty of Engineering, University of Diyala, Baqubah, Iraq
{dhiaahmed_eng,eng_grad_mech035}@uodiyala.edu.iq
[2] College of Technical Engineering, Al-Farahidi University Iraq, Baghdad, Iraq
muhsin.jweeg@uoalfarahidi.edu.iq

Abstract. Damage identification is an essential program in monitoring the health structures of the mechanical, civil, and aeronautical structures. The vibration measurements are one of the techniques used in this respect on the basis that the stiffness is reduced due to any defect and therefore, the natural frequency is reduced correspondingly. In this work, a rectangular composite cantilever plate was designed and analysed using the ANSYS package employing an experimental data program. Vibration sensors were used sensor used to detect the effect of crack length in carbon/Kevlar hybrid fiber-reinforced polymer composites, and a numerical analysis was performed via Ansys software to examine different crack scenarios. The effect of damage length was experimentally detected in carbon/Kevlar hybrid fiber-reinforced polymer composites using the sensor vibration measurements in the present work. The Finite element analysis (FEA) was applied to determine the most significant notch length in the carbon/Kevlar hybrid fiber-reinforced polymer composite structure. The investigation was performed at several different lengths and locations, including vertical and horizontal incisions. The experimental were agreed with those obtained numerically with a discrepancy not more than 5% which proves that the numerical cheap solution is an adequate in predicting failure mode of the structure and can be used to health monitoring of composite structures.

Keywords: composites · Natural frequency · FEA · Defect · Cantilever plate

1 Introduction

Composite panels are widely used in a variety of industries due to their high strength-to-weight ratio and good mechanical qualities. However, they are vulnerable to damage from collisions, fatigue, and environmental deterioration. Dynamic analysis is a common approach for evaluating the health of composite panels, using the theoretical and experimental techniques to create effective health monitoring strategies. Theoretical research

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 198–211, 2024.
https://doi.org/10.1007/978-3-031-56728-5_18

involves creating mathematical models that characterize the behavior of composite panels under various loading circumstances, which can be used as a baseline for health monitoring and to mimic the reaction of the panels to various forms of damage.

The damage detection in carbon fiber reinforced plastics were investigated numerically and experimentally using vibration measurements [1]. The empirical research program was used in deducting damage in the plate structures using the simplified energy principles for locating the damage location at low, medium, and high frequency levels [2]. The free vibration of the circular cylindrical shell structures with the crack presence was studied experimentally. The sensitivity analysis was used to deduct the damage in addition to investigate the damage estimation in glass fiber-reinforced polymer composite panels [3]. The vibration of damaged structures was investigated deeply using the experimental programs [4–8]. were The damage deduction study in structures was achieved using the remaining error method, especially in the bar and plate elements. The performance of the method is compared to the performance of ADMSC (absolute difference between the damage mode shape curvature) in two dimensions to assess the strength and restrictions of the method in the case of the three structures [9]. The damage identification technique and characterization procedure for beam structures by introducing a damage index was achieved based on the geometry of the beam's first mode [10]. The sensitivity of the inverse algorithms to detect damage in the constructs was employed to describe a two-step approach to accurately predicting discharge parameters that requires sensitivity analysis using measurements of noise levels, location, and size of discharge to evaluate prediction errors [11]. The experimental study on damage identification in composite plates reinforced with hemp fiber (Cannabis sativa) was presented using modal analysis techniques [12]. The acoustic-laser technique for defect detection in FRP-bonded structural systems in the presence of induced airborne noise was studied to establishes a quantitative correlation [13]. The dynamic analysis was presented in many related literatures [14–19]. The method for the crack detection in plate-like structures using a typical strain energy and the detection force index was employed to evaluate the accuracy of crack location and length detection. The method was tested on a rectangular aluminum plate with three different boundary conditions [20]. The study of the defect detectability of time- and frequency-domain analysis techniques was conducted for flash thermography using three CFRP samples with different defect types, sizes, and depths [21].

In the present work, sensor vibration was used to detect the effect of crack length in carbon/Kevlar hybrid fiber-reinforced polymer composites by measuring the change in natural frequencies, and a numerical analysis was performed via ANSYS software to examine different crack scenarios.

2 Mathematical Model

The natural frequencies of the plate shown in Fig. 1 are obtained using the following frequency equation [22]:

$$\text{Natural frequency(Hz)}\, f_{ij} = \frac{\lambda_{\mathrm{ij}}^2}{2\pi a^2}\left[\frac{Eh^3}{12\gamma\left(1-\nu^2\right)}\right]^{\frac{1}{2}} \quad \mathrm{i} = 1, 2, 3\ldots;\ \mathrm{j} = 1, 2, 3\ldots$$

The chosen case study is the composite mate material Mat type and it is considered as an isotropic properties with the following geometry and material properties:

a = 200 mm Length of plate.
b = 200 mm Width of pate.
h = 2 mm Thickness of plate.
v = 0.3 Poisons ratio of composite material.
ρ = 1500(Kg/m3) Density of composite material.
E = 10 (GPa) Young's modulus of elasticity.

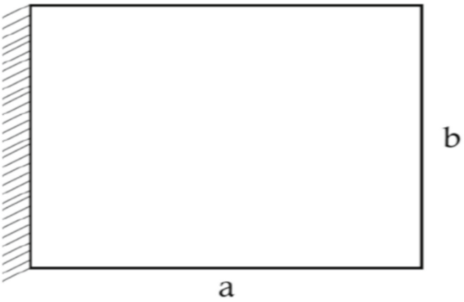


Fig. 1. A cantilever plate

The natural frequencies of various conditions can be determined mathematically using this general formula. Both the software analysis and the experimental modal analysis can be used to compare all of the natural frequencies. Theoretical natural frequencies are calculated from plate dimensions and material parameters. Table 1 shows the theoretical natural frequencies of a rectangular cross-sectioned cantilever plate. When a/b = 1.

Table 1. Theoretical Natural Frequencies

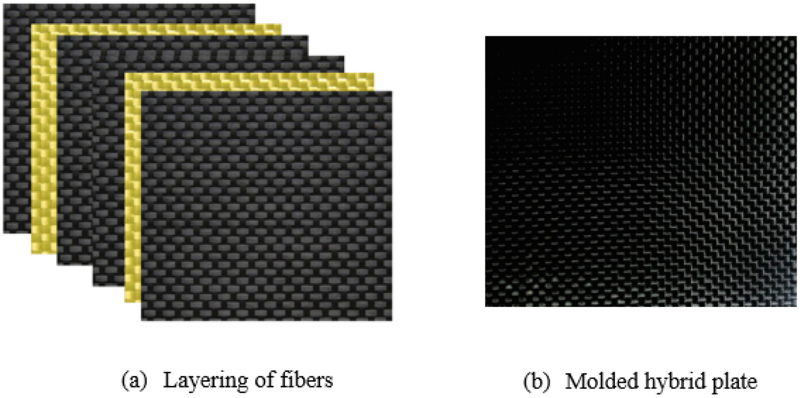
Mode	Natural frequency (HZ)
1	21.72339239
2	53.03319591
3	133.3139458
4	170.017272
5	193.5322844
6	338.6659455

3 Experimental Work

The aim of this experimental work is to manufacture a sample of a composite material and analyze its natural frequencies in both its pure and defective versions. The comparison of the results allows us to assess the impact of the flaw on the composite structure's dynamic properties [23–26].

3.1 Manufacture of Samples

A hybrid sample consisting of carbon fiber fabric and Kevlar fiber fabric coated with epoxy resin is manufactured. The bag-forming process is used to manufacture the composite plate. A layer of carbon fiber is placed, then a layer of Kevlar fiber, then two layers of carbon fiber mat, then a layer of Kevlar fiber mat and a layer of carbon fiber are placed, respectively, as shown in the Fig. 2.a, then the epoxy resin matrix material was extruded, after that, by drawing air through the vacuum pressure process. The mold from the blanks is emptied completely. The casting process takes place at room temperature, then a load was applied and heat the mold, about 60 °C, for a period of three hours, then the mold was taken out of the mold and put it in the oven at a temperature of 100 °C for three hours. The mold is then removed from the oven and is ready, as illustrated in Fig. 2.b.


(a) Layering of fibers

(b) Molded hybrid plate

Fig. 2. Composite plate manufacturing

3.2 Specimens Tests

In order to analyze the mechanical properties and behavior of the material, it is necessary to conduct an experiment on a sample of the material in a laboratory. The tensile and bending test is one of these tests that is considered to be among the most significant. These tests are often carried out to assess the strength, elasticity, and ductility of a material, as well as any other properties it may possess, under a wide variety of loads and conditions.

A. **Test tensile**

The ASTM D 638–03 standard test technique for assessing the tensile characteristics of plastics was utilized for the tensile testing that was carried out. The sample that was used in the test was cut from a composite sheet material that was previously manufactured, and the dimensions of the sample that were utilized in the test are shown in Fig. 3 as follows:

Length: 165 mm.
Maximum width: 19 mm.
Minimum width: 13 mm.
Thickness: 2 mm.

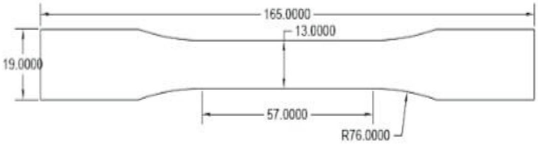
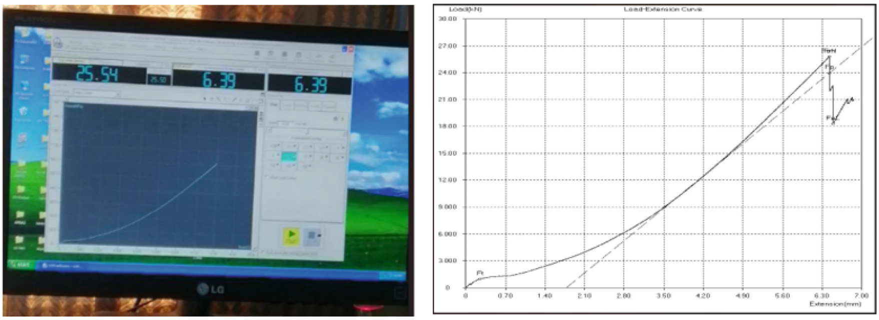


Fig. 3. Tensile Test Specimen

These measurements are in reference to the physical characteristics of the specimen that have been put through a tensile examination. The experiment was carried out in the Materials Engineering Department at the University of Technology. Figure 4.a.b depicts the procedure that must be followed in order to calculate the modulus of elasticity, which can be found in Table 2. The specimen must be loaded into the tensile testing equipment in a longitudinal orientation before it can be pulled hydraulically with a big steel strip at a velocity of 5 m/s.



(a) Tensile data reading

(b) Flexibility curve

Fig. 4. The flexibility curve

Table 2. Mechanical properties of chopped composite plate.

Properties	Value
Elastic modules (E) (GPa)	10
Shear modulus (G) (GPa)	3.85
Poisson ratio	0.3

B. **Flexural test**

The standard dimensions of the bending test sample plate for composite materials can vary depending on the specific test. The sample that was used in the test was cut from a composite sheet material that we had previously manufactured, and the dimensions of the sample that were used in the test are shown in Fig. 5 as follows:

Thickness: 2 mm.

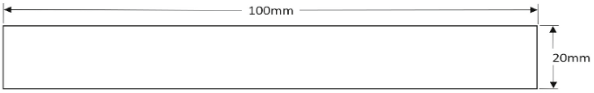


Fig. 5. Standard flexural test sample

The experiment was carried out in the Department of Materials Engineering at the University of Technology. Table 3 presents the mechanical properties of the hybrid materials. The sample is placed in the testing machine, where the type of test is triple, consisting of two supports and a force applied from the middle applied to the sample, as shown in Fig. 6.

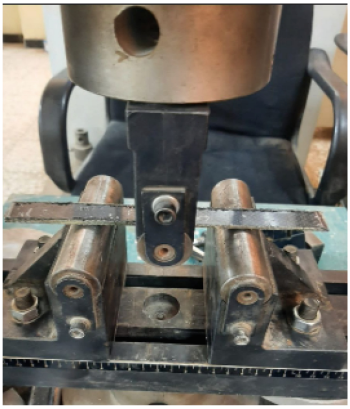


Fig. 6. Flexural test machine

Table 3. Mechanical properties of chopped composite plate.

Properties	Value
Elastic modules Bending (Eb) (GPa)	24
Shear modulus (G) (GPa)	9.23
Poisson ratio	0.3

C. **Free Vibration Test**

The dimensions of a composite cantilever slab that was subjected to a vibration test are as follows: Length of 200 mm, width of 200 mm, the thickness is 2 mm. These instruments are used for this examination Fig. 7:

1. Impact Hummer.
2. Accelerometer.
3. Charge Amplifier.
4. Digital Oscilloscope.
5. Flash Memory

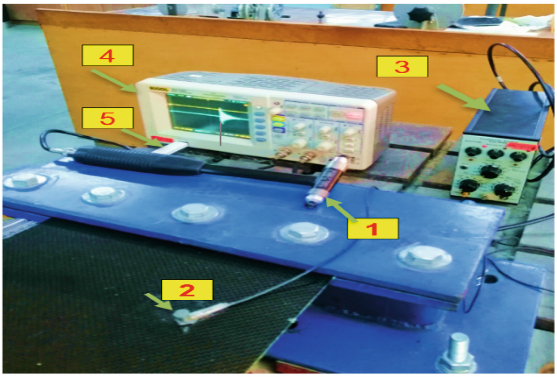


Fig. 7. The cantilever plate test rig.

To determine the natural frequency and response of the undamaged and damaged cantilever plates in order to detect the damage and characterize the material based on the difference in frequency values. The variety of impact hammer (patent NO 4799375) is used to excite the plate with a pulse signal, which causes the plate to vibrate. In addition, an accelerometer of type (4368) is installed on the circuit board in order to acquire the signal for an oscilloscope of type (DS1102E). The oscilloscope displays the response of the sample supplied by the oscilloscope-connected impact hammer that produces the load applied to the plate. The oscilloscope displayed the signal as a sine wave, whose peaks indicated the signal's inherent frequency. As shown in Fig. 8 depicts the experimentally determined response spectrum of a cantilever deflection plate made of isotropic material for plates with a deflection angle between 0 and 90 °. It was deduced from the Fourier transform of the temporal history of the transient response of the excited plate. Figure 8 shows the results of the response with time picked up the accelerometer signals which are converted into frequency with time using the SIGVIEW program.

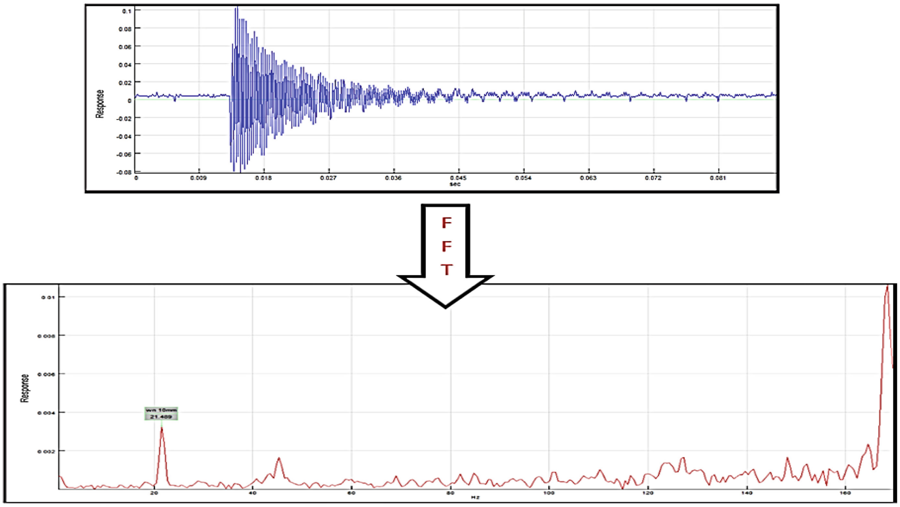


Fig. 8. The FFT analyses of the response wave (SIGVIEW program)

4 Numerical Analysis

The study conducted a numerical analysis utilizing the ANSYS commercial FE code to develop a finite element model of both the undamaged and damaged composite plates. The purpose of this analysis was to investigate the free vibration characteristics of the plates. The laminated composite board is composed of carbon fiber and Kevlar as the reinforcing materials, while the matrix is made of epoxy resin. Tables 2 and 3 enumerate the material properties that were computed and utilized for finite element analysis. The dimensions of the composite plate are as follows: length (L) of 200 mm, width (W) of 200 mm, and depth (h) of 2 mm. The current problem is being addressed through the utilization of shell element 281 in the implementation of FEA [27]. The structure is comprised of eight distinct nodes, each possessing six degrees of freedom. The analysis employs six degrees of freedom, namely translations and rotations around the X, Y, and Z axes. The determination of the quantity of elements in the finite element models was accomplished through the utilization of a convergence test (as depicted in the figure). The outcome of this test yielded a mesh size of 108 x 108, which is presented in Fig. 9. The convergence study shown in this figure gives the suitable mesh used in the finite element discretization according to the number of finite element used in the numerical modelling. The Table 4 shows the results obtained for the model numerically.

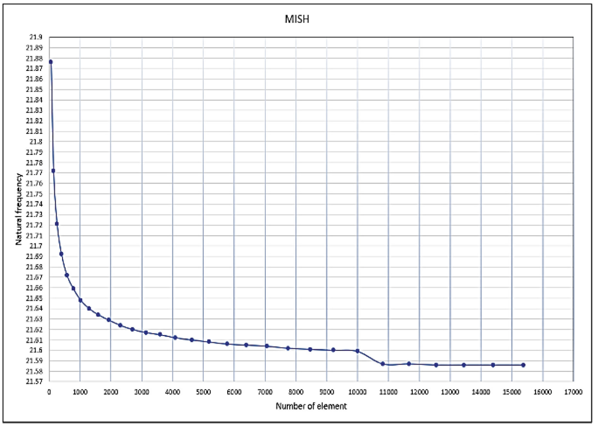


Fig. 9. Convergence of natural frequencies

Table 4. Numerical Natural Frequencies (Hz)

Mode	Natural frequency (HZ)
1	21.587
2	52.746
3	132.27
4	168.79
5	191.88
6	335.4

The results of finite element method shown in Table 4 are greed very well with those obtained from the theoretical findings presented in Table 1. The natural frequencies were calculated up to the sixth mode. In fundamental mode the percentage of discrepancy is not more than 0.5% and in the sixth mode the discrepancy is less than 4%. This gives a confidence in dealing with the finding the natural frequency with and without defects.

5 Results and Discussions

The experimental and numerical analysis of free vibration was undertaken on a cantilever deflection plate made of isotropic composite materials The board has a thickness (2 mm) and is constructed out of carbon fibers, Kevlar, and epoxy resin. It has cracks of varying lengths (1, 2,3…, 8) centimeters and varying numbers (1, 2, 3). These fractures may be found in a variety of places on the plate, such as on the side edge, in the middle, or at the end of the plate. The next section will describe the influence that the aforementioned parameters have on the free vibration qualities. Figure 11 shows the numerical and experimental results for the fundamental frequency of cantilever deflection plates for slit length. These results demonstrate that the natural frequency decreases with increasing slit length due to the fact that an increase in slit length decreases the hardness of the material, which in turn leads to a decrease in the natural frequency. And the amount of increase in the crack length is the increase in the difference ratio at a rate of (0.48) for the same site, at one centimeter. The increase in the number of cracks leads to a decrease in the natural frequency values, as shown in Fig. 12, and this is due to the same previous reason. As for the difference in the location of the crack in the two examples that came before, as shown in Fig. 12, the natural frequency increases as we travel away from the fixation axis. According to what is shown in Fig. 13, the natural frequency goes up as it gets closer to the center orthogonal axis of the fixation. As can be seen in Fig. 14, the difference between the frequency values is almost negligible when the length of the slit is aligned perpendicular to the axis of the fixation. The greatest difference percentage of (7.09%) between numerical and experimental results indicates that there is a discrepancy between the predicted values obtained through numerical analysis and the actual measurements obtained through experimental testing as shown in Fig. 15. This disparity is attributable to a number of factors, including manufacturing and implementation conditions, as well as the precision of the measuring devices (Fig. 10).

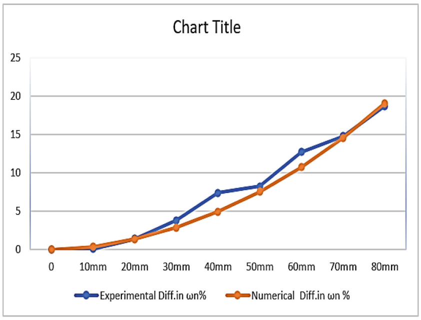


Fig. 10. The percentage increase in the change in frequency according to the length of the crack.

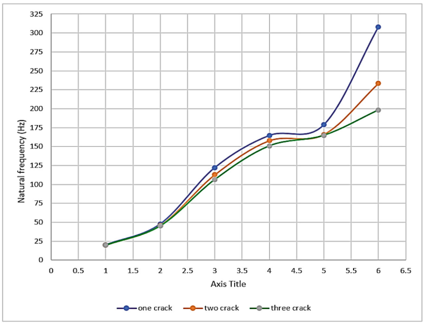


Fig. 11. The change in the frequency value in the presence of one, two, or three cracks.

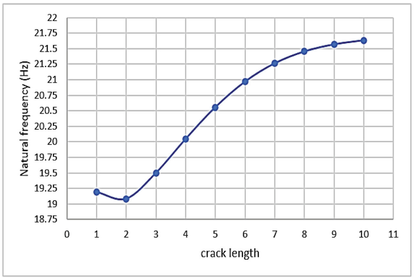


Fig. 12. The change in the frequency value according to the distance between the crack and the fixation axis.

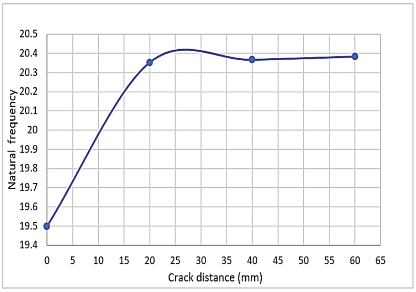


Fig. 13. The change in natural frequency as the crack approaches the mid-axis

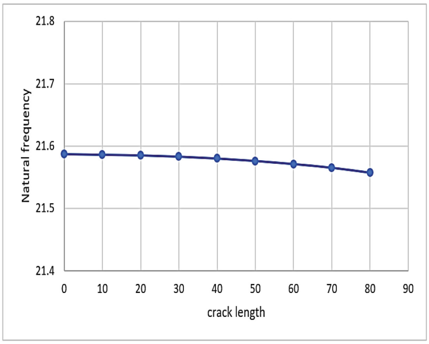


Fig. 14. The frequency value at the crack perpendicular to the fixation axis

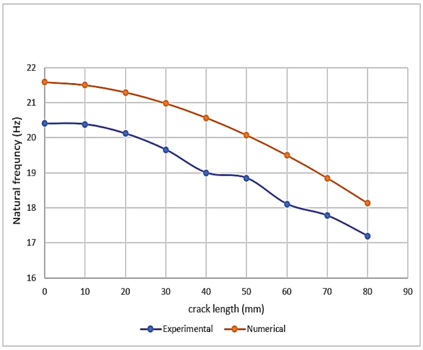


Fig. 15. Experimental and Numerical Results.

6 Conclusions

The primary gials of this study are to determine the effects of changes in natural frequencies for detecting the damage in composite panel structures and monitoring their healths. From the present work's numerical and experimental investigations, the following conclusions drawn:

1. The diffrernce between the natural frequencies of the damaged and undamaged plates depends upon the direction of the crack. When the longitudinal axis of the crack is parallel to the fixing direction, the change in natural frequencies is significant, opposite to that when the axis of the crack is perpendicular to fixation.
2. The change in natural frequencies is incresed between the damaged and undamaged plates in case if increasing the crack length, this is due the reduction in stiffness due to the increase in the crack length with the same mass.
3. The change in natural frequencies is increased as the crack position approahes the clamping axis. This is an indication of the crack position in the case of health monitoring of the strucures in service. The site engineer will be able to estimate approximately the crack position.
4. The size of the difference in the value of the natural frequency increases as the mode is increased. The lowest value of the frequency changes appears in the first mode. The change in natural frequencies between the damaged and undamaged natural frequencies becomes significant in higher vibration modes between the damaged and virgin plate strucure.
5. The maximum discrepancy between the experimental and numerical results was 7.09%.

References

1. Gomes, F.G., et al.: A numerical–experimental study for structural damage detection in CFRP plates using remote vibration measurements. J. Civ. Struct. Health Monit. **8**, 33–47 (2018)
2. Samet, A., et al.: Experimental investigation of damage detection in plate-like structure using combined energetic approaches. Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci. **233**(4), 1193–1203 (2019)
3. Husain, M.A., Al-shammari, M.A.: Effect of cracks on the natural frequency of cylindrical shell structures. Eng. Technol. J. **38**(12) Part A, 1808–1817 (2020)
4. Jweeg, M.J., Hussein, E.Q., Mohammed, K.I.: Effects of cracks on the frequency response of a simply supported pipe conveying fluid. Int. J. Mech. Mechatron. Eng. **17**(5) (2017)
5. Al-Shammari, M.A., Al-Waily, M.: Theoretical and numerical vibration investigation study of orthotropic hyper composite plate structure. Int. J. Mech. Mechatron. Eng. **14**(6) (2014)
6. Al-Shammari, M.A., Husain, M.A., Al-Waily, M.: Free Vibration analysis of rectangular plates with cracked holes. In: 3rd International Scientific Conference of Alkafeel University, A.I.P. Conference Proceedings, p. 2386 (2022)
7. Abbas, E.N., Jweeg, M.J., Al-Waily, M.: Analytical and numerical investigations for dynamic response of composite plates under various dynamic loading with the influence of carbon multi-wall tube Nano materials. Int. J. Mech. Mechatron. Eng. **18**(6), 1–10 (2018)
8. Al-Baghdadi, M., Jweeg, M.J., Al-Waily, M.: Analytical and numerical investigations of mechanical vibration in the vertical direction of a human body in a driving vehicle using biomechanical vibration model. Pertanika J. Sci. Technol. **29**(4), 2791–2810 (2021)
9. Katunin, A., Ručevskis, S.: Effectiveness of damage identification in composite plates using damage indices based on smoothing polynomials and curvelet transform: a comparative study. Procedia Struct. Integr. **37**, 292–298 (2022)
10. Gorgin, R.: Damage identification technique based on mode shape analysis of beam structures. In: Structures, vol. 27. Elsevier (2020)
11. Iqbal, A., et al.: Damage detection in one-and two-dimensional structures using residual error method. J. Sound Vib. **462**, 114908 (2019)
12. Zhang, Z., et al.: Sensitivity analysis of inverse algorithms for damage detection in composites. Compos. Struct. **176**, 844–859 (2017)
13. Patil, S., Reddy, D.M.: Damage identification in hemp fiber (cannabis sativa) reinforced composite plates using MAC and COMAC correlation methods: experimental study. J. Nat. Fibers **19**(4), 1249–1264 (2022)
14. Njim, E.K., Bakhy, S.H., Al-Waily, M.: Free vibration analysis of imperfect functionally graded sandwich plates: analytical and experimental investigation. Arch. Mater. Sci. Eng. **111**(2), 49–65 (2021)
15. Jweeg, M.J., Hammood, A.S., Al-Waily, M.A.: Suggested analytical solution of isotropic composite plate with crack effect. Int. J. Mech. Mechatron. Eng. **12**(5) (2012)
16. Al-Waily, M., Al-Shammari, M.A., Jweeg, M.J.: An analytical investigation of thermal buckling behavior of composite plates reinforced by carbon Nano particles. Eng. J. **24**(3) (2020)
17. Njim, E.K., Bakhy, S.H., Al-Waily, M.: Analytical and numerical investigation of buckling load of functionally graded materials with porous metal of sandwich plate. Materials Today: Proceedings (2021)
18. Jweeg, M.J., Hammood, A.S., Al-Waily, M.: Experimental and theoretical studies of mechanical properties for reinforcement fiber types of composite materials. Int. J. Mech. Mechatron. Eng. **12**(4) (2012)
19. Jweeg, M.J., Al-Waily, M., Deli, A.A.: Theoretical and numerical investigation of buckling of orthotropic hyper composite plates. Int. J. Mech. Mechatron. Eng. **15**(4) (2015)

20. Qiu, Q., Lau, D.: Defect detection of FRP-bonded civil structures under vehicle-induced airborne noise. Mech. Syst. Signal Process. **146**, 106992 (2021)
21. Le, T.C., et al.: Crack detection in plate-like structures using modal strain energy method considering various boundary conditions. Shock and Vib. 1–17 (2021)
22. Poelman, G., et al.: An experimental study on the defect detectability of time-and frequency-domain analyses for flash thermography. Appl. Sci. **10**(22), 8051 (2020)
23. Abbas, H.J., Jweeg, M.J., Al-Waily, M., Diwan, A.A.: Experimental testing and theoretical prediction of fiber optical cable for fault detection and identification. J. Eng. Appl. Sci. **14**(2), 430–438 (2019)
24. Al-Waily, M., Jweeg, M.J., Al-Shammari, M.A., Resan, K.K., Takhakh, A.M.: Improvement of buckling behavior of composite plates reinforced with hybrids nanomaterials additives. Mater. Sci. Forum **1039**, 23–41 (2021)
25. Kadhim, A.A., Al-Waily, M., Abud Ali, Z.A.A., Jweeg, M.J., Resan, K.K.: Improvement fatigue life and strength of isotropic hyper composite materials by reinforcement with different powder materials. Int. J. Mech. Mechatron. Eng. **18**(2) (2018)
26. Abbas, E.N., Al-Waily, M., Hammza, T.M., Jweeg, M.J.: An investigation to the effects of impact strength on laminated notched composites used in prosthetic sockets manufacturing. Mater. Sci. Eng. **928** (2020)
27. Abud Ali, Z.A.A., Kadhim, A.A., Al-Khayat, R.H., Al-Waily, M.: Review influence of loads upon delamination buckling in composite structures. J. Mech. Eng. Res. Dev. **44**(3), 392–406 (2021)

Computer Modelling of the Gait Cycle Patterns for a Drop Foot Patient for the Composite a Polypropylene Ankle-Foot Orthoses

Maryam I. Abduljaleel[1], Muhsin J. Jweeg[2](✉), and Ahmed K. Hassan[3]

[1] Department of Mechanical Engineering, Faculty of Engineering, University of Kerbala, Karbala, Iraq
m09152086@s.uokerbala.edu.iq

[2] College of Technical Engineering, Al-Farahidi University, Baghdad, Iraq
muhsin.jweeg@uoalfarahidi.edu.iq

[3] Prosthetics and Orthotics Engineering Department, College of Engineering, University of Kerbala, Karbala, Iraq
dr.ahmed_kh74@uokerbala.edu.iq

Abstract. Ankle-foot orthoses (AFO) is a device that supports the ankle and foot part of the body when there is a muscle weakness or a nerve damage, Ankle-foot orthoses are prescribed to individuals with minimal spinal cord injury and excellent trunk muscle control (AFOs). In this work design and manufacturing of Ankle Foot Orthosis (AFO) was achieved experimentally to obtain the gait cycle shape using the suggested types of composite layering. The experimental program includes the manufacture two type of AFO: Carbon, Perlon, Kevlar, Kenaf fibers. Two Sequences are used in this works based on the results of the examinations of the samples. The sequence 1 is the order of the layers in it (2Perlon + 1Carbon fiber + 2Perlon + 1Kevlar + 2Perlon) and the Sequence 2 is (2Perlon + 3Carbon fiber + 2Perlon + 3Kevlar + 2Perlon). The patient's gait cycle data (including pressure distribution using an F-socket and Ground Reaction Force [GRF] using a force plate) were presented here. The patient's height (176 cm), weight (78 kg), and approximate age of 39. The case study was a patient suffered from drop foot. The patient's gait cycle data (including pressure distribution using F-socket and using a force plate, ground reaction force) have been collected. All the readings of the gait cycle became different when the patient was without wearing AFO because of the foot drop due to severe damage to the nerves, so there was a difference in the readings of the right and left foot, while the difference decreased in the case of the patient wearing AFO because the AFO helped the patient to walk, and therefore the readings converged with the readings of the normal person. Finally, the findings of patient balance and gait cycle show that composite materials AFO superior than PP.

Keywords: AFO · Polypropylene · Composite material · GRF · Drop foot · Tensile test · Fatigue test · Gait cycle

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 212–225, 2024.
https://doi.org/10.1007/978-3-031-56728-5_19

1 Introduction

The human foot plays a very important role in human movements such as standing, running, walking, jumping, etc. The human foot in general made up of three sections: the forefoot, the middle of the foot, and the hind foot. AFOs are used to help people walk by clearing the path for their feet as they progress through the various stages of walking (called "gait stages"). They can also be used as rehabilitation devices to correct motor patterns, such as limiting plantar flexion or dorsiflexion, decreasing the risk of falling, increasing balance, or strengthening the muscles in the lower legs [1–3]. Patients with drop foot commonly have a complete or partial weakness in muscles that dorsiflexion the foot at the ankle joint, so that the main aim of AFO is to supply patients with a convenient orthosis that will grant them the most normal gait possible [4–7].

The acceleration was measured at the trunk to record the gait cycle characteristics. Signals from each trial were transformed into a horizontal-vertical coordinate system and analyzed using an unbiased autocorrelation method to derive metrics such as step length, cadence, and measures of gait regularity and symmetry. Data acquired from a timing device and a portable sensor at free walking speeds demonstrated that gait cycle characteristics analysis, which previously required fixed laboratory equipment and timed walking methods, is now possible [8–12].

The impact of using on-body accelerometers was used to analyze gait. A novel method was proposed whereby wearable sensor units are used to analyze human walking posture. The sensor modules consist of three gyro sensors that are oriented along three axes and a tri-axial acceleration sensor. The angular velocity and acceleration during walking were measured using seven sensor units that were worn on the abdomen and lower limb segments (shanks, thighs, and feet). In this way, the angles and lengths of the segments at each joint can be used to calculate their coordinates in three spaces. Mechanically, joint angle can be determined by observing the gravity acceleration along "the segment's anterior axis. Thus, it was proposed to separate the gravitational acceleration from the acceleration data using an optimization approach. Since the cyclic pattern of acceleration data may be uncovered through continuous walking, an FFT analysis was utilized to extract particular differentiating frequencies from it [13–17].

The force analysis is essential for amputation patients undergoing rehabilitation because overloading might endanger the bone-implant contact and under loading could lengthen their already lengthy recovery times.

Force analysis is made simple and accurate by modern sensor and data collecting technologies together with the background information on orthotics and prosthetics, including Osseo integration, is provided. An integration platform and appropriate sensor were chosen to construct a new gait analysis system based on the criteria that are provided. The prototype is suggested for use in a number of experiments with various gait states for gait analysis, and the outcomes are then presented and discussed. Due to the prototype's success, plans have been made to create an OPRA ("Osseo integrated Prostheses for the Rehabilitation of Amputees") product [18–22].

The Immediate Effects and Gait Deviations of Ankle-Foot Orthotics and "Functional Electrical Stimulation" was presented to analyse the connection between weakened sensorimotor control and dropped foot gait abnormalities and analyse the immediate effects of a “functional electrical stimulation” (FES) device and an ankle-foot orthotic (AFO) on stroke survivors with dropped foot disability [23–25].

In this work, a patient suffered from a drop foot was chosen to wear the suggested types of laminations in order to test the effectiveness of the AFO manufactured here from the durability side and the gait cycle pattern in addition to compare the biomechanics of walking in stroke with and without impaired dropped foot. This article aimed to present a novel technique for collecting gait cycle measurements with a minimal data collection effort.

2 Theoretical Consideration

Drop foot often shows several gait problems, including slower walking, longer steps, and a lower vertical peak force. The absence of active dorsiflexion movement in the ankle joint is thought to be the cause of the abnormal gait patterns. The anatomical outline of the lower leg will be shown in this chapter, highlighting the key components that support ambulation. The gait cycle, kinetic and kinematic analyses, and “ground reaction force” are all investigated to determine the abnormal gait characteristics. The gait cycle determination is so useful to the clinician to provide him with the information about the ability of the patient to walk safely using the AFO. The gait cycle divisions are the initial heel contact to the final toe-off and swing form toe-off to heel contact as shown in Fig. 1. The ground reaction forces during the gait cycle using the force are measured using the force plate. It is an instrument to read the force and moment using the sensors attached to the required positions. The signals from the sensors during the gait cycle were measured and interfaced with a computer to facilitate the ground reaction forces values. Having obtained these forces, the stress analysis may be achieved to ensure the safety requirements.

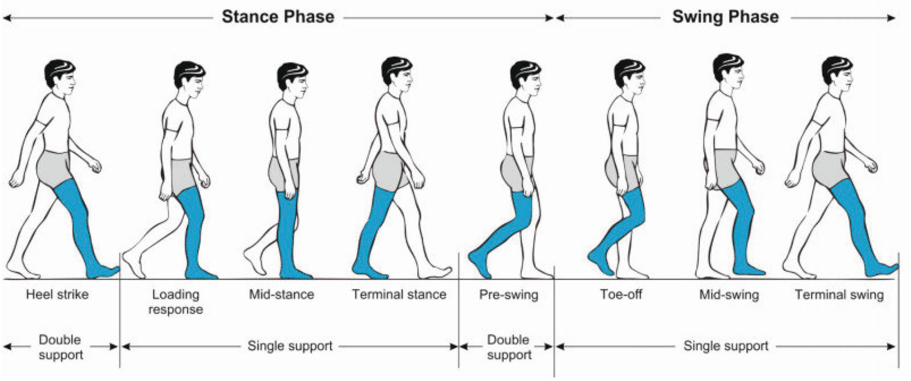


Fig. 1. Gait Cycle

2.1 The Ground Reaction Force

GRF, or ground reaction forces, are created throughout gait as a response of the force the foot exerts on the ground as it comes into contact with it. The force exerted by the foot on the ground is equal to and opposite to GRF. The GRF is of major relevance in gait analysis since it is an external force that acts on the body as it is moving. The measurements obtained from the of the GRF are valuable to assess the applicability of using the AFO and gives an indication of the developed forces and their directions during the gait cycle directly from the computer and stored which may be used later in the gait cycle analysis. The M curve, which resembles the shape of the letter M, is a common representation of the vertical ground response force over one gait cycle. (Fig. 2) Fz decreases to roughly 80% body weight during single stance and reaches a maximum of 120% body weight during the double stance phase [26].

According to the second Newton's law of motion was.

$$\sum \boldsymbol{F} = \boldsymbol{Ma} \tag{1}$$

$$\sum \boldsymbol{F} = \sum \boldsymbol{GRF} - \boldsymbol{Mg} = \boldsymbol{Ma} \tag{2}$$

$$\mathrm{GRF} = \mathrm{M(g + a)} \tag{3}$$

where M: mass of the person, GRF: ground reaction force, and (g, a): gravity and the centre of mass vertical accelerations respectively

According to the equation above, the value of GRF is dependent on vertical acceleration since M and g have constant values. Therefore, depending on the body weight and centre of gravity, the net force may be positive, negative, or zero [27].

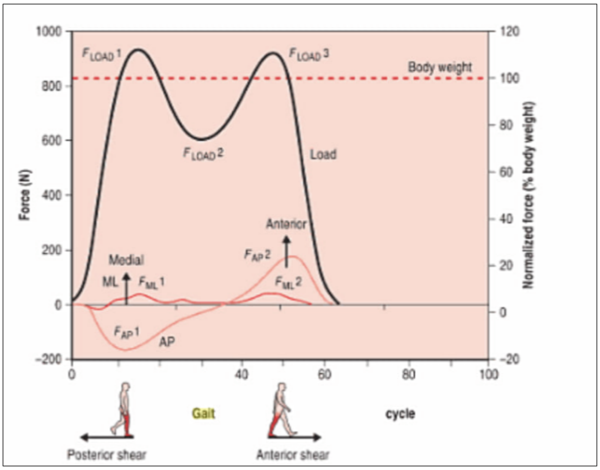


Fig. 2. The three elements of GRF during a typical gait cycle. Here, Fz, the vertical part of the GRF, is referred to as FLOAD. FML is the medial force component of GRF, while FAP is its anterior/posterior force component [4]

3 Experimental Work

This includes the apparatus and materials utilized in this paper in addition the experimental process, the materials used in AFO are carbon, Kevlar, kenaf fibers, and perlon, Polyvinylalcohol (PVA), Polypropylene and hardener.

3.1 Manufacturing of Ankle Foot Orthosis (AFO)

a. The cast was connected to the vacuum machine forming system through the pressure tubes, then dressed the mold a sock for give a smooth surface and easy to separate the orthosis from mold then pull the PVA bag to the positive mold and open the pressure valve to make the PVA stuck on the cast, after that drag the PVA bag to the positive mold and release the pressure to allow the PVA to stick to the cast. As shown in Fig. 3.a.

b. Putting the perlon stockinet (2 perlon) layers Fig. 3.b, put 1 layers of carbon fiber Fig. 3.c, putting (2 perlon) layers, put 1 layers of Kevlar Fig. 3.d, and the end layer is (2 perlon) layers.

c. After finishing arranging the layers by pulling the outer (PVA) bag while maintaining the smaller end over the value region, and tying off the (PVA) bag using cotton string. As shown in Fig. 3.e.

d. The C-orthocryl lamination resin is combined with the hardener in a ratio of around 800 g of resin to 6.84 g of hardener. The resultant matrix combination is then placed

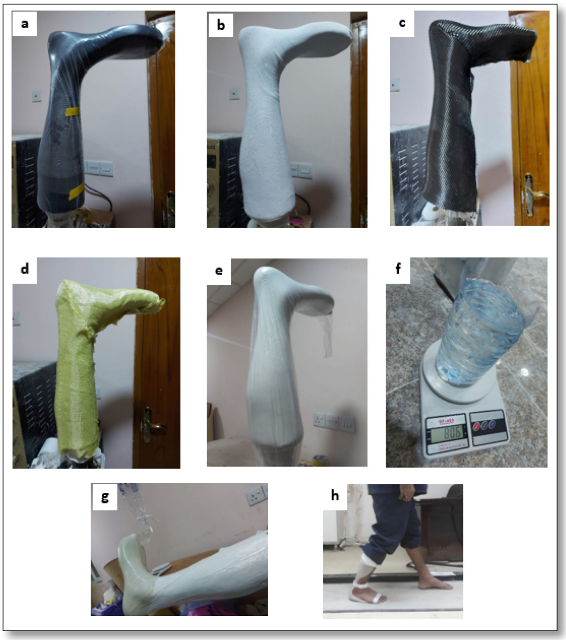


Fig. 3. Preparing the Negative Gypsum Pattern

inside the outer (PVA) bag, where it is evenly distributed across the whole lamination surface. As shown in Fig. 3.f.

e. Maintaining a steady vacuum on until composite materials become cool it was cooled after about 3 h, leaving the resultant lamination, and then cutting it to the requisite dimensions for an ankle-foot orthosis. As shown in Fig. 3.g.
f. Figure 3.h shows the image of the patient after wearing the AFO

4 Results and Discussion

The gait cycle's findings (pressure distribution, GRF, peak contact pressure, force distribution, center of pressure (COP), peak contact area, gait analysis, gait cycle tables, step-stride, and foot print analysis) and the F-socket device to measure the pressure at the patient's leg's interface with the AFO using a sensor.

The main difference between the diseased subject (drop foot patient) will be displayed, when he wearing an ankle foot orthosis (AFO) made of composite material and plastic (Polypropylene), and not wearing an AFO. The examination of the kinematics data will show the changes in gait cycle behavior between the diseased and normal individual. Six different Sequences were tested and only two Sequences (Sequences1 and Sequences 2) were used in paper based on the results of the examinations of the samples.

5 Tensile Properties Results

The mechanical parameters (Ϭy, Ϭultand E) of each sample are obtained from these curves and listed in Table 1. The results show that increasing the layers of carbon and kevlar fibers while keeping a constant number of layers of perlon affected mechanical characteristics and increased Ϭyand Ϭultas shown in Figs. 4 and 5 for the sequences 1 and 2 respectively.

Table 1. The sequence's mechanical characteristics

Serial	Number of layers	Ϭy (MPa)	Ϭult (MPa)	E (GPa)
Sequences1	8	50	67	2.22
Sequences2	12	63	80	3.7

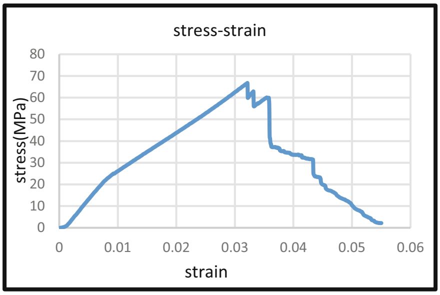


Fig. 4. Stress-strain curve for a sample of the sequence 1

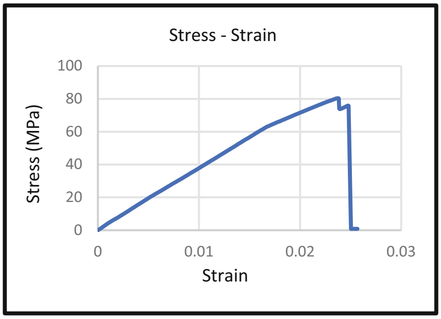


Fig. 5. Stress-strain curve for a sample of the sequence 2

6 Fatigue Characteristics Results

Figures 6 and 7 show the fatigue test results of 21 samples for sequence 1 and sequence 2. The number of cycles required to achieve the failure sites is increasing while the failure stresses are decreasing. Where we note that the highest number of cycles obtained in the first sequence is 1358489, in the event that it was in the second sequence 5587528. This indicates that the second sequence achieves a longer lifetime for the AFO.

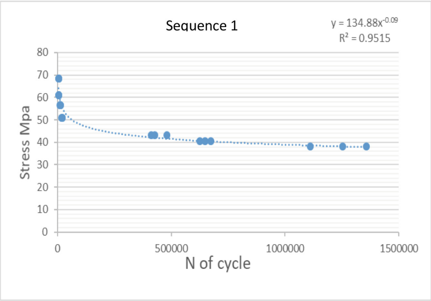


Fig. 6. S-N curves for sequence 1

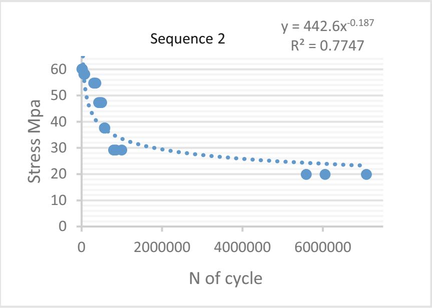


Fig. 7. S-N curves for sequence 2

7 The Gait Cycle and Step Table Parameters' Results and Discussion

Table 2 show the gait cycle parameters for the patient with and without AFO.

The results shown in Table 2 indicate that first, when the patient is walking without an orthosis; second, when the patient is wearing an orthosis while walking. The outcomes show that, when the AFO orthosis is not worn, there is a difference in the gait cycle data between the foot with the abnormality and the normal foot. As a result, take notice that the orthosis-wearing patient's gate cycle data shows very little difference between the abnormal foot and the normal foot.

Table 2. The Gait Cycle Parameters' Results

Gait cycle table(sec)	Patient without AFO			Patient with PP AFO			Patient with Composite material AFO		
	Left leg	Right leg	Difference	Left leg	Right leg	Difference	Left leg	Right leg	Difference
Gait cycle Time	1.42	2.36	0.94	1.63	1.85	0.22	1.71	1.9	0.19
Stance Time(sec)	0.77	1.29	0.52	0.87	0.92	0.05	0.79	0.85	0.06
Swing Time(sec)	0.65	1.07	0.42	0.76	0.93	0.17	0.92	1.05	0.13
Single support Time	0.8	1.85	1.05	0.97	1.16	0.19	0.86	1.11	0.25
Initial Double support time	0.28	0.87	0.59	0.34	0.32	0.02	0.38	0.33	0.05
Terminal Double support time	0.87	0.28	0.59	0.32	0.34	0.02	0.33	0.38	0.05
Total Double support time	1.15	1.15	0	0.75	0.75	0	0.71	0.71	0
Heel Contact Time	0.42	0.83	0.41	0.67	0.72	0.05	0.61	0.69	0.08
Foot Flat Time	0.23	0.67	0.44	0.21	0.35	0.14	0.24	0.32	0.12
Mid Stance Time	0.47	0.75	0.28	0.57	0.66	0.09	0.53	0.64	0.11
Propulsion Time	0.55	0.82	0.27	0.52	0.63	0.11	0.49	0.61	0.12
Active Propulsion	0.38	0.43	0.05	0.31	0.33	0.02	0.34	0.37	0.03
Passive Propulsion	0.17	0.39	0.22	0.21	0.28	0.07	0.15	0.24	0.09
Step table	Patient without AFO			Patient with PP AFO			Patient with Composite material AFO		
	Left leg	Right leg	Difference	Left leg	Right leg	Difference	Left leg	Right leg	Difference
Step time (sec)	0.78	1.4	0.62	0.82	0.98	0.16	0.80	0.95	0.15
Step length (cm)	31.4	42.1	10.7	29.2	33.4	4.2	28.8	30.4	2.6
Step velocity(cm\sec)	40.2	30	10.2	35.6	34.08	1.51	34.7	32	2.7
Step width (cm)	12.6	13.1	0.5	11.7	11.4	0.3	10.8	11.3	0.5

There are several causes for the variations in phases, stride length, and stride time. The main factor is wearing an ankle foot orthosis in the right leg, which has drop foot and calf muscles that are atrophying. This supports the legs and improves the condition. The step width of right leg is reduced from 13.1 cm to 11.3 cm by using composite materials AFO. The step length of right leg is reduced from 42.1 cm to 30.4 cm by using composite materials AFO. In addition, the step time was also reduced from 1.4 to 0.95 s with composite AFO.

Note that the difference in the heel contact time, foot flat time and mid stance time is reduced when the patient wears the AFO from 0.41, 0.44 and 0.28 respectively to 0.05, 0.14 and 0.09 in PP AFO and 0.08, 0.12 and 0.11 in Composite material AFO.

In the case of a normal person, the readings of the right and left feet are very close. All the readings became different when the patient was without wearing AFO because of the footdrop due to severe damage to the nerves, so there was a difference in the readings of the right and left foot, while the difference decreased in the case of the patient wearing AFO because the AFO helped the patient to walk, and therefore the readings converged with the readings of the normal person.

Figure 8 demonstrates that the deformed foot's footprint is less complete than the healthy foot's footprint, changing the value of ground reaction force on the foot as a result. While in the Fig. 9 demonstrate the deformed foot's footprint is close to the healthy foot's footprint. Figure 10 demonstrate footprints are more visible than they are in the Fig. 8 and the deformed foot's footprint is very close to the healthy foot's footprint.

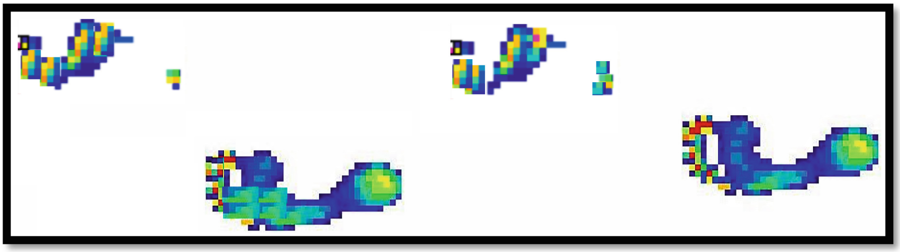


Fig. 8. The foot Print without wearing the AFO

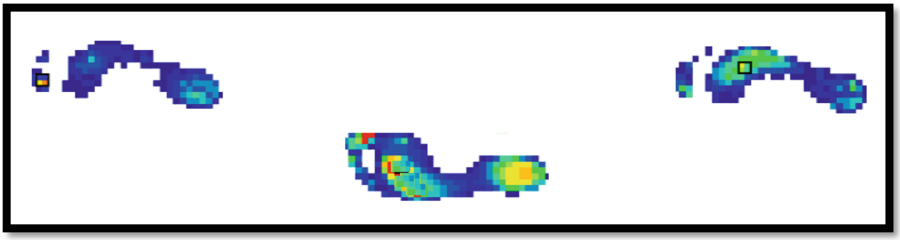


Fig. 9. The foot Print with wearing PP AFO

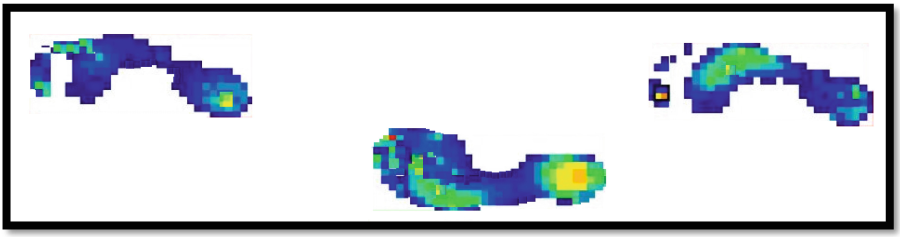


Fig. 10. The foot Print with wearing composite material AFO

8 The Ground Reaction Force

From Fig. 11 we notice that there is a difference between the heel strike of the right and left foot, where the force value of the left heel strike was approximately equal to 740 N, while for the injured (right) foot it was very little, equal to 580 N, while the value of the toe increased in the case of the injured foot and became 780 N, while for a healthy foot it was 700 N. We also note that in the event that the AFO was not worn, the patient needed a longer time to complete one cycle of his gait.

While we can notice a very clear difference in the case of the patient wearing a polypropylene AFO, where the values of the heel strike of the healthy and injured feet converged and became approximately 800 N. The time it took to complete one cycle of walking became less than it had previously been as shown in Fig. 12 (Fig. 13).

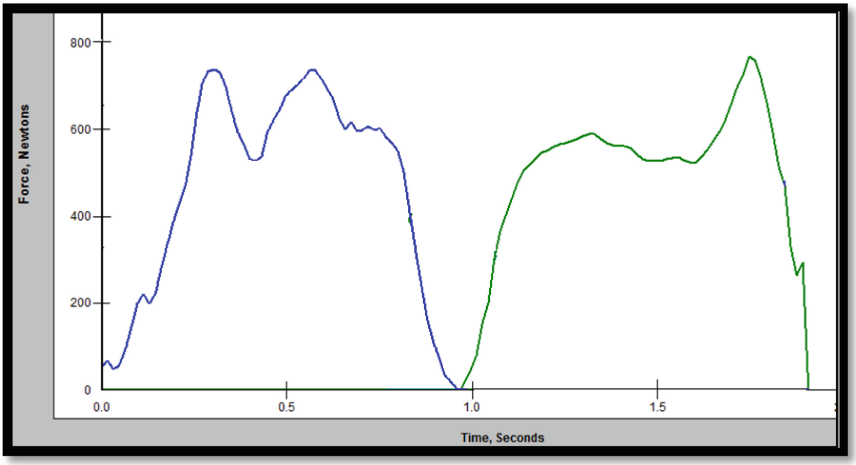


Fig. 11. The ground reaction force of the left and right leg when the patient walks without the AFO

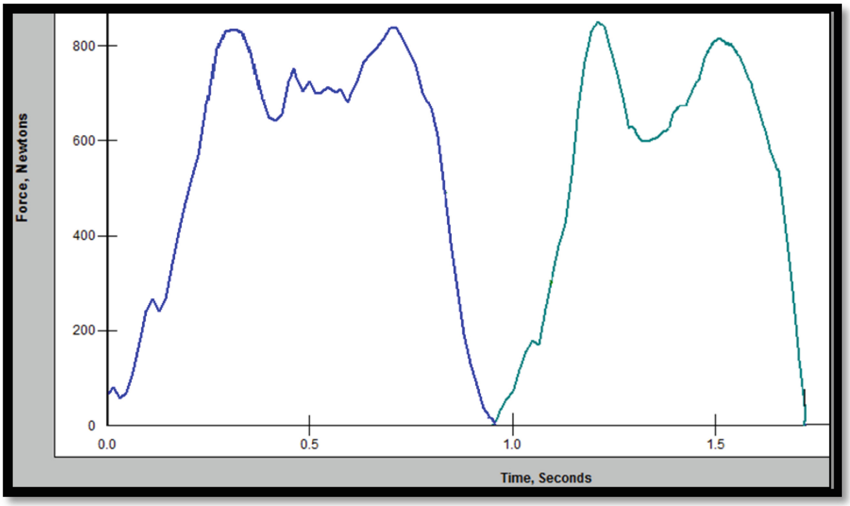


Fig. 12. The ground reaction force of the left and right leg when the patient walks with the PP AFO

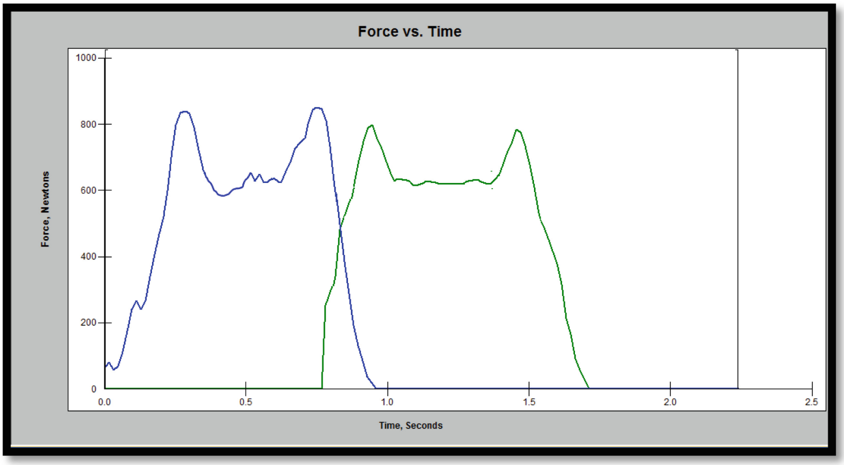


Fig. 13. The ground reaction force of the left and right leg when the patient walks with composite material AFO

9 Conclusions

1. The manufacturing of composite ankle-foot orthoses using perlon, carbon fiber, Kevlar fiber, and C-orthocryl lamination resin offers the patient's leg excellent control throughout the gait cycle. The patient's gait cycle is corrected using ankle-foot orthoses.

2. When the gait cycle of patients wearing plastic AFOs, composite material AFOs, and those without AFOs was compared, it was apparent that the three instances were very different from one another.
3. The ultimate strength and the measured yield of the composite AFO (sequence 1) were Ϭult= 67 MPa and Ϭy= 50 MPa, and for the (sequence 2) were Ϭult= 80 MPa and Ϭy= 63 MPa, respectively. This shows that the AFO has an acceptable stiffness in comparison with polypropylene AFO (Ϭy= 24.73 MPa, Ϭult = 35.76 MPa) that is often used for AFO fabrication.
4. The deformed foot's footprint is less complete than the healthy foot's footprint without wearing AFO, changing the value of ground reaction force on the foot as a result. In the event that the patient wears the PP AFO the foot's footprint is close to the healthy foot's footprint. And if the patient wears composite material AFO, the footprints become clear and very close to the healthy footprints.
5. The gait cycle obtained for the patient wearing the AFO shows a significant difference for a patient without using the AFO. This gives a confidence for the patient wearing AFO using the materials and suggested lamination sequence.
6. The suggested laminations for the composite sequences gave good results and safe the Von-Mises stress and safety factors in using the alternating stresses obtained using the fatigue results. This will allow more time to wear the suggested AFO, therefore a recommendation is given for the rehabilitation centers to manufacture such types of AFO.

References

1. B.A.: The Biomechanics Of The Foot, Vol. 10, Clinical Prosthetics And Orthotics (1986)
2. Jweeg, M.J., Ameen, S.H.: Experimental and theoretical investigations of dorsiflexion angle and life of an ankle-Foot-Orthosis made from (Perlon-carbon fibre-acrylic) and polypropylene materials. In: 10th IMEKO TC15 Youth Symposium on Experimental Solid Mechanics (2011)
3. Jweeg, M.J., Resan, K.K., Ismail, M.T.: Study of creep-fatigue interaction in a prosthetic socket below knee. In: ASME International Mechanical Engineering Congress and Exposition (2012)
4. Douglas, H., Lawrence, H.: Treating Dropfoot With Ankle-Foot Orthosis. Foot And Ankle Center Of Washington (2009)
5. Jweeg, M.J., Alhumandy, A.A., Hamzah, H.A.: Material Characterization and Stress Analysis of Openings in Syme's Prosthetics International Journal of Mechanical & Mechatronics Engineering IJMME-IJENS, Vol. 17, No. 04, (2017)
6. Jweeg, M.J., Hammoudi, Z.S., Alwan, B.A.: Optimised analysis, design, and fabrication of trans-tibial prosthetic sockets. In: IOP Conference Series: Materials Science and Engineering, 2nd International Conference on Engineering Sciences, vol. 433 (2018)
7. Jweeg, M.J., Ahumdany, A.A., Mohammed Jawad, A.F.: Dynamic stresses and deformations investigation of the below knee prosthesis using CT-Scan modeling. Int. J. Mech. Mechatron. Eng. IJMME-IJENS **19**(01) (2019)
8. Rolf, M.N., Jorunn, L.H.: Estimation of gait cycle characteristics by trunk accelerometry. J. Biomech. 121–126 (2004)
9. Abbas, S.M., Takhakh, A.M., Al-Shammari, M.A., Al-Waily, M.: Manufacturing and analysis of ankle disarticulation prosthetic socket (SYMES). Int. J. Mech. Eng. Technol. (IJMET) **09**(07), 560–569 (2018)

10. Abbas, S.M., Resan, K.K., Muhammad, A.K., Al-Waily, M.: Mechanical and fatigue behaviors of prosthetic for partial foot amputation with various composite materials types effect. Int. J. Mech. Eng. Technol. (IJMET) **09**(09), 383–394 (2018)
11. Jweeg, M.J., Al-Waily, M., Muhammad, A.K., Resan, K.K.: Effects of temperature on the characterisation of a new design for a non-articulated prosthetic foot. In: IOP Conference Series: Materials Science and Engineering, vol. 433, 2nd International Conference on Engineering Sciences, Kerbala, Iraq, 26–27 March 2018
12. Al-Waily, M., Hussein, E.Q., Al-Roubaiee, N.A.A.: Numerical modeling for mechanical characteristics study of different materials artificial hip joint with inclination and gait cycle angle effect. J. Mech. Eng. Res. Dev. (JMERD) **42**(04), 79–93 (2019)
13. Ryo, T., Shigeru, T., Masahiro, T., Manabu, M., Minoru, N.: Gait analysis using gravitational acceleration measured by wearable sensors. J. Biomech. 223–233 (2009)
14. Abbas, E.N., Jweeg, M.J., Al-Waily, M.: Fatigue characterization of laminated composites used in prosthetic sockets manufacturing. J. Mech. Eng. Res. Dev. **43**(5), 384–399 (2020)
15. Al-Waily, M., Al Saffar, I.Q., Hussein, S.G., Al-Shammari, M.A.: Life enhancement of partial removable denture made by biomaterials reinforced by graphene nanoplates and hydroxyapatite with the aid of artificial neural network. J. Mech. Eng. Res. Dev. **43**(6), 269–285 (2020)
16. Al-Waily, M., Tolephih, M.H., Jweeg, M.J.: Fatigue characterization for composite materials used in artificial socket prostheses with the adding of nanoparticles. In: IOP Conference Series: Materials Science and Engineering, 2nd International Scientific Conference of Al-Ayen University, Vol. 928, (2020)
17. Mechi, S.A., Al-Waily, M., Al-Khatat, A.: The mechanical properties of the lower limb socket material using natural fibers: a review. Mater. Sci. Forum **1039**, 473–492 (2021)
18. Weizhen, M.: Instrumentation of Gait Analysis, Master of Science Thesis Stockholm, Sweden (2010)
19. Mechi, S., Al-Waily, A.M.: Impact and mechanical properties modifying for below knee prosthesis socket laminations by using natural kenaf fiber. In: 3rd International Scientific Conference of Engineering Sciences and Advances Technologies, Journal of Physics: Conference Series, vol. 1973 (2021)
20. Al-Waily, M., Jweeg, M.J., Jebur, Q.H., Resan, K.K.: Creep characterization of various prosthetic and orthotics composite materials with nanoparticles using an experimental program and an artificial neural network. Materials Today: Proceedings (2021)
21. Jweeg, M.J., Hamdan, Z., Majeed, A.H., Resan, K.K., Al-Waily, M.: A new method for measurement the residual stresses in friction stir welding. Arch. Mater. Sci. Eng. **112**(2), 63–69 (2021)
22. Haider, S.M.J., Takhakh, A.M., Al-Waily, M.: A review study on measurement and evaluation of prosthesis testing platform during gait cycle within sagittal plane. In: 14th International Conference on Developments in eSystems Engineering. IEEE Xplore (2021)
23. Amanda, E.C.: Gait Deviations and the Immediate Effects of Ankle-Foot Orthotics and Functional Electrical Stimulation, PhD Thesis, Graduate Department of Rehabilitation Science, University of Toronto (2012)
24. Haider, S.M.J., Takhakh, A.M., Al-Waily, M., Saadi, Y.: Simulation of gait cycle in sagittal plane for above-knee prosthesis. In: 3rd International Scientific Conference of Alkafeel University, AIP Conference Proceedings, vol. 2386 (2022)
25. Haider, S.M.J., Takhakh, A.M., Al-Waily, M.: Designing a 3D virtual test platform for evaluating prosthetic knee joint performance during the walking cycle. Open Eng. **12**, 590–604 (2022)

26. Kirtley, C.: Clinical Gait Analysis: Theory And Practice. Elsevier, Edinburgh, New York (2006)
27. E.M.: Gait Analysis Of Normal And Differently Abled Subjects For Rehabilitation, Msc. Thesis In Department Of Biotechnology And Medical Engineering National Institute Of Technology Rourkela (2014)

Arabic Sign Language Alphabet Classification via Transfer Learning

Muhammad Al-Barham[1], Osama Ahmad Alomari[1], and Ashraf Elnagar[2](✉)

[1] MLALP Research Group, University of Sharjah, Sharjah, United Arab Emirates
{malbarham,oalomari}@shrajah.ac.ae

[2] Department of Computer Science, University of Sharjah, Sharjah, United Arab Emirates
ashraf@shrajah.ac.ae

Abstract. The integration of artificial intelligence (AI) has addressed the challenges associated with communication with the deaf community, which requires proficiency in various sign languages. This research paper presents the RGB Arabic Alphabet Sign Language (ArASL) dataset, the first publicly available high-quality RGB dataset. The dataset consists of 7,856 meticulously labeled RGB images representing the Arabic sign language alphabets. Its primary objective is to facilitate the development of practical Arabic sign language classification models. The dataset was carefully compiled with the participation of over 200 individuals, considering factors such as lighting conditions, backgrounds, image orientations, sizes, and resolutions. Domain experts ensured the dataset's reliability through rigorous validation and filtering. Four models were trained using the ArASL dataset, with RESNET18 achieving the highest accuracy of 96.77%. The accessibility of ArASL on Kaggle encourages its use by researchers and practitioners, making it a valuable resource in the field (https://www.kaggle.com/datasets/muhammadalbrham/rgb-arabic-alphabets-sign-language-dataset).

Keywords: Sign-Language · Dataset · Deaf · Arabic · Alphabet

1 Introduction

Arabic Sign Language (ArSL) serves as the predominant communication method for individuals with hearing impairments within the Arab world. As a visual language, ArSL employs a combination of hand gestures, facial expressions, and body movements to convey thoughts and meaning. Similar to spoken languages, sign languages possess distinctive grammatical structures, vocabulary, and phonological characteristics.

Deaf individuals face significant challenges when it comes to communicating with the hearing community. The communication barrier often isolates them from fully participating in social, educational, and professional activities. While written Arabic can be used for written communication, it does not adequately

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 226–237, 2024.
https://doi.org/10.1007/978-3-031-56728-5_20

address the need for real-time interaction. Consequently, the ability to effectively express thoughts, emotions, and intentions may be limited for deaf individuals.

Deep learning, a subfield within the domain of artificial intelligence (AI), has brought about significant transformations in numerous fields, including computer vision and natural language processing. At its core, deep learning entails training intricate neural network models to autonomously discern patterns and construct representations from input data. Within the context of sign language recognition, deep learning models, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have proven to be effective tools in the identification and interpretation of diverse sign languages utilized across the globe.

The application of artificial intelligence (AI) and deep learning methodologies has emerged as a promising avenue to address the communication barrier that exists between the deaf and hearing communities. These technological advancements offer the opportunity to create Arabic sign language recognition systems capable of automatically translating sign language into written or spoken languages. Such systems not only facilitate seamless communication but also enhance the overall communication experience for both deaf and hearing individuals by enabling real-time interpretation and understanding. Leveraging the capabilities of deep learning techniques in sign language recognition holds immense promise for transforming communication experiences for individuals employing Arabic Sign Language. Through the utilization of artificial intelligence (AI) and deep learning methodologies, we can develop resilient and precise systems that enable instantaneous translation of sign language into written or spoken languages. These advancements will play a pivotal role in fostering effective communication, promoting inclusivity, and establishing a foundation of equality between deaf and hearing individuals.

Moreover, deep learning-based sign language recognition systems have the potential to assist in educational settings, providing deaf students with access to information and resources typically delivered through spoken or written language. By automating the interpretation process, these systems can contribute to a more inclusive learning environment.

Research and development efforts have been made to apply deep learning techniques to recognize sign languages, including American Sign Language (ASL)[2,10], British Sign Language (BSL) [8], and Japanese Sign Language (JSL) [7]. These systems strive to utilize AI to empower the deaf community by eliminating communication barriers.

For instance, the study in [1] aims to address the educational challenges faced by deaf and dumb Muslims in their pursuit of advanced education and understanding of the Holy Qur'an. have developed a model that utilizes deep learning algorithms to recognize Qur'anic sign language, enabling individuals with hearing impairments to learn, recite, and comprehend the profound meanings and interpretations of the Holy Qur'an. This study can help deaf people by providing them with a tool to learn and understand the Arabic alphabet, which is the language of the Holy Qur'an, and overcome communication challenges with others.

In Summary, Arabic Sign Language recognition systems based on deep learning techniques hold tremendous promise for overcoming communication barriers faced by deaf individuals. By harnessing the power of AI and deep learning, accurate and efficient systems can be developed, enhancing the communication experience and promoting inclusivity. The success of deep learning in recognizing sign languages such as ASL and BSL serves as evidence of the potential for similar advancements in Arabic Sign Language recognition.

This study presents a comprehensive exploration of various models that have been trained on a novel, meticulously curated RGB dataset specifically tailored for Arabic sign language (ArSL). The dataset, characterized by its high quality, serves as a valuable resource to facilitate the advancement of ArSL and its associated applications.

The paper is structured as follows: Sect. 2 provides an overview of the related work in the field. Section 3 delves into the research methodology, encompassing the dataset description, collection process, novelty aspects, pre-processing techniques, and data splitting strategies. Moreover, the section discusses the pre-trained models employed in this study. Section 4 presents the experimental setup and showcases the obtained results. Lastly, Sect. 5 encapsulates the principal discoveries and formulates definitive conclusions derived from the research outcomes.

2 Related Work

Researchers worldwide are currently demonstrating considerable interest in the development of tools that facilitate communication and alleviate barriers faced by individuals with hearing impairments. In this context, it is pertinent to discuss pertinent research endeavors specifically focused on the Arabic Sign language.

The research presented in [11] delves into the utilization of transfer learning and fine-tuning techniques within deep convolutional neural networks to enhance the accuracy of recognizing 32 distinct hand gestures from the Arabic sign language. To address the issue arising from class size inconsistencies, the authors leveraged the ArASL dataset and implemented random under-sampling. Through this approach, the resulting model achieved remarkable validation accuracies of 99.4% for VGG16 and 99.6% for ResNet152. The study not only demonstrated the significant potential of employing Arabic sign language imagery for effective classification but also introduced the concept of fine-tuning to obviate the necessity of amassing an extensive dataset for training purposes. By harnessing the power of deep learning, the authors propose that the benefits of machine learning technology can be effectively transferred to provide superior solutions to challenges encountered in the realm of Arabic sign language.

The research conducted in [6] introduces an innovative deep learning-based approach for effectively recognizing letters of the Arabic Sign Language. The proposed system combines image pre-processing techniques with a robust Deep Convolutional Neural Network architecture to facilitate automatic detection and recognition of hand-sign letters corresponding to the Arabic alphabet. By training the system on a dataset comprising 5000 RGB images and integrating a purpose-designed pre-processing stage, the model achieved an impressive accuracy rate of 97.07% in accurately identifying static Arabic sign letters. The study concludes that the proposed system holds tremendous potential in overcoming communication barriers faced by the deaf-mute community and can be further extended to recognize dynamic signs in future applications.

The study presented in [4] introduces a comprehensive system designed for recognizing and classifying Arabic Sign Language (ArSL) gestures by employing machine learning algorithms. The system incorporates five distinct classification algorithms, namely Decision Tree, K-Nearest Neighbor, Naive Bayes classifier, Random forest, and Stochastic Gradient Descent. To facilitate accurate classification, the authors curated a dataset comprising 32 unique ArSL characters, with each image encompassing a feature vector containing 31 distinct features. Through thorough testing and comparison, the authors determined the optimal methods that yielded the most favorable outcomes. Notably, the experimental findings revealed that an image size of 20×20 proved to be optimal for the dataset. The results demonstrated that the k-Nearest Neighbor (k-NN) algorithm exhibited the highest accuracy of 86%. Furthermore, the authors extended their investigations to apply the machine learning algorithms to the initial data collected from multiple videos, affirming the superior performance of the k-NN algorithm. The authors propose that their system has the potential to recognize finger movements in videos and subsequently translate the identified Arabic signs into text or voice. To facilitate further research, the authors have developed two databases-one containing still images and the other containing videos-both of which will be made freely accessible to the research community.

The study presented in [5] conducts a comparative analysis of three prevalent deep learning models, namely AlexNet, VGGNet, and GoogleNet/Inception, for the offline classification and recognition of Arabic sign language alphabets. To train and evaluate these models, the study utilized the ArSL2018 dataset, which represents the most up-to-date publicly available collection of Arabic sign language images, consisting of 54,000 samples. The findings revealed that the VGGNet model surpassed other convolutional neural network (CNN) models, achieving an impressive accuracy of 97% on the test dataset. To further enhance model performance and optimize hyperparameters, the study employed k-fold cross-validation with 10 iterations for data splitting during training and testing phases. This iterative approach yielded improved accuracy results. The paper concludes that the proposed approach exhibits promising outcomes and can serve as a valuable tool for individuals with specific needs. Additionally, the study acknowledges the challenges associated with sign language recognition and suggests future directions for research in this domain.

The research paper discussed in [12] introduces an innovative system designed for recognizing Arabic sign language through the utilization of transfer learning techniques. The proposed system comprises several crucial steps, including image acquisition, preprocessing, and feature extraction, leveraging pretrained models. The authors conducted experiments with various Keras pretrained models and discovered that the EfficientNetB4 model outperformed the others in terms of performance. The system underwent training using a diverse dataset and integrated several data augmentation techniques to enhance its adaptability across different environmental conditions. The results demonstrated that the proposed method surpassed previous approaches in terms of class recognition and showcased greater practicality in real-time settings. The authors conclude that their system represents a leading solution for Arabic sign language recognition, offering significant potential to enhance communication within the deaf and mute communities.

From the previous work, we observe that there are numerous needs from the researchers for a dataset with high quality that can be utilized in real-life applications. Therefore, our paper aims to show a new dataset with trained models to support this field.

3 Research Methodology

In this section, a detailed examination of the dataset utilized in this paper is provided, encompassing its collection methodology and relevant characteristics. More details can be found in [3]. Additionally, the section delves into the data pre-processing and division process, incorporating the utilization of pre-trained models.

3.1 Data Description

The ArASL (Arabic Alphabet Sign Language) dataset is an outcome of a collective undertaking involving over 200 contributors who shared their knowledge of various alphabets. Numerous images in the dataset were captured using diverse devices, such as webcams, digital cameras, and mobile phone cameras. Within the ArASL dataset, there are 7,856 labeled images that represent the Arabic sign language alphabet. A team of proficient Arabic sign language specialists oversaw the dataset, validating and refining the images to guarantee a dataset of superior quality.

The dataset is structured into 31 distinct folders, with each folder corresponding to a specific alphabet. The distribution of images across these folders is summarized in the accompanying Table 1. Furthermore, Fig. 1 showcases a collection of sample images representing different alphabets.

3.2 Data Collection

To facilitate the process of data collection, an online form accompanied by a comprehensive set of instructions was meticulously prepared. The alphabets were systematically categorized into five distinct groups for the participants, with the first four categories consisting of six alphabets each, while the fifth and final category encompassed the remaining seven alphabets. Notably, among the thirty-one ArSL alphabets, sixteen have been assigned special instructions derived from expert insights, aimed at mitigating common errors encountered

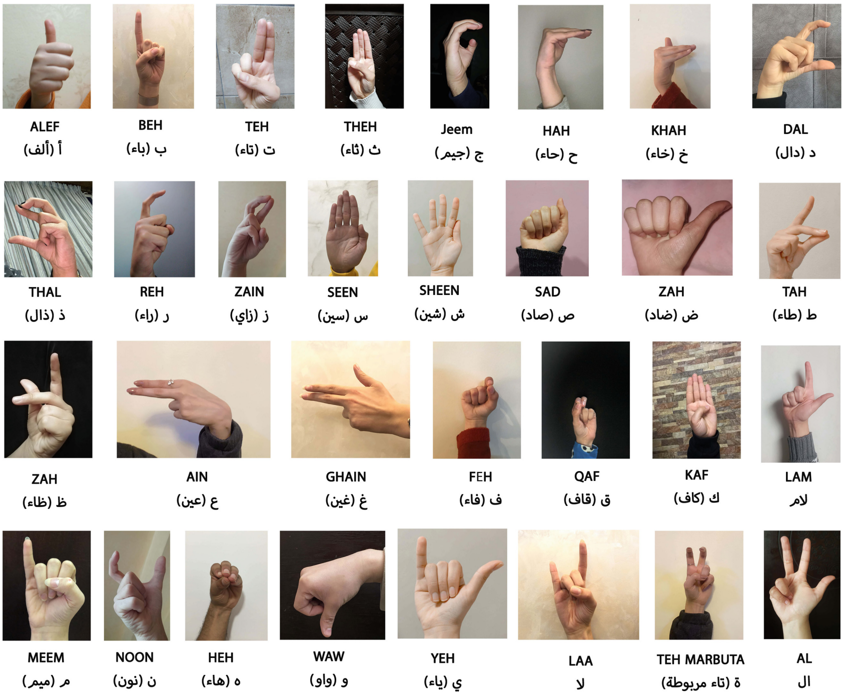


Fig. 1. A subset of the alphabets extracted from the ArASL dataset.

during the process of capturing the alphabets. These instructions are informed by the collective experience and expertise of the professionals involved in the development of the dataset. Participants were afforded the flexibility to submit images of the alphabets they were most comfortable performing. As a result, there were no constraints imposed on the number of images that a participant could submit. The link to access the online form was disseminated across various social media platforms, attracting a diverse range of participants from educational institutions such as schools and universities, spanning different age groups and genders. The images captured by the participants exhibited considerable variability, owing to the utilization of different camera types, backgrounds, lighting conditions, and image sizes. Throughout the data collection process, utmost care was taken to ensure the anonymity and confidentiality of the participants' identities.

3.3 Data Novality

The dataset exhibits versatility as it encompasses a wide range of data collected under varying settings, including diverse lighting conditions, backgrounds, image orientations, sizes, and resolutions. This comprehensive dataset proves to be well-suited for the development and training of machine learning algorithms aimed at Arabic sign language classification. It is important to note that this dataset

Table 1. Distribution of the alphabets in the dataset.

#	Letter name in English Script	Letter name in Arabic Script	# of Images	#	Letter name in English Script	Letter name in Arabic Script	# of Images
1	ALEF	أ (أْلف)	287	17	ZAH	ض (ظاءْ)	232
2	BEH	ب (باءْ)	307	18	AIN	ع (عين)	244
3	TEH	ت (تاءْ)	226	19	GHAIN	غ (غين)	231
4	THEH	ث (ثاءْ)	305	20	FEH	ف (فاءْ)	255
5	JEEM	ج (جيم)	210	21	QAF	ق (قاف)	219
6	HAH	ح (حاءْ)	246	22	KAF	ك (كاف)	264
7	KHAH	خ (خاءْ)	250	23	LAM	ل (لام)	260
8	DAL	د (دال)	235	24	MEEM	م (ميم)	253
9	THAL	ذ (ذال)	202	25	NOON	ن (نون)	237
10	REH	ر (راءْ)	227	26	HEH	ه (هاءْ)	253
11	ZAIN	ز (زاي)	201	27	WAW	و (واو)	249
12	SEEN	س (سين)	266	28	YEH	ي (ياءْ)	272
13	SHEEN	ش (شين)	278	29	TEH MARBUTA	ة (تاءْ مربوطة)	257
14	SAD	ص (صاد)	270	30	AL	ال	276
15	DAD	ض (ضاد)	266	31	LAA	لا	268
16	TAH	ط (طاءْ)	227				

has undergone rigorous verification and validation processes by domain experts, ensuring its reliability and accuracy. Notably, this dataset represents a pioneering contribution as it is, to the best of our knowledge, the first publicly available RGB high-resolution dataset specifically tailored for Arabic sign language research.

On a final note, images of our dataset are raw in nature, and thus interested researchers are left to perform any necessary processing they may need. Also, this work has been inspired by ArASL (Arabic Alphabets Sign Language) Dataset [9].

3.4 Data Pre-processing and Splitting

In deep learning, preprocessing plays a vital role in preparing images for input into the model. The dataset used in this study consists of images with varying sizes, necessitating preprocessing as a crucial preparatory step. The objective of preprocessing is to ensure that the images are compatible with the requirements of the model. To achieve this, a series of standardized preprocessing steps were applied to each image in the dataset. Firstly, all images were resized to a dimension of 224 pixels, which is a common size used in many deep learning architectures. This resizing step ensures that the images have a consistent size, regardless of their original dimensions.

Additionally, center cropping was performed on the resized images. Center cropping involves removing the outer regions of the image to focus on the central part, which often contains the most relevant information. This step helps in reducing the influence of irrelevant background or peripheral elements, thereby enhancing the model's ability to capture the essential features. Furthermore, the preprocessed images were converted into a Tensor format suitable for further analysis. Tensors are multi-dimensional arrays that serve as the fundamental data structure in deep learning frameworks. By converting the images into tensors, they can be efficiently processed and manipulated by the deep learning model during training and evaluation.

By applying these standardized preprocessing steps to all the images in the dataset, uniformity is achieved, ensuring that each image is represented in a consistent format. This standardized preprocessing not only facilitates the training and evaluation processes but also helps in maintaining the integrity and comparability of the results obtained from the model.

The dataset was divided into three subsets: a training set comprising 70% of the data, a validation set consisting of 15%, and a testing set also accounting for 15%. This partitioning scheme ensures a representative distribution of the dataset across the different phases of model development and evaluation.

To gain insights into the dataset's characteristics, the histogram distribution was analyzed and visualized. Figure 2 showcases the histogram distribution, providing a graphical representation of the frequency distribution of the dataset.

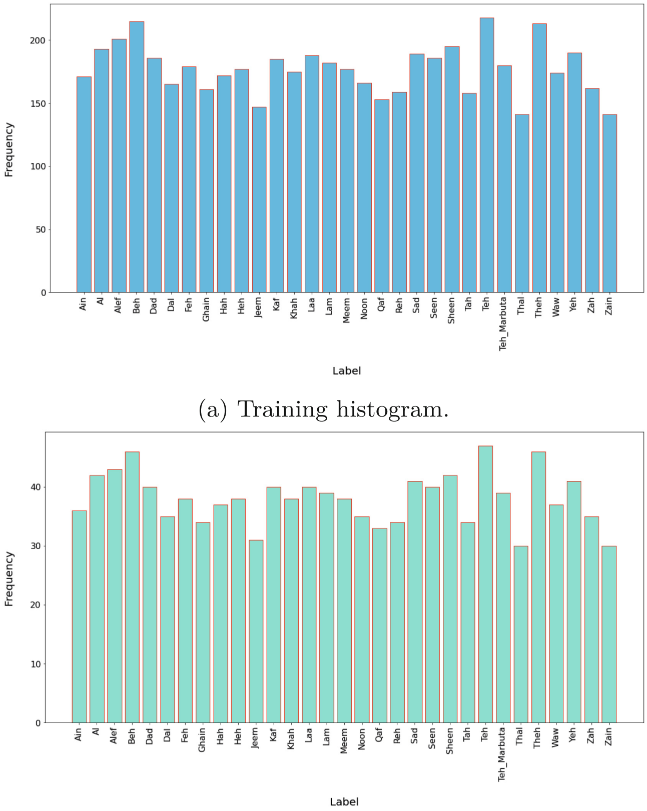
3.5 Pre-trained Models

This research incorporates a range of models, including RESNET18, VGG16, AlexNet, and SqueezeNet, utilizing their respective weights and architectures acquired through PyTorch. The study explores two distinct approaches following the removal of the head layer from these models. The first approach involves freezing all layers except the head layer, while the second approach maintains the trainability of all model parameters throughout the training process.

4 Experiments and Results

Several experiments were conducted on the datasets using pre-trained models, The results of the experiments, as shown in Table 2, indicate that keeping all model parameters trainable yields superior performance.

To prevent overfitting and improve generalization, the early stopping regularization technique was implemented. A patience parameter of 10 epochs was used, monitoring the validation loss during training. All models were trained using the SGD optimizer, which demonstrated excellent performance throughout the training process. The batch size used is 64. Evaluation of the testing results for all models, including average testing accuracy, precision, recall, and F1 score, was performed using Sklearn metrics with the macro average method. Only models with an accuracy above 70% were reported in Table 2.



(a) Training histogram.

(b) Validation/Testing histogram.

Fig. 2. Histogram for the splitted dataset.

A learning rate of 0.1 and 0.01 was utilized across all models. Based on the results, the Resnet18 model without freezing the layers and with a learning rate of 0.1 emerged as the best model. All training and evaluations codes are available in Github[1].

Figure 3 presents the validation accuracy of all models throughout the training process. The results clearly indicate that the top-performing model exhibits superior performance compared to the other models, achieving the highest accuracy rate while also demonstrating the fastest convergence.

Figure 4 displays the confusion matrix of the best model, highlighting its proficiency in accurately classifying the alphabet even though there are some mismatches, it is still an excellent model.

[1] https://github.com/mohammad-albarham/Arabic-Sign-Language-Alphabet-Classification-via-Transfer-Learning.

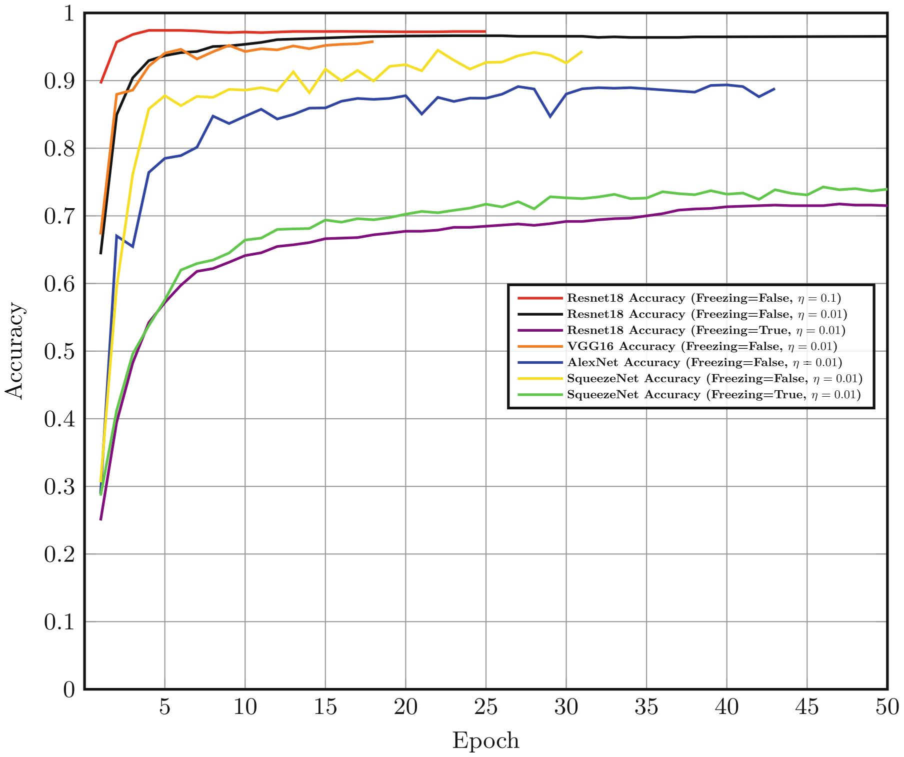


Fig. 3. Validation Accuracy of the models listed in Table 2.

Table 2. Testing Results for Various Models.

Model	Freezing	lr	Accuracy	Precision	Recall	F1 score
RESNET18	False	1.0×10^{-1}	96.77%	96.85%	96.66%	96.71%
	False	1.0×10^{-2}	95.76%	95.73%	95.62%	95.63%
	True	1.0×10^{-2}	71.05%	71.28%	70.70%	70.74%
VGG16	False	1.0×10^{-2}	94.74%	94.78%	94.55%	94.55%
AlexNet	False	1.0×10^{-2}	89.13%	89.23%	88.95%	88.95%
SqueezeNet	False	1.0×10^{-2}	93.55%	93.55%	93.35%	93.36%
	True	1.0×10^{-2}	72.67%	74.39%	72.43%	72.67%

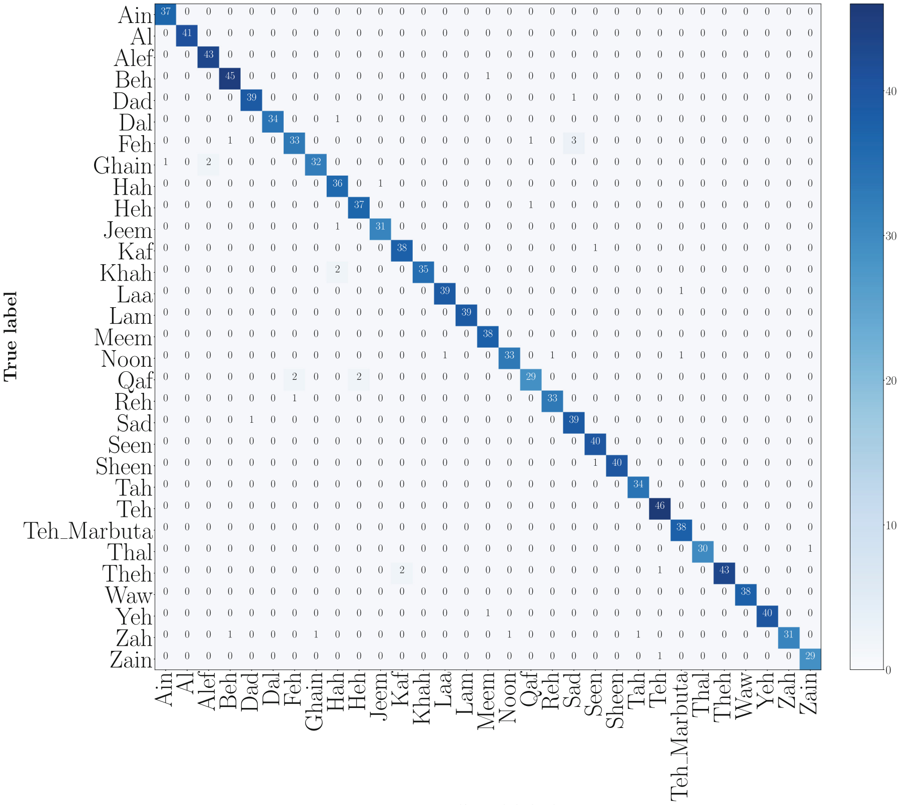


Fig. 4. Confusion Matrix of Resnet18 (Freezing = False, $\eta = 0.1$)

5 Conclusion

Through extensive experimentation, it has been demonstrated that the utilized dataset performs exceptionally well during testing, thereby showcasing its potential for practical applications in bridging the communication gap between individuals utilizing Arabic Sign Language and others. Notably, the findings indicate that maintaining trainable parameters enables the model to acquire improved learning capabilities. Among the models evaluated, ResNet18 exhibited the highest accuracy on the testing dataset when the layers were not frozen and a learning rate of 0.1 was employed. Looking ahead, future endeavors could involve expanding the dataset to encompass Object Detection tasks through the process of annotation and modeling. Additionally, the development of a real-life application stands as a viable avenue for further exploration.

Acknowledgment. We extend our sincere appreciation to the Student Counseling Department at the University of Jordan for their valuable guidance and expertise, which greatly contributed to acquiring accurate and appropriate images based on their extensive experience. Furthermore, we express our gratitude to Jana M. AlNatour and Raneem F. Abdelraheem for their valuable assistance in the meticulous data collection process. Additionally, we would like to extend our special thanks to the Research Institute for Sciences and Engineering (RISE) at the University of Sharjah for their support and collaboration.

References

1. AbdElghfar, H.A., et al.: A model for qur'anic sign language recognition based on deep learning algorithms. J. Sensors **2023** (2023)
2. Abdulhussein, A.A., Raheem, F.A.: Hand gesture recognition of static letters American sign language (ASL) using deep learning. Eng. Technol. J. **38**(6), 926–937 (2020)
3. Al-Barham, M., et al.: RGB Arabic alphabets sign language dataset. arXiv preprint arXiv:2301.11932 (2023)
4. ALtememe, M.S., El Abbadi, N.K.: Gesture interpreting of alphabet Arabic sign language based on machine learning algorithms. In: 2022 Iraqi International Conference on Communication and Information Technologies (IICCIT), pp. 177–183. IEEE (2022)
5. Duwairi, R.M., Halloush, Z.A.: Automatic recognition of Arabic alphabets sign language using deep learning. Int. J. Electr. Comput. Eng. (2088-8708) **12**(3) (2022)
6. Hdioud, B., Tirari, M.E.H.: A deep learning based approach for recognition of Arabic sign language letters. Int. J. Adv. Comput. Sci. Appl. **14**(4) (2023)
7. Ito, S.i., Ito, M., Fukumi, M.: Japanese sign language classification based on gathered images and neural networks. Int. J. Adv. Intell. Inform. **5**(3) (2019)
8. Kumar, K.: DEAF-BSL: deep learning framework for British sign language recognition. Trans. Asian Low-Resour. Lang. Inf. Process. **21**(5), 1–14 (2022)
9. Latif, G., Mohammad, N., Alghazo, J., AlKhalaf, R., AlKhalaf, R.: Arasl: Arabic alphabets sign language dataset. Data Brief **23**, 103777 (2019)
10. Pansare, J.R., Gawande, S.H., Ingle, M.: Real-time static hand gesture recognition for American sign language (ASL) in complex background (2012)
11. Saleh, Y., Issa, G.: Arabic sign language recognition through deep neural networks fine-tuning. Int. J. Online Biomed. Eng. (iJOE) **16**, 71 (2020).https://doi.org/10.3991/ijoe.v16i05.13087
12. Zakariah, M., Alotaibi, Y.A., Koundal, D., Guo, Y., Mamun Elahi, M., et al.: Sign language recognition for Arabic alphabets using transfer learning technique. Comput. Intell. Neurosci. **2022** (2022)

Evaluation of Chemical Data by Clustering Techniques

Gonca Ertürk[1](✉) and Oğuz Akpolat[2]

[1] Institute of Natural Sciences, Department of Chemistry, Muğla Sıtkı Koçman University, Muğla, Turkey
erturk.gonca@gmail.com

[2] Faculty of Science, Department of Chemistry, Muğla Sıtkı Koçman University, Muğla, Turkey

Abstract. Obtaining more useful information by applying mathematical techniques from chemical data obtained by different methods can be defined as chemometry. The development of computer-equipped devices allows for obtaining a large number of data in the field of chemistry. Statistical methods and data mining principles are needed for the processing and evaluation of these data. Chemometry is briefly the investigation of how to perform meaningful calculations on data fly the investigation of how to perform meaningful calculations on data. In most cases, these calculations are too complex to be performed manually, and many different techniques are used in these processes, such as classification, clustering, data summarization, learning classification rules, finding dependency networks, variability analysis, and abnormal detection. In data mining, classification, and curve fitting are defined as prediction methods, while methods such as clustering and association analysis are described as descriptive. Classification is the examination of the attributes of data and assigning this data to a predefined class. The important thing here is that the specialties of each class are determined in advance. Clustering is the grouping of data according to their proximity or distance to each other, and there are no pre-defined group boundaries here, but it can be optimized by giving the number of groups. In given context, this study was aimed to group the measurement data obtained as a result of analyses with samples taken from raw wastewater from wastewater treatment plants using the clustering method to determine which cluster the new data to be measured are in and to estimate the BOD_5 value related to these data without experimental measurement.

Keywords: Chemical analysis · Classification · Clustering · BOD5 · Estimation

1 Introduction

In a conceptual sense, data is any kind of event, situation, or idea that has been recorded. Data mining can be defined as the acquisition of previously unknown, valid, and applicable information from data stacks by a dynamic process. Data mining is the name given to data analysis techniques that are kept among very large data volumes, the meaning of which is extracted from potentially useful and understandable information that has not

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 238–250, 2024.
https://doi.org/10.1007/978-3-031-56728-5_21

been discovered before, and database management systems, statistics, artificial intelligence, machine learning, parallel and decoupled operations are in the background. In this process, many techniques are used, such as classification, clustering, data summarization, learning classification rules, finding dependency networks, variability analysis, and abnormal detection. In data mining, classification, and curve fitting are defined as prediction methods, while methods such as clustering and association analysis are described as descriptive. The main classification methods are decision trees, Bayesian classification, artificial neural networks, and decision support machinery. Classification is examining the attributes of a new object and assigning that object to a predefined class. The important thing here is that the specialties of each class are determined in advance. Clustering, on the other hand, is the grouping of data according to the proximity or distance of each other, there are no predetermined group limits, but it can be optimized by giving the number of groups [1].

1.1 Clustering Methods

The techniques used in data mining can be divided into models according to the type of data at hand and the intended use of the results obtained. These models can be grouped under two headings. These are predictive and descriptive models. Descriptive models extract relationships from the data set. The data mining techniques used in descriptive models are clustering, summarization, association rules, and sequential sequences. Predictive models, on the other hand, develop a model from situations whose results are already known, and with this model, they obtain new results from data sets whose results are unknown. Data mining techniques used in predictive models are classification, curve fitting, and time series. Clustering analysis divides the information in a data set into groups according to certain proximity criteria. In the clustering process, the similarity of the elements in the cluster should be high, and the similarity between the clusters should be low. Dec. Clustering falls into descriptive models from data mining techniques, i.e., unattended classification. In unattended classification, the goal is to separate an initially given and yet unclassified set of data in such a way that they form meaningful subsets. The clustering process is performed entirely according to the characteristics of the incoming data [2].

Some alternative measurements and methods can be used by taking into account similar distances in the use of clustering analysis. Euclidean, Standardized Euclidean, Manhattan Square, Decayed Euclidean, Minkowski, or Canberra measurements can be used for distances between units. This makes it necessary to act carefully in the use of clustering analysis in practice. The clustering algorithm divides the database into subsets. The elements in each cluster have common properties that distinguish the group they are included in from other groups. In clustering models, the goal is to find clusters in which the cluster members are very similar to each other, but whose properties are very different from each other, and to divide the records in the database into these different clusters [3].

There are many clustering algorithms used in data mining, and they are determined according to the structure of the data to be analyzed. Clustering methods are generally the following:

- Hierarchical Method: Before analyzing, the objects are organized according to a hierarchical structure. Various methods are used to translate the data into a hierarchical structure. Among them are the BIRCH and CURE methods.
- Model-Based Method: A model is determined for each cluster, and the data corresponding to this model is placed in the appropriate cluster (Fig. 1).

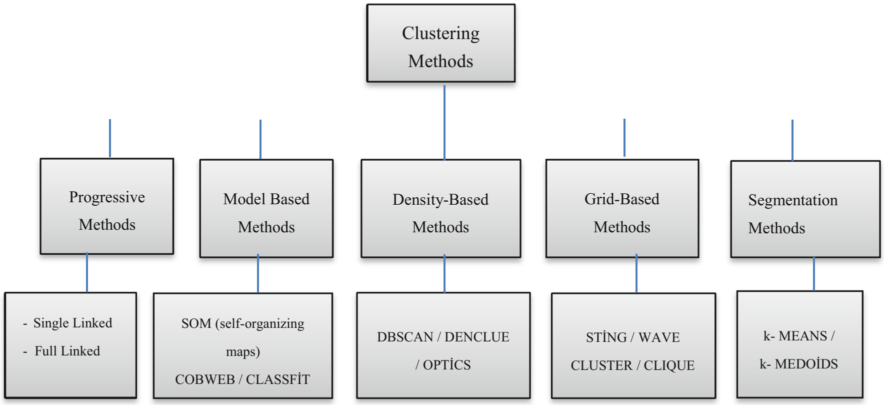


Fig. 1. Clustering Methods.

- Density-Based Method: Many clustering methods perform clustering according to the differences of objects Decoupled from each other. This method makes grouping according to the density of objects. Density is the number of objects analyzed. DBSCAN can be given as an example of density-based methods.
- Grid-Based Method: Classifies objects by their number to create a Grid structure. The main advantage is that it is completed quickly and is independent of the number of objects. Sting can be given as an example.
- The Segmentation Method: İn a database with n objects, the objects are analyzed by dividing them into logical groups. While several groups may exist in small and medium-sized databases, more groups may form when the database size increases. Different criteria can be evaluated while making the grouping. The grouping made affects the quality of the analysis.

In the hierarchical clustering method, the dendrogram is used especially for easier understanding of the functioning. The most commonly used methods are Single-link, Full-link, Average link, Central, and Ward methods. The non-hierarchical clustering method is preferred if there is preliminary information about the number of clusters or if the researcher has decided on the number of clusters that will be meaningful. The two most preferred methods in the non-hierarchical clustering method, are the k-means technique developed by Mac Queen and the likelihood technique [4].

1.2 K-Means Algorithm

It is one of the most widely used unsupervised learning methods. The K-means assignment mechanism allows each data to belong to only one set [5]. Therefore, it is a sharp clustering algorithm.

The general logic of the K-means algorithm is to partition a dataset consisting of "n" data objects into "k" sets given as input parameters. The aim is to ensure that the intra-cluster similarities of the clusters obtained at the end of the Decoupling process are maximum and the inter-cluster similarities are minimum. In this method, clustering is performed based on the Euclidean distance formula (1) [6].

$$P = (p1, p2, \ldots, pn)veq = (q1, q2, \ldots, qn) \tag{1}$$

$$\sqrt{\sum\nolimits_{i=1}^{n} (p_i - q_i)^2} = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_n - q_n)^2} \tag{2}$$

The K-Means algorithm starts with a randomly selected K (number of clusters) number of center points. Each point in the dataset is assigned to the set of the center point closest to it. The value of the cluster center is calculated by taking the average of its points. This process continues until the values of the centers do not change [7]. The process steps of the K-means algorithm are as follows:

Step 1. k objects are randomly selected. The selected k objects represent the cluster centers. $\mathrm{M_1}$, $\mathrm{M_2}$… $\mathrm{M_k}$. The sample midpoint is calculated as follows (2) [8].

$$M_k = \frac{1}{n_k} \sum\nolimits_{i=1}^{n_k} x_{ik} \tag{3}$$

Step 2. Intra-cluster changes are calculated using the Quadratic Error Formula (3) as shown in the formula. $\mathrm{e_1}$, $\mathrm{e_2}$,…$\mathrm{e_k}$ [9].

$$e_i^2 = \sqrt{\sum\nolimits_{i=1}^{n_k} (x_{ik} - M_K)^2} \tag{4}$$

For a space of all sets containing a set K, the square error is the sum of changes in the set. In this case, the square-error value in question (4) is calculated as shown in the formula.

$$E_k^2 = \sqrt{\sum\nolimits_{k=1}^{k} e_k^2} \tag{5}$$

Step 3. Each data is assigned to the cluster closest to it.

Step 4. When all the data are assigned to the nearest clusters, the centers are calculated for k clusters again.

Step 5. The second and third steps are repeated until there are no changes in the Cluster Centers.

2 Methodology

One of the areas of chemistry where a large number of data produced is environmental chemistry. The largest part of the pollution in wastewater consists of detergents, organic substances, and oils. The basic processes in the treatment of wastewater are to remove biodegradable organic substances, suspended solids, harmful heavy metals and toxic compounds, nitrogen, and phosphorus depending on the ambient conditions, and to destroy pathogenic organisms. The ability to monitor the treatment processes and provide the necessary controls is based only on the continuous determination of the characteristics of wastewater and activated sludge. The main measurement criteria for determining the properties of wastewater are biochemical oxygen demand (BOD_5), chemical oxygen demand (COD), and determination of total organic carbon (TOC) and dissolved oxygen (DO) amounts. Of these parameters, BOD_5 measurement takes at least 5 days, while others can be measured in short periods such as 1–2 h. If BOD_5 values can be mathematically related to other parameters, depending on them, it will be able to provide a great advantage in terms of controlling the estimated process in a short time. In this study, a data set was created by measuring the specified parameters from 334 samples taken from a treatment facility for statistical evaluation, and the interaction of the parameters contained in this data set with each other by the clustering method was examined. Thus, taking into account the weighted effects of the parameters, it was tried to estimate the probable BOD_5 value of a sample whose result is unknown. The algorithm selected for this data mining study was modeled with the Python program and the performance of the algorithm was examined in estimating the BOD_5 parameter depending on other parameters by extracting clustering rules.

In this section, it was measured for 334 days at a wastewater treatment plant used in cluster modeling performed using the Python programming language and presented in Table 1 [10]. These measured magnitudes are chemically: acidity (pH: Ph), Temperature (TemperC: °C), Total phosphate (TotPhosmgPL: mg/L), Suspended solids (CVLmgPL: mg/L), Chemical oxygen demand (CODmgPL: mg/L), Total Nitrogen (TotNitromgPL: mg/L**)** and Biological oxygen demand (BOD5mgPL: mg/L).

Table 1. The measured analytical values of chemical substances in wastewater samples.

Lab	Ph	TemperC	ToPhosmgPL	CVLmgPL	COmgPL	BOD5mgPL
1	7.3	8.7	17.7	310	920	19.41
2	7.55	9.7	15.9	150	495	169
3	7.47	10.3	11.6	180	401	209
4	8.03	9.7	5.2	130	433	272
EXPERIMENTAL DATA 334*7 In Size (**# Data: EnvirodataI.txt)**						
331	8	17	1.95	154	474	120
332	7.77	17.2	0.55	8.4	142.65	36.4
333	7.76	30	0.31	42	162.12	45.6
334	7.41	24.4	3.43	33	153.5	38.4

The PYTHON commands required for the analysis of the algorithm written for clustering in data evaluation are supplemented (**Ek**).

3 Results and Discussion

BOD5_PYTHON/Environment_Clustering_00_K_Means_med, written with Python commands for clustering and performance evaluation. The graphs of the program outputs are given in Fig. 2, Fig. 3, Fig. 4, and Fig. 5, respectively; "Distribution of samples according to BOD5 values", "Showing sample distributions in box graphics", "Histograms of distributions related to chemical measurement values", "Linking chemical measurement values to each other" and "Sample sets created according to BOD$_5$ values". In this section, the outputs and results of the clustering application selected for solving grouping problems are examined, and the BOD$_5$ numerical, and % distribution of the samples are given in Table 2 (Fig. 6).

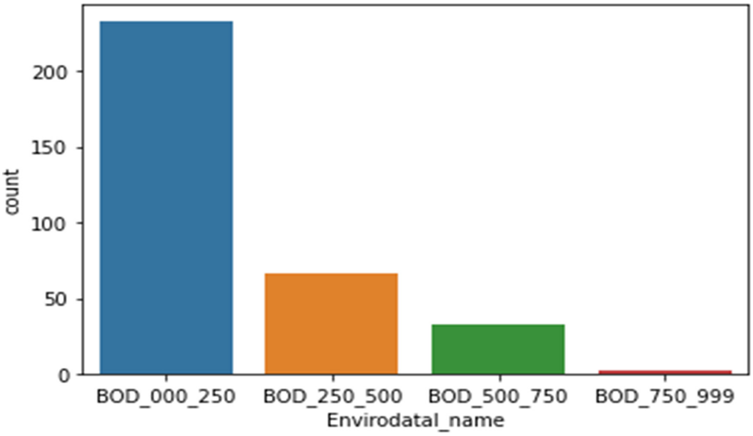


Fig. 2. BOD_5 distribution of samples according to their values.

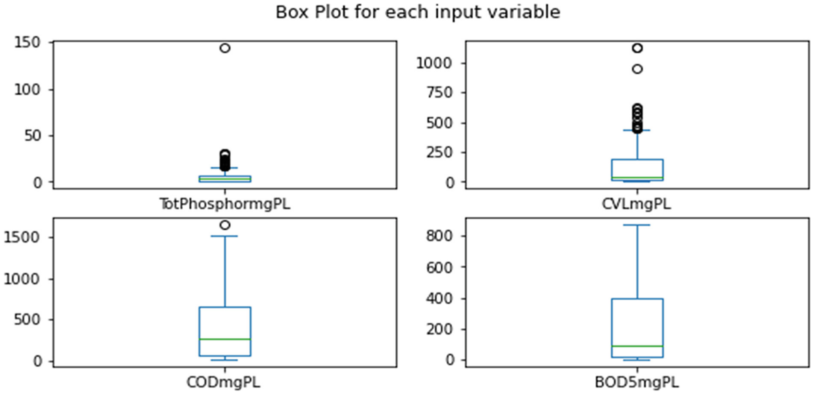


Fig. 3. Showing the sample distributions in box graphs.

Table 2. The determined BOD_5 percent distributions of all data

BOD_5 intervals	Numerical value	%Distribution
0–250	279	72
251–500	70	18
501–750	30	8
751–999	5	2
0–999	384	100

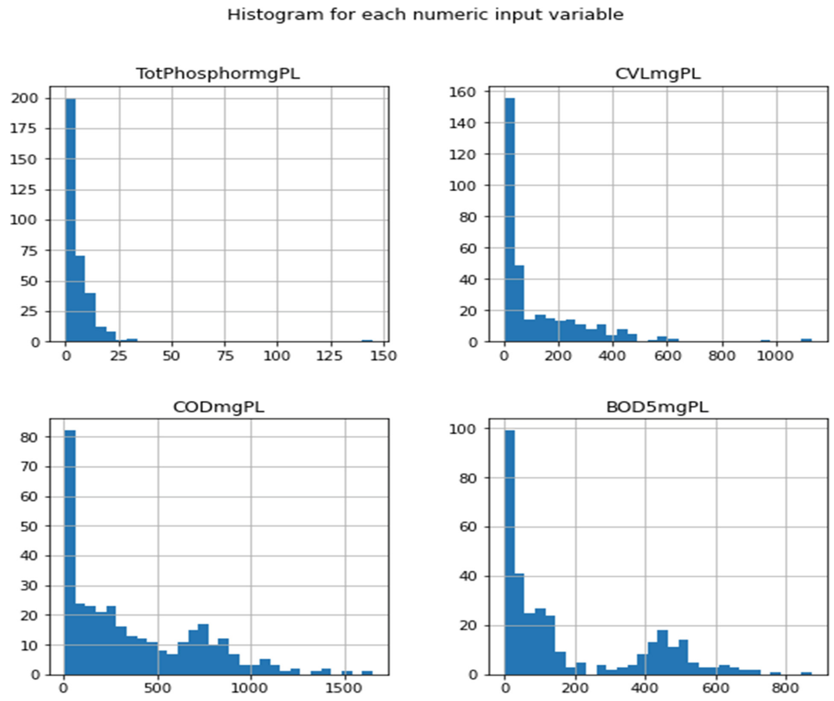


Fig. 4. Histograms of distributions related to chemical measurement values.

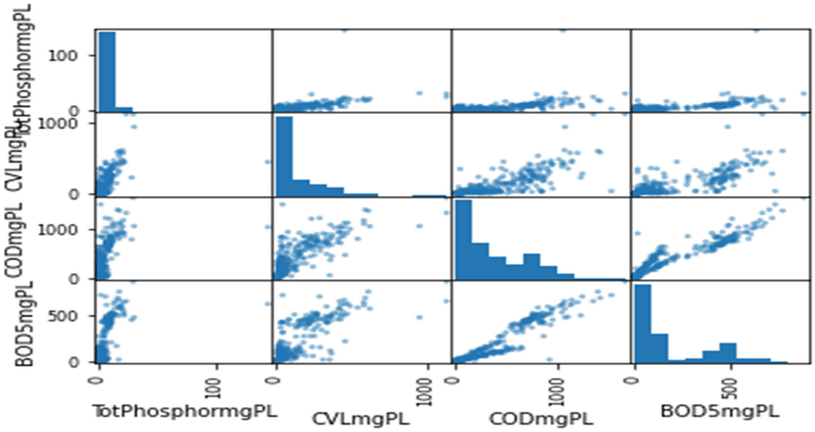


Fig. 5. Correlation of chemical measurement values with each other.

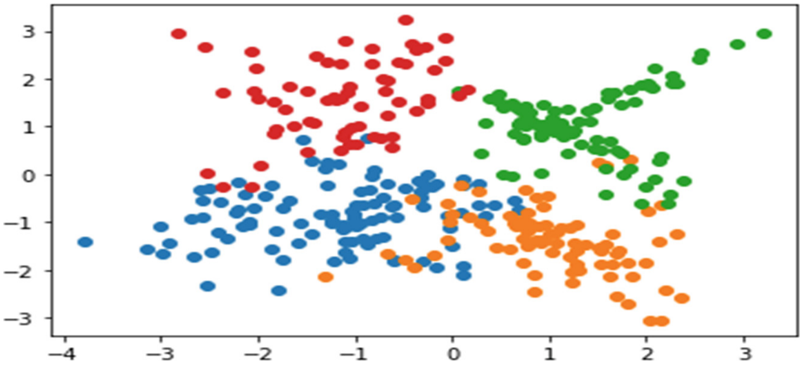


Fig. 6. BOD5 sample sets created according to their values.

When the results obtained from the data set showing the analysis results of 4 parameters related to 334 domestic qualified wastewaters evaluated by the K-Means Clustering Method, which was taken from a previous study in this study and written in Python programming language, were examined; the distribution of BOD_5 (Biological Oxygen Demand) value of 334 samples was found to be lower than 250 by 72%. The proportion of those with a BOD_5 value between 251–500 is 18%, while the proportion of those between Dec 501–750 is 6%. The percentage of those with a BOD_5 value between Dec 501–500 is 6%. When Fig. 5 is examined, it is understood that the variable that affects the BOD_5 value the most is COD (Chemical Oxygen Demand). The results of this study are published in Güller, compared to those of (2019) [10], it is clear that it is very close, as expected. In the future, in the clustering study conducted with this selected data group, ANN will be used instead of KNN and the results will be compared.

Appendix: Python Commands for Clustering and Performance Evaluation

```
#Environment_Clustering_00_K_Means, py
#Created on 2023

"""
print('# Environment_Clustering_00_K_Means')
print('# 2023- SEPTEMBER')
print('# MSKU FACULTY OF SCIENCE')
print('-------------------------')

# Data: EnvirodataI.txt
#
print('#################################')
qualifications = ["EnvirodataI_label","EnvirodataI_name","pH","TemperC","TotPhosphormgPL",
          "CVLmgPL","CODmgPL","BOD5mgPL"]
target qualification = ["BOD5mgPL"]
#  Installation of the Necessary Libraries)
import pandas as pd
import numpy as np
envirodata=pd.read_table('envirodataI.txt')
print(envirodata.head())
print(envirodata.shape)
print(envirodata['EnvirodataI_name'].unique())
print(envirodata.groupby('EnvirodataI_name').size())
import seaborn as sns
sns.countplot(envirodata['EnvirodataI_name'],label="Count")
# Distribution Measures
# -------------
import matplotlib.pyplot as plt
envirodata.drop('EnvirodataI_label', axis=1).plot(kind='box',subplots=True, layout=(4,2),
                                    sharex=False, sharey=False,figsize=(9,9),
                                    title='Box Plot for each input variable')
plt.savefig('envirodata_box')
plt.show()
import pylab as pl
envirodata.drop('EnvirodataI_label' ,axis=1).hist(bins=30, figsize=(9,9))
pl.suptitle("Histogram for each numeric input variable")
plt.savefig('envirodata_hist')
plt.show()
import matplotlib.pyplot as plt
from pandas.plotting import scatter_matrix
 # selecting six numerical features
qualification = ["TotPhosphormgPL","CVLmgPL","CODmgPL","BOD5mgPL"]
# plotting the scatter matrix
# with the features
scatter_matrix(envirodata[nitelik])
plt.show()
################################
# Statistical Study
# -----------------
qualification_ cluster = ["TotPhosphormgPL","CVLmgPL","CODmgPL"]
import pandas as pd
import numpy as np
envirodata_Clustering=pd.read_table('EnvirodataI_Clustering.txt')
print(envirodata_Clustering)
# Models
# ------------
# K Means
# ------------------
# k-means clustering
from numpy import unique
from numpy import where
from sklearn.datasets import make_classification
from sklearn.cluster import KMeans
from matplotlib import pyplot
# define dataset
X, _ = make_classification(n_samples=334, n_features=3,
                        n_informative=3, n_redundant=0,
                        n_clusters_per_class=4, random_state=4)
# print(X, _)
# define the model
model = KMeans(n_clusters=4)
# fit the model
model.fit(X)
# assign a cluster to each example
```

```
yhat = model.predict(X)
# retrieve unique clusters
clusters = unique(yhat)
# create scatter plot for samples from each cluster
for cluster in clusters:
    # get row indexes for samples with this cluster
    row_ix = where(yhat == cluster)
    # create scatter of these samples
    pyplot.scatter(X[row_ix, 0], X[row_ix, 1])
# show the plot
pyplot.show()
# Performance For K_Means
# ----------------------------------
```

Program Outputs:

```
    Python 3.9.12 (main, Apr  4, 2022, 05:22:27) [MSC v.1916 64 bit (AMD64)]
    Type "copyright", "credits" or "license" for more information.
    IPython 8.2.0 -- An enhanced Interactive Python.
    runfile('C:/Users/er-
tur/OneDrive/Masaüstü/MED_BOD5_PYTHON/Environment_Clustering_00_K_Means_med.
py', wdir='C:/Users/ertur/OneDrive/Masaüstü/MED_BOD5_PYTHON')
    # Environment_Clustering_00_K_Means
    # 2023-SEPTEMBER
    # MSKU FACULTY OF SCIENCE
    -------------------------
    ################################
      EnvirodataI_label EnvirodataI_name  ... CODmgPL  BOD5mgPL
    0              1     BOD_000_250  ...  449.50     107.6
    1              1     BOD_000_250  ...  393.38     100.0
    2              1     BOD_000_250  ...  371.90     108.0
    3              1     BOD_000_250  ...  560.41     155.0
    4              1     BOD_000_250  ...  350.00     165.0

    [5 rows x 7 columns]

    (334, 7)
    ['BOD_000_250' 'BOD_250_500' 'BOD_500_750' 'BOD_750_999']
    EnvirodataI_name
    BOD_000_250    233
    BOD_250_500     66
    BOD_500_750     33
    BOD_750_999      2
    dtype: int64
```

```
C:\Users\ertur\anaconda3\lib\site-packages\seaborn\_decorators.py:36: Future Warning: Pass
the following variable as a keyword arg: x. From version 0.12, the only valid positional argument
will be `data` and passing other arguments without an explicit keyword will result in an error or
misinterpretation.
```

```
        1   32.8   449.5
0    2.00  137.0  393.38
1    1.00   90.6  371.90
2    0.22  290.0  560.41
3    6.59   33.4  350.00
4    1.39  123.0  578.23
..    ...    ...     ...
328 10.30  120.0  792.00
329 11.10  273.0  803.00
330 14.20  420.0  951.00
331 18.10  610.0 1386.00
332 30.00 1130.0 1650.00
[333 rows x 3 columns]
```

References

1. Silahtaroğlu, G.: Veri Madenciliği Kavram ve Algoritmaları, Papatya Yayıncılık (2016)
2. Dinçer, E.: Veri Madenciliğinde K-Means Algoritması ve Tıp Alanında Uygulanması. Yüksek Lisans Tezi, s.101, Kocaeli Üniversitesi, Fen Bilimleri Enstitüsü, Kocaeli (2006)
3. Arslan, H.: Sakarya Üniversitesi Web Sitesi Erişim Kayıtlarının Web Madenciliği İle Analizi. Yüksek Lisans Tezi, s.80, Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Sakarya (2008)
4. Han, J., Kamber, M.: Data Mining Concepts and Techniques. Morgan Kaufman Publishers, Academic Press, USA (2001)
5. Evans, S., Lloyd, J., Stoddard, G., Nekeber, J., Samone, M.: Risk factors for adverse drug events. Ann. Pharmacother. **39**, 1161–1168 (2005)
6. Dinçer, E.: Veri Madenciliğinde K-Means Algoritması ve Tıp Alanında Uygulanması. Yüksek Lisans Tezi, Kocaeli Üniversitesi, Fen Bilimleri Enstitüsü, Kocaeli, s.101 (2006)
7. Amasyalı, F.M., Ersoy, O.: Kümeleyici topluluklarının başarısını etkileyen faktörler. In: IEEE 16th Signal Processing and Communication Applications Conference, SIU, Aydın (2008)
8. Gersho, A., Gray, R.M.: Vector Quantization and Signal Compression, p. 738. Kluwer Academic Publishers Norwell, USA (1991)
9. Linde, Y., Buzo, A.: An algorithm for vector quantizer design. IEEE Trans. Commun. 702–710 (1980)
10. Güller, S., Silahtaroğlu, G., Akpolat, O.: Analysis wastewater characteristics via data mining: a Muğla province case and external validation. Commun. Stat. Case Stud. Data Anal. Appl. **5**(3), 200–213 (2019)

Novel Quantum Key Distribution Method Based on Blockchain Technology

Faruk Takaoğlu(✉), Mustafa Takaoğlu, Taner Dursun, and Tolga Bağcı

TÜBİTAK-BİLGEM, Kocaeli, Turkey
faruk.takaoglu@tubitak.gov.tr

Abstract. Quantum key distribution (QKD) methods, one of the working subjects in the field of quantum technologies, are among the solutions that have great potential and will be used both today and in the post-quantum era. Because it is one of the unique solutions that come to the fore in both true random number generation and key distribution in a secure manner. In addition, quantum attack methods are also being studied against the security of quantum key distribution techniques. There are different suggestions against man-in-the-middle attacks, which is one of the important attack methods. In this study, a protocol consisting of two stages and foreseeing the use of blockchain technology without affecting quantum mechanics is proposed in order to be a solution against attacks such as man-in-the-middle on QKD systems. In the first phase of the protocol, a basic identification capability between quantum devices (parties) is gained using the Hyperledger Fabric permissioned blockchain protocol. In the second stage, the standard QKD reconciliation process for BB84 will be performed. In the study, blockchain technology has been added to the quantum key distribution process in the most optimal way, and both the identification of the parties has been brought to the quantum systems and a viable solution has been proposed against man-in-the-middle attacks.

Keywords: Quantum Key Distribution · BB84 · Blockchain Technology · Hyperledger Fabric · Smart Contacts

1 Introduction

In today's digitized world, all online transactions performed by users involve layered security operations and procedures that are abstracted from users in the background. Data is encrypted to enhance data security by preventing unauthorized or malicious users, or in other words, attackers, from accessing it. These security operations and procedures primarily involve the use of cryptographic algorithms to ensure that data cannot be obtained by attackers. They include different key generation and key usage scenarios. The algorithmic complexity and potential for providing high-level security of a cryptographic algorithm are directly related to the secure sharing of the key used in encryption. In other words, the security of a cryptographic algorithm depends on the security of key distribution. Additionally, the advancements in quantum technologies and the computational power of quantum computing, when compared to the computational power of

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 251–263, 2024.
https://doi.org/10.1007/978-3-031-56728-5_22

today's traditional systems, raise questions about the reliability of our current encryption and key distribution systems. However, the post-quantum era also brings threats to the reliability of existing systems, while offering significant contributions from quantum technologies. Firstly, based on the characteristics derived from the rules of quantum physics, it is known that quantum systems provide complete randomness, which is a crucial aspect of data security. The random number generators used in current encryption algorithms actually contains artificial randomness. This artificial randomness can be influenced by multiple factors and can be increased, but it should never be compared to the proven randomness generated by photons used in quantum systems. Therefore, the randomness generated by quantum photons can increase the security level of existing encryption algorithms.

Another outstanding feature derived from the laws of quantum physics is the uncopiable nature of photons used in quantum communication. It is not possible to read or clone a photon transmitted along a fiber optic cable without any physical intervention, as is the case with current technology. In order to obtain information from photons, they need to be physically manipulated. The photons are directed through beam splitters, polarizing filters, or crystals and processed with sensitive photodiodes to be read. All the routing, beam-splitting, and photon reading-interpretation processes fundamentally affect the physical state of the photon. The photon sent by the sender undergoes routing through directed filters, which are referred to as "deterministic" or "probabilistic," and acquires meaning and becomes a qubit. Deterministic filters can attenuate the photons when used in different directions, while "probabilistic" filters can change the spin direction of the photons or leave them in the same direction with a probability. During this photon communication between the sender and receiver, an eavesdropper who physically intercepts the communication channel and performs photon reading cannot manipulate the current state of the photons or make any operations without knowing the polarization filters used by the parties. In a key distribution system where the communication channel is physically interrupted and photon reading occurs, with a data block that can contain 256 qubits or more, each qubit value obtained from every photon used for communication is predetermined. Even a single photon exchange made by the eavesdropper can easily be detected and exposed between the sender and receiver, revealing the presence of the eavesdropper, as long as the reconciliation process is completed in the key distribution system established between the sender and receiver.

One of the most likely products to be implemented after random number generators in today's quantum technologies is the key distribution algorithms that were first developed in the 1980s. The fundamental purpose of quantum key distribution methods is to allow the sender and receiver to agree on an encryption key without sharing critical information and without compromising the confidentiality of key distribution. The agreed-upon encryption key can then be used for various purposes. Examples of successful algorithms, such as BB84 [1], E91, and B92, achieve key distribution between parties by using a quantum fiber optic cable for photon transmission and a conventional classical communication channel, ensuring the conditions mentioned above are met.

Despite the advantages derived from the rules of quantum physics, quantum key distribution systems, which are still under development, have not yet produced a universally accepted solution against known attack methods adapted to quantum systems, such as “man in the middle,” and against “harvest now attack later” attacks by quantum computers. Currently, proposed solutions only address limited scenarios, such as a single sender, a single receiver, or eavesdropping on the channel. In other cases, security against attacks is believed to be achieved by introducing additional components called quantum nodes, which increase the setup costs of the systems. Other proposed solutions generally suggest using low-coherent photons in combination with different systems, considering that these photons can be attenuated during attack operations, and the probability of attenuation increases as the distance extends. It should be noted that in these proposed systems, data transmission along the channel will be heavily influenced by noise. Consequently, in a nationwide key distribution system involving low-coherent photons, multiple amplification processes will be required, and the need to improve the quantum bit error rate (QBER) will arise.

The fundamental problem can be summarized as the inability of the sender and receiver to identify each other’s identities. To address this issue, the addition of quantum units between the sender and receiver, as well as the direct and indirect quantum and classical communication between these units and the sender and receiver, will increase the complexity and challenges of implementing additional protocols, as well as the design costs. Furthermore, it is necessary to ensure the software and physical security of the electronic, optomechanical, and electro-optical circuits to which the photon generators and detectors, as well as each module added to the quantum network, are connected.

The concept of a decentralized, immutable, and distributed ledger, introduced with the whitepaper of the Bitcoin payment system in 2008, has attracted significant attention from people across various sectors following the financial success of Bitcoin. Subsequently, these technologies, known as blockchain technology and more broadly distributed ledger technology, have been implemented in many fields, particularly in finance. Quantum technology is also one of the areas of exploration within these domains [11]. Upon reviewing the literature, it is evident that many studies have been conducted without the contribution of expert researchers and scientists in the field of quantum, resulting in proposals that are impractical or impossible to implement. Various ideas have been examined, ranging from the use of blockchain protocols in quantum computers to performing quantum key distribution processes on the blockchain. However, it has been observed that many of these ideas overlook fundamental concepts and are infeasible. Therefore, the researchers at the Quantum Technologies Department and Blockchain Technologies Department (BZLAB) within the Informatics and Information Security Research Center (BİLGEM) of The Scientific and Technological Research Council of Turkey (TÜBİTAK), possessing the necessary expertise, have contributed to the preparation of an applicable hybrid study. The aim of this study is to incorporate blockchain technology in a manner that does not compromise the functioning mechanisms of quantum key distribution protocols and to achieve the most optimal contribution offered by the inherent properties of the technology.

After the introduction, the study further elaborates on the literature review, quantum attacks, blockchain technology, and Hyperledger Fabric sections to provide a better presentation. Subsequently, the proposed method, system evaluation, and conclusion sections are included to complete the study.

1.1 Literature Review

In [2], an improvement was attempted for QKD using Turbo Code. Before sending quantum keys, parity bits were added, and adjustments were made with an additional module during the decoding process.

In [3], an interesting study was conducted to transform classical computers into an emulator-like system capable of performing QKD operations through an organization conducted on the blockchain. A hybrid system was developed to test and improve the existing technology, possibly increasing its familiarity.

In [4], it was conveyed that the developed QKD systems and the devices used are not as reliable as the security provided by quantum physics at the core of the system. It was noted that some devices cannot operate independently of the QKD algorithms due to their inherent properties (design, production, noise, etc.). It can be inferred that in a network where different brands and QKG are used, the key distribution in the network would likely involve a high level of redundant bits.

In [5], multiple quantum nodes are placed between Alice and Bob to XOR the obtained quantum values and record them in the blockchain system. Despite the "man in the middle" attack, communication can continue, and key sharing remains unaffected. However, the performance and costs of the repetitive processes in the performed operations and the presence of multiple nodes, as well as the applicability of these systems over long distances in real-life scenarios, need to be considered.

In [6], they utilize a feature called Physical Unclonable Functions (PUFs), which uses random variations in the production of physical devices as a distinguishing characteristic to differentiate each device from others. As a result, the outputs obtained from each device will be entirely different. They aim to perform device identification without using protocols such as identification in quantum memory, using qPUF-based devices and have developed a protocol suitable for this purpose to reduce attacks. The basic structure is the device identification protocol scheme.

In [7], this article explains how the secure and reliable operation of QKD is ensured within the network layer structure established based on SDN (Software-Defined Networking).

In [8], an attack was conducted on an example QKD network to obtain information from the network by targeting the avalanche photodiode data reading and operational time intervals. Similar to other studies, the aim may have been to obtain data from the network without being noticed. However, in the study conducted by the authors, an attack method was developed to disable the QKD system.

In [9], using the BB84 quantum key distribution algorithm, a systemic vulnerability was created by manipulating the recharge times of Avalanche Photodiodes used for photon reading in the network. The relevant article made suggestions on how to take measures against such attacks that exploit this type of system vulnerability.

In [10], examples and information were provided about the fake state attack, which is one of the quantum attack methods, and quantum cryptography.

2 Quantum Attacks

Before discussing attacks on quantum key distribution (QKD) systems, let's briefly summarize the BB84 quantum key distribution system as an example. The shared BB84 algorithm, depicted in Fig. 1, was developed in 1984 by Charles H. Bennett and Gilles Brassard. Alice serves as the sender, and Bob as the receiver, located at the two ends of the communication channel. Both parties are aware that they will communicate using the BB84 method and will follow its steps. To establish communication between the parties, both a quantum fiber optic cable and a classical communication channel are used. Initially, Alice selects a series of quantum polarization filters to send to Bob and sequentially passes these photons through the polarization filters to transmit them through the fiber optic cable. Since Bob doesn't know which polarization filters Alice has chosen, he randomly selects " 45°", "135°", "90°", and "0°" polarization filters to pass the incoming photons through. Based on this reading process, the directions of the obtained photons are determined, and they are interpreted as "+", "–", "1", and "0" values based on their obtained directions. Bob communicates the obtained values to Alice using the classical channel. Alice evaluates Bob's results and identifies the matching results to find the corresponding selected polarization filters. Alice then communicates the index/sequence numbers of the matching polarization filters to Bob using the classical channel. Through this process, both parties determine which polarization filters they will use for data transmission, thereby establishing the key. Now, let's consider a scenario where there is an unauthorized party without access to the channel. This process is known as eavesdropping, and the person performing this process is commonly referred to as Eve. Eve knows that according to the laws of quantum physics, photons cannot be cloned or copied. Therefore, Eve can physically integrate into the fiber optic cable and start reading the photons using optical components. For this process, she can read the photons by sequentially selecting random polarization filters. Since Eve doesn't know the polarization filters chosen by Alice and Bob, if she mistakenly selects a polarization filter in a different order, she can distort, change, or even attenuate the direction of the photon. In this case, when Alice and Bob obtain conflicting data while using the agreed-upon filters, they will realize the presence of an eavesdropper/listener.

Like any other systems, QKD systems can have vulnerabilities and exploitable areas. Quantum attackers target these points to extract data from the network and disrupt its operation. Examples of these attacks include the "photon splitting" attack, "man-in-the-middle" attack, "fake state" attack, and "time shifting" attack.

As examples of attacks aimed at extracting information from the system, the "man-in-the-middle" attack and "fake state" attack methods can be mentioned. In the man-in-the-middle attack method, the attacker, known as Eve, can enter the network using a beam splitter and similar optical components. It has been reported [8] that current QKD test operations are conducted over telecommunication fiber optic cables with a wavelength range of 1550 nm or 1310 nm. In such a situation, even before the reconciliation process takes place between the parties in the communication channel, the attacker named Eve can position herself on the channel, split the photons with a beam splitter, and read the photons from both sides to obtain information about the key.

With the fake state attack method, Eve can both extract data from the communication channel and sabotage data communication. In pursuit of this goal, Eve will read the

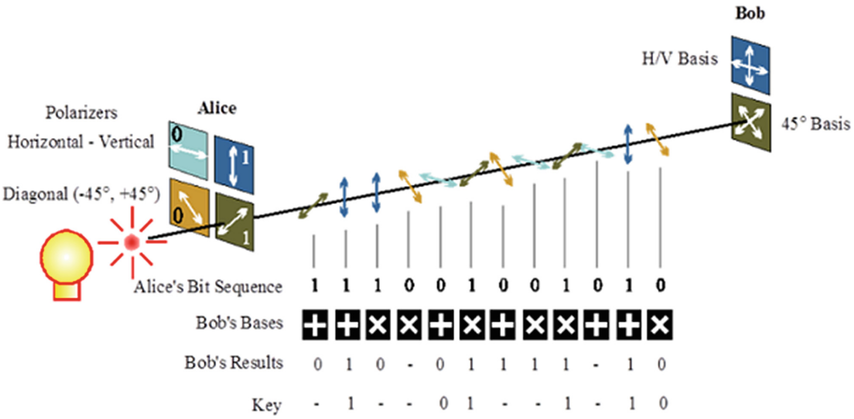

Bob
H/V Basis
45° Basis
Polarizers
Alice
Horizontal - Vertical
0
1
Diagonal (-45°, +45°)
0
1
Alice's Bit Sequence 1 1 1 0 0 1 0 0 1 0 1 0
Bob's Bases + + × × + × + × × + + ×
Bob's Results 0 1 0 - 0 1 1 1 1 - 1 0
Key - 1 - - 0 1 - - 1 - 1 0

Fig. 1. BB84 Quantum Key Distribution Algorithm [1].

photons sent by Alice and transmit the data to Bob in the opposite direction of the obtained values. The objective is to send fake states to disrupt Bob's connection with Alice and ensure the matching of intentionally corrupted bits with Bob's bits. While it is known that having a high number of errors in the transmitted bits between Alice and Bob would be problematic, as stated in [10], real-world test operations have reported that many photons are not highly readable, and this situation can be attributed to system glitches and synchronization problems between sensors.

To completely sabotage data transmission, methods like time-shifting attacks can be employed. In this attack method, the attacker infiltrates the communication lines between the parties in the communication channel without considering the meaning and interpretation of the obtained photons and exploits the synchronization vulnerability of avalanche photodiodes (APDs). APDs used for sensitive photon reading in the communication channel are used to read data at certain time intervals and are rested for recharge after the data reading process within a specific time interval. If APDs do not operate and recharge at the same time intervals, proper data reading cannot be performed mutually. If APDs are not properly maintained according to the ideal working and recharge times, they will lose their operational sensitivity and will be no different from normal photodiodes [8].

3 Blockchain Technology

Blockchain technology is a decentralized, distributed, and tamper-resistant digital ledger protected by cryptographic algorithms such as hashing and digital signatures. With the introduction of smart contract development capabilities through the Ethereum protocol, blockchain technology has made it possible to be utilized in almost every field, unlike the single-purpose solution proposed in the Bitcoin payment system. Thus, the unified blockchain technology introduced in 2008 with the Bitcoin whitepaper has evolved over time into the concept of the world state machine, or in other words, blockchain virtual machine solutions. Nowadays, many Layer 0 and Layer 1 protocols have the ability to develop smart contracts, enabling the application of blockchain technology, or more broadly, distributed ledger technology, in almost all areas, both in terms of scalability and smart contract upgradeability [12].

The capability of developing smart contracts is a significant development. Smart contracts, in simple terms, are pieces of code that are pre-programmed for a predetermined condition and deterministically produce a result when that condition is met. Different blockchain protocols give different names to smart contracts. For example, they are called programs in Solana and chaincode in Hyperledger Fabric. However, they are all code blocks running on their respective blockchain protocols. Through smart contracts, decentralized applications (DApps) can be developed, and smart contracts can trigger other smart contracts. This enables the creation of decentralized autonomous organizations (DAOs). The opportunities provided by smart contract capabilities open the way for many companies, public institutions, or software projects to be implemented on a blockchain-based platform [13].

Blockchain solutions can be entirely public, or they can be permissioned or private, depending on the access rights, transaction capabilities, or transaction history visibility requirements of the targeted user base in the projects. Nowadays, it can be observed that many private sector organizations or public institutions prefer permissioned or private structures in their ongoing or completed projects. Central Bank Digital Currency projects are a good example in this context [11].

3.1 Hyperledger Fabric

Hyperledger Fabric (HLF), a product of the Hyperledger Foundation family, is a permissioned blockchain protocol and a widely preferred Layer-1 solution in the digital transformation of businesses. HLF has the potential for smart contract development. The concept of smart contracts in the protocol is referred to as chaincode. HLF is an open-source platform with high-quality documentation and a community-driven structure. Unlike counterparts such as Bitcoin, Ethereum, or AVA, HLF does not have a native coin designed within the protocol. HLF has a multi-version concurrency control (MVCC) system that ensures ledger consistency and makes it difficult for Distributed Denial-of-Service (DDoS) attacks to be successful. Therefore, a native coin is not deemed necessary in HLF. However, HLF is a protocol that allows account-based and UTXO-based projects to be developed. HLF is a modular protocol and can be considered successful in terms of interoperability with other blockchain platforms through oracles. The protocol provides system monitoring and analysis tools. The biggest challenge faced by the deterministic and fault-tolerant HLF protocol is the issue of speed. While HLF is widely preferred in small and medium-scale projects, it may encounter scalability issues in projects generating large amounts of data and involving numerous transactions. This can be attributed to factors such as the endorsement policy used, the CouchDB or LevelDB databases that can be selected in the network, and the architecture of the system designed according to specific requirements (e.g., the use of Decentralized Identifiers (DIDs) based on requirements, etc.). Especially in projects with complex requirements, high performance (TPS) should not be expected from the HLF protocol. However, it can be highly beneficial in academic research or Proof-of-Concept focused investigations where speed is not of utmost importance [14].

In this context, the proposed protocol in the study utilizes the HLF protocol. Through the developed chaincodes in the HLF protocol, the identification and definition of the sender and receiver parties in the quantum key distribution process are ensured. The chaincode facilitates the random generation of 128-bit prefixes for the parties, which will be used for photon generation and for identification purposes, at specified intervals. Additionally, the values representing the working and recharging times of Avalanche Photodiodes (APD), which enhance security against time-shifting attacks, are stored in the blockchain as variables.

4 Proposed System

In the proposed system, the initial phase involves obtaining a permissioned ledger system using the Hyperledger Fabric protocol. The parties involved in the quantum key distribution within the system are defined as users, and they are granted access permissions to the HLF. At this stage, devices that will interact with HLF are assigned a Decentralized Identifier (DID), which is a blockchain-based digital identity. This enhances the security of device access to the system and strengthens overall system security. BİLGEM's DID Proof of Concept (PoC) solution is utilized as the DID solution. The developed chaincode generates random prefixes for each user at intervals determined and modifiable based on the system's needs, and associates them with the users. These prefixes involve the usage of polarization filters in different sequences, which are utilized in the conversion of quantum photons to qubits, and the order of usage is randomly generated. Furthermore, working and recharge time values of Avalanche Photodiodes (APD) are recorded in the ledger. If there is a desire to initiate a quantum key distribution process between parties, another chaincode comes into play, transmitting prefix values and APD values of the sender and recipient. The proposed method's initial phase concludes with the establishment of the necessary blockchain infrastructure, user identification, and definition of operations that will be performed at the chaincode level when interacting with HLF.

In the second phase of the proposed method, the sender initiates the BB84 key distribution process using their own prefix. The recipient waits for a photon to arrive, which starts with the prefix value shared with them through HLF prior to the operation. It is expected that the first 128 bits of the incoming photon will perfectly match the pre-shared prefix between the parties. During this process, if a faded photon or a photon with a different polarization than expected is received, it indicates the presence of an eavesdropper. In the case of a mismatch, the parties will communicate over a classical channel and repeat the procedures mentioned in the first phase by changing the fiber channel. This process effectively eliminates attacks such as man-in-the-middle and fake states.

If the attacker intends to exploit the vulnerabilities of the existing system and sabotage the network rather than leaking data, they can ensure data transmission with opposite values of the APD (Avalanche Photodiode) values on the recipient side instead of reading the photons. In this case, the time intervals set for the proper functioning of the APDs on the recipient side are disrupted, and the dark count values obtained on the recipient side increase. In the proposed system, when the dark count values observed on the recipient side exceed the ideal values, users can communicate over a classical channel and indicate the presence of an eavesdropper, thereby changing the fiber channel and repeating the procedure mentioned in the initial phase.

If all these stages are successfully completed, the steps of the well-known BB84 algorithm are executed, enabling secure key sharing. Figure 2 illustrates the overall representation of the proposed system.

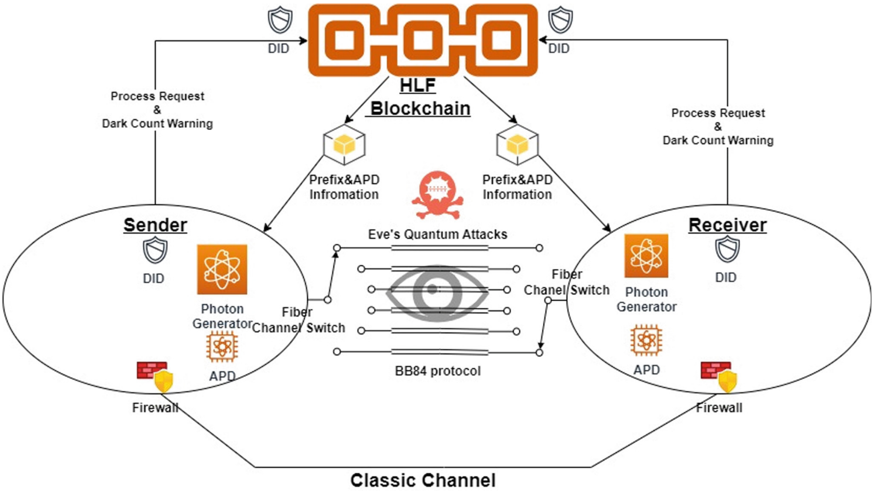


Fig. 2. Proposed blockchain-assisted QKD method.

The provided visual depicts the fundamental operations in a scenario involving two communication channels. In a system that enables communication across the entire country and between provinces, it is not feasible for a photon emitter to send photons of a certain speed and quality to all distances using current devices. The efficient communication distances vary depending on the type of fibers used in the communication infrastructure and the size of the data to be transmitted. In experiments, a data block of 1.26 megabits per second has been transported over a distance of 50 km using a standard optical fiber communication infrastructure. By utilizing ultra-low-loss fiber cables, data transmission speeds of 1.16 bits per hour have been achieved over a distance of 404 km. [15] Therefore, in a quantum network established within a country, due to the long distances between parties, the message needs to be transferred between the sender and recipient through one or several quantum nodes. In the Demo QuanDT project [16], planned to be completed in Germany in 2024, a budget of €15.2 million has been allocated for quantum key distribution between Berlin and Bonn using 18 trusted quantum nodes. It should be noted that the implementation of the proposed systems mentioned in previous sections, even in experimental projects, may lead to significant cost increases by adding additional quantum nodes to address the current problem. The main objective is to enable "any-to-any" communication with N devices. It is necessary to establish a quantum communication network among all quantum nodes in different locations to provide a secure and reliable communication infrastructure against attackers. Through the proposed blockchain-integrated system, not only can reliability be ensured against known attacks, but also the secure addition of new nodes to the network and their integration into the blockchain system can be easily achieved.

As depicted in Fig. 3 below, in the proposed system, prefix values can be generated and managed in a multi-instance manner by the blockchain system. These prefix values can be transmitted to all other nodes responsible for reading and transmitting the signal between the sender and the recipient, with different values due to the distance limit between them. This enables the tracking of any mismatches or deviations in prefix or dark count rates among nodes when transmission is conducted over long distances. This feature provides insight into which local area Eve is attempting an attack.

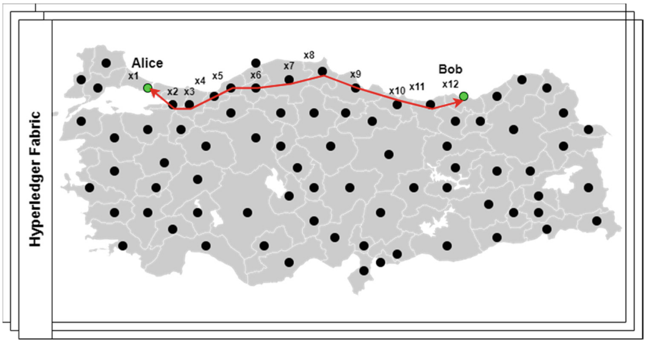


Fig. 3. A sample representation of the proposed method at the country level.

5 System Evaluation

Low-coherent photons can be used as a preventative measure against the vulnerability of fiber lines to splitting or any other assaults in QKD systems. Low-coherent photons are easily attenuated against such assaults and reduce the receiver's dark count performance. These low-coherent photons, however, cannot be employed successfully over long distances and are sensitive to noise within the channel. High-coherent photons, on the other hand, transported across greater distances while being less susceptible to noise. As a result, periodic photon regeneration/amplification must be used in systems employing low-coherent photons to ensure appropriate transmission. These solutions to the current scenario result in large cost increases when installing a QKD system.

Another issue that QKD systems face is the exposure to attacks like fake-state and man-in-the-middle. The primary objective in such cases is for Eve, playing the role of an eavesdropper, to leak data over the relevant channel and deduce information about the parties' key sharing. Additionally, Eve may sabotage the data transfer. Proposed solutions to prevent man-in-the-middle attacks include sharing entangled photons, known as Quantum Registers, from a trusted third-party source, integrating error correction and code systems into the quantum system, and typically utilizing classical channels for information exchange. The deployment of each additional quantum module in the system would require environmental and fiber optic infrastructure, security measures for established systems, preview processes, and performance measurements.

Time-shifting attacks utilized in QKD systems can exploit the inherent limitations of avalanche photodiodes (APDs) involved in key distribution, thus allowing the sabotage and exploitation of the entire network, as mentioned in previous sections.

The fundamental cause of all the aforementioned problems lies in the lack of a pre-QKD identification process to counteract sabotage and information leakage between the parties. Although low-coherent photon sources appear to be successful solutions, it is anticipated that implementing this solution on a nationwide scale would significantly increase system expenses. To address this issue, a blockchain-based identification system has been proposed to ensure pre-QKD identification between the parties and enhance attacker detection. A permissioned blockchain system created on Hyperledger Fabric, which all quantum nodes would be connected to, could be implemented nationwide. The structure of this blockchain system would inherently prevent unauthorized access. In the communication process, the parties would retrieve specific-length random polarization filter combinations from the blockchain system to identify each other. With knowledge of these prefix polarization values, both sender and receiver would establish an identification process.

Furthermore, the blockchain system would share APD recharge and working time durations, which can be adjusted based on the dark count value, to prevent time-shifting attacks. In this scenario, Eve, not being part of the permissioned blockchain system, would be unable to obtain the prefix values required for data leakage. Additionally, when attempting to exploit or sabotage the channel, if the dark count rate between the parties falls below the desired level, they would switch to a new channel and, if needed, acquire shared parameters for APD synchronization. These enhancements would remove obstacles to using high-coherent photon generators, eliminating the need for additional elements, quantum nodes, and cost-increasing amplification circuits in the system.

With the proposed innovative blockchain-supported method, it is possible to obtain a feasible system that can provide solutions to all the mentioned problems. The blockchain solution does not affect the quantum mechanism in the quantum key distribution process. There is no usage scenario against the principles of quantum physics. The identification capabilities of the parties are included in the system with the blockchain capability before the photon generation. Furthermore, the use of Hyperledger Fabric (HLF) as the choice of blockchain protocol provides permissioned access, allowing only approved parties to access the system and inherently reducing the risks of insider attacks. In addition, transaction histories in HLF can also be hidden between authorized users using cryptographic algorithms in case it is desired to abstract and limit data access. Another advantage of using a blockchain platform is to reduce the risk of a single point of failure. Although the level of decentralization is not as high as that of public protocols, blockchain structures are more resilient to a single point of failure than centralized solutions. By randomly generating prefix values at regular intervals in HLF, the system reduces the risk of successful eavesdroppers exploiting an overlooked vulnerability. If an eavesdropper successfully exploits such a vulnerability, the system will switch to a new prefix before any recognizable pattern is detected, thus reducing the risk of a successful attack.

Known attack types and security measures taken in similar studies are explained above. The system proposed in our study and the security elements it provides do not cause any change in the evaluation procedures of the BB84 algorithm. The most fundamental element of the security provided by the system is to ensure that the "coherent" position of the generated photons is not disturbed. In order for the qubits to be coherent, they must be transmitted at ideal distances, because the transmission distance of the qubits is not infinite. There will be a certain and measurable amount of noise in the system and this is expected. It is important that the unexpected noise in the system is detectable and understandable. For this purpose, a security system that can reveal all kinds of attacks in the current situation should be integrated into quantum key distribution systems.

Our system conducts a testbed to reveal the presence of eavesdropping and to evaluate the protocol's ability to detect and counter eavesdropping attempts, thus helping the system validate the protocol's security claims.

6 Conclusion

In the proposed system design, unlike other systems, the emphasis is on feasibility and the ability to generate hybrid solutions using existing and emerging technologies. The goal is to achieve a reliable solution against current attack methods while keeping the cost of system implementation at a minimum through proposed solution recommendations. The sharing of known prefixes in the proposed system architecture can provide a solution to the authentication issue between parties in all n-to-n and side-to-side communication types. This allows for easy detection of attempts to leak data from quantum key distribution systems. The visibility of APD recharge and working time values through the blockchain enables the use of these values for system improvements and performance enhancements. The adjustability and determinability of APD values contribute to the system's reliability against time-shifting attacks.

From a design perspective, ensuring security and reliability through the blockchain allows for easy integration of new nodes into the system and enhances manageability. The generation and transfer of prefix values to devices by the blockchain system, as well as the ability for authorized users to examine past records, can be an ideal input for forensic analysis. In long-distance transmission, the ability to synchronize with multiple devices and assign different prefix values between all device pairs ensures security during operation.

References

1. Bennett, C.H., Brassard, G., Mermin, N.D.: Quantum cryptography without Bell's theorem. Phys. Rev. Lett. **68**(5), 557–559 (1992)
2. Vatta, F., Romano, R., Alizo, M.T.D.: Turbo codes for quantum key distribution (QKD) applications. In: Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL '11), pp. 162–166 (2011)
3. Krishnaswamy, D.: Quantum blockchain networks. In: Proceedings of the Twenty-First International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing (Mobihoc '20), pp. 327–332 (2020)

4. Vazirani, U.V., Vidick, T.: Fully device-independent quantum key distribution. Phys. Rev. Lett. **113**, 140501 (2014)
5. Chen, H.: Quantum relay blockchain and its applications in key service. In: Proceedings of the 2020 4th International Conference on Cryptography, Security and Privacy (ICCSP 2020), pp. 95–99 (2020)
6. Doosti, M., Kumar, N., Delavar, M., Kashefi, E.: Client-server identification protocols with quantum PUF. ACM Trans. Quantum Comput. **2**(3), Article 12 (2021)
7. Wang, H., Zhao, Y., Nag, A.: Quantum-key-distribution (QKD) networks enabled by software-defined networks (SDN). Appl. Sci. **9**(10), 2081 (2019)
8. Qi, B., Fung, C.F., Lo, H., Ma, X.: Time-shift attack in practical quantum cryptosystems. Quantum Inf. Comput. **7**, 73–82 (2005)
9. Makarov, V., Anisimov, A., Skaar, J.: Effects of detector efficiency mismatch on security of quantum cryptosystems. Phys. Rev. A **74**, 022313 (2006)
10. Denny, T.: Faked states attack and quantum cryptography protocols. ArXiv preprint arXiv:1112.2230 (2011)
11. Takaoğlu, M., Özer, Ç., Parlak, E.: Blockchain technology and possible application areas in Turkey. Int. J. Eastern Anatolia Sci. Eng. Des. **1**(2), 260–295 (2019)
12. Marengo, A., Pagano, A.: Investigating the factors influencing the adoption of blockchain technology across different countries and industries: a systematic literature review. Electronics **12**(14), 3006 (2023)
13. Park, J., Jeong, S., Yeom, K.: Smart contract broker: improving smart contract reusability in a blockchain environment. Sensors **3**(13), 6149 (2023)
14. Clarke, R., McGuire, L., Baza, M., Rasheed, A., Alsabaan, M.: Online voting scheme using IBM cloud-based hyperledger fabric with privacy-preservation. Appl. Sci. **13**(13), 7905 (2023)
15. Lucamarini, M., Yuan, Z.L., Dynes, J.F., et al.: Overcoming the rate–distance limit of quantum key distribution without quantum repeaters. Nature **557**, 400–403 (2018)
16. DemoQuanDT Homepage. https://www.forschung-it-sicherheit-kommunikationssysteme.de/projekte/demoquandt. Accessed 10 July 2023

Smart Parking System Based on Dynamic and Optimal Resource Allocation

Khadidja Tair(✉), Lylia Benmessaoud, and Saida Boukhedouma

LSI Laboratory, Faculty of Computer Science, USTHB, Algiers, Algeria
{ktair,lbenmessaoud1,sboukhedouma}@usthb.dz

Abstract. Smart parking uses the Internet of Things (IoT) to collect and control data to efficiently manage parking spaces. Indeed, Smart Parking Systems (SPS) in smart cities can reduce traffic congestion, reduce the time for searching available parking spots, reduce the CO2 emissions, and offer other benefits, but their efficient resource (parking spot) allocation remains a challenge. In this article, we present our allocation approach using genetic algorithms (GA) with a multi-criteria objective function. We managed to optimize the four criteria of distance, parking space utilization, user preferences and travel time, while considering a set of parkings in a smart city. As a conceptual support of our approach, we propose a meta-model of concepts for smart parking management in IoT context; we also propose a four-layer architecture of SPS in a smart city. Our goal is to maximize parking space utilization, minimize travel time and distance, and satisfy driver preferences.

We conducted a set of simulations to highlight significant improvements in the overall optimization of parking spots allocation. The results showed that the GA algorithm satisfied driver's preferences while minimizing distance and travel time to get a parking spot and also, maximizing parking spots utilization. Furthermore, comparisons of the results obtained with those derived from the FCFS algorithm showed the effectiveness of the GA algorithm.

Keywords: smart city · smart parking · dynamic resource allocation · IoT · optimization · genetic algorithm · multi-objective function

1 Introduction

Smart parking is an essential component in smart city. It refers to the use of advanced technologies, such as sensors, cameras, actuators and data analytics methods, to optimize parking management, reduce congestion, and improve the overall driving experience. This can be achieved through an effective resource management system.

Smart parking involves the use of virtual resources in the form of protocols enabling communication and data exchange between system components, and physical resources such as networking resources, parking spaces, sensors, and actuators. The parking spaces are the central resources to manage, sensors and actuators monitor their availability.

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 264–278, 2024.
https://doi.org/10.1007/978-3-031-56728-5_23

Resource allocation is essential in IoT applications and in smart parking particularly to optimize the use of available spaces (parking spots), improving traffic flow and reducing waiting times. Automated systems detect spots occupancy and provide real-time information, enhancing the quality of life for residents and visitors. However, densely populated cities with high traffic face challenges like congestion, pollution, accidents, and wasted time and fuel due to parking difficulties.

Several studies have explored various approaches to tackle the challenges associated with efficient resource management in this domain. After exhaustive exploration of the literature, we concluded that notable works focused on four main categories of resource allocation aiming to optimize one or more criteria: travel time optimization [7, 22], parking space optimization [6, 9], distance optimization [8] and user preferences-oriented [17], In our approach, we aim to simultaneously optimize all four of these criteria.

The main objective of this article is to focus on optimization of resource allocation in smart parking systems. Specifically, after a clear review of the literature and comparison of approaches, we propose an approach aiming to enhance the overall user experience by optimizing the travel time, distance traveled, and parking space utilization while taking into account user preferences. To achieve this, by employing genetic algorithm, we define a multi-objective fitness function based on four criteria: distance, travel time, vehicle size for parking space utilization, and user preferences. After several experiments, we selected the best combination of functions (selection, mutation, and crossover) that provided the highest fitness level.

In our study, we simulated sensors' data from several smart parkings in a smart city. Using these simulated data, we conducted extensive tests and evaluations on multiple scenarios. The results consistently demonstrated that the Genetic Algorithm (GA) successfully converged towards the optimal solution. We also compared the algorithm with the First-Come, First-Served (FCFS) approach and found that the GA algorithm yields better results than FCFS. The GA algorithm prioritizes efficiency and maximizes the satisfaction of drivers by considering distance, travel time, vehicle size, and user preferences in the parking assignment process. Let's notice that the FCFS can be simply used in case of sequential users' requests (smooth traffic), where each request is served in the order it arrives. And the Genetic Algorithm (GA) is utilized in scenarios where there are simultaneous requests (heavy traffic).

The rest of the paper is structured as follows: Sect. 2 defines basic concepts attached to IoT in smart parking and resource allocation challenges in smart parking systems (SPS). Section 3 presents existing works around resource allocation optimization in SPS according to the four categories identified. Section 4 provides a comparison and a discussion around the approaches presented. Section 5 presents conceptual and implementation elements of our solution. Section 6 presents the evaluation of our solution and experimental results. Section 7 concludes the paper and evokes starting points for future works.

2 Smart Parking in IoT and Challenges

Smart parking solutions use new technologies, including IoT devices such as smart cameras and sensors, to collect and transmit data to a mobile application, optimizing parking utilization, improving urban mobility, reducing congestion, and enhancing quality of life [1]. The Internet of Things (IoT) enables the creation of Smart Parking Systems (SPS) by facilitating device communication through unique identifiers (UIDs), allowing wireless or wired data transfer, leading to increased efficiency [4].

Smart cameras detect license plates of vehicles in car parks to provide information on parking availability, waiting times, and prices; they can also authenticate users for reserved parking spaces. Similarly, sensors detect vehicle presence and provide real-time information on space availability, optimizing parking space use and improving efficiency, resulting in easier access for users and shorter waiting times [4].

According to [1–3], types of sensors used in smart parking systems are ultrasonic sensors, infrared sensors, microwave radar sensors, and magnetic sensors. Actuators can be of two types: electromechanical or hydraulic.

IoT resources are of two categories: physical resources and virtual resources. Among these resources, we focus on physical resources, especially parking spaces.

Smart parking using IoT is a potential solution, but it presents challenges like overcrowding, overuse, and underuse of parking areas, leading to high costs, delays, and negative environmental impacts. Effective resource allocation is crucial for addressing these challenges [4, 5].

- Overcrowded parking areas which can cause heavy traffic and stress drivers.
- Overuse of parking areas forcing drivers to look for a parking space unnecessarily.
- Underutilization of parking lots resulting in loss of revenue due to poor management.
- Searching for parking spaces can cause traffic jams and block waiting vehicles.
- Vehicle collisions in parking lots can occur due to simultaneous arrivals or failure to follow traffic rules.
- Drivers waste time searching to find a parking space.
- User preferences such as proximity to transportation or accessible parking for individuals with disabilities.

3 Literature Review on Smart Parking Systems

In our analysis of the existing literature on SPS, we identified four main categories of approaches that focus on optimizing one or more criteria mainly: the travel time to access a parking space, the use of parking spaces, the distance between the driver and the parking space or the driver's preferences.

3.1 Travel Time Optimization

This category includes research works that aim to optimize the travel time between the driver and the parking space, taking into account traffic density. In [7], the authors developed a cloud-based intelligent parking mobile application that uses deep learning and an LSTM-based model. In [13], the authors proposed an algorithm, a mathematical model, and an IoT-based architecture for an intelligent parking system to detect free spots in less time. In [15], a dynamic allocation method for smart parking management was proposed to reduce travel time using an event-based allocation algorithm. The authors of [22] developed a novel mobile deep learning architecture-based approach (MDLpark) to predict parking occupancy. This approach, based on Temporal Convolutional Network (TCN), improves prediction accuracy and reduces driver travel time to access a parking space.

3.2 Parking Space Utilization Optimization

The main objective of this category is to maximize parking space utilization, for example by avoiding to assign a small vehicle to a larger parking space. In [6], the authors developed an automated method for detecting vacant parking spaces using machine-learning algorithms (bag-of-features representation technique and customized background subtraction algorithm). The authors of [9] used genetic algorithms and particle swarm optimization algorithms to optimize parking space utilization of electric vehicles and Demand Response Resources (DRR) in the power distribution network. The authors of [10] proposed the edPAS system for efficient management of dynamic parking allocation requests in vehicular networks, with two parking allocation schemes, FCFS and PR (priority algorithms). By maximizing the use of parking spaces, the travel time is also reduced. In [16], the authors proposed a scheme to select optimal parking sites that reduce CO2 emissions in urban road networks using genetic algorithms. The authors of [18] proposed an intelligent IoT-based service that can forecast parking space occupancy in real-time using a Deep Learning-based ensemble technique; they used genetic algorithms to optimize predictor parameters. In [19], the authors suggested a smart parking architecture and a solution that requires minimal financial investment and can increase capacity by 31.66%. In [21], a system was proposed integrating IoT and ensemble-based regression models to predict parking space availability. The authors used the Bagging method to optimize the prediction in order to reduce urban congestion and pollution. In [23], the authors proposed a framework for smart Parking Data Management and Prediction (SPDMP) system, to address the challenges of long-term parking lot occupancy prediction using a custom PAP (Parking Availability Prediction) neural network.

3.3 Distance Optimization

This category of approaches focus on minimizing the distance between the driver and the assigned parking space. In [8], the authors proposed an algorithm based on Dijkstra to minimize the distance between the driver and the assigned parking space; travel time was also minimized. In [12], the authors proposed a method minimizing distance and travel time by employing Steady State Evolutionary Algorithm (SSEA), Random Search (RS), and Simulated Annealing (SA).

3.4 User Preferences-Oriented

The goal of these approaches is to satisfy drivers' needs and preferences such as the per hour price of a parking space, the nearest parking space, the parking space that is close to the driver's destination. The authors of [17] proposed a Multiple Criteria-based Parking Space Reservation (MCPR) algorithm using multiple criteria decision analysis (MCDA) method that allocates spots based on user's needs, which has led to a reduction in distance and travel time.

4 Comparison of Approaches

Table 1 provides a comparison of the different approaches previously evoked, according to a set of criteria: objective/category, proposed and used methods, data sources (for experimentation) and performance metrics.

Table 1. Comparison of resource allocation approaches.

Reference	Category/objective	Proposed and used methods	Data sources	Performance metrics
(Varghese et al., 2019) [6]	Parking space utilization	Machine learning, Bag-of-features and BS algorithm	MATLAB simulation	Model Accuracy
(Canli et al., 2021) [7]	Travel time	Deep learning	Istanbul parking dataset	MAE, MSE, RMSE and Model Accuracy
(Zajam et al., 2018) [8]	Distance Travel time	Dijkstra's Algorithm	Real-time traffic data	Average waiting time
(Amini et al., 2016) [9]	Parking space utilization	Genetic Algorithm and Particle Swarm Optimization Algorithms	Simulation	Charging rate of EV parking lots
(Raichura et al., 2014) [10]	Parking space utilization	FCFS and PR	Simulation	Parking Utilization rate Fair Parking Allocation rate Communication Cost in terms of messages
(Arellano-Verdejo et al., 2016) [12]	Distance Travel time	SSEA, RS and SA	Simulation	Classic statistical indicators

(*continued*)

Table 1. (*continued*)

Reference	Category/objective	Proposed and used methods	Data sources	Performance metrics
(Pham et al., 2015) [13]	Travel time	A mathematical model	Simulation	Average waiting time
(Nugraha et al., 2017) [15]	Travel time	Event-driven algorithm	Simulation	Average finding parking lot time
(Shen et al., 2019) [16]	Parking space utilization	Genetic Algorithm	Jiefang commercial center parking data	X
(Rehena et al., 2018) [17]	User preferences Travel time Distance	Multicriteria algorithm	MATLAB simulation	Average distance, resource utilization rate
(Piccialli et al., 2021) [18]	Parking space utilization Travel time	Genetic Algorithm and Deep learning	Collected data	MAE, RMSE, R^2
(Marcu et al., 2019) [19]	Parking space utilization	Dynamic allocation algorithm based on vehicle size	Simulation	Percentage of Assigned Vehicles
(Tekouabou et al., 2022) [21]	Parking space utilization	Bagging method	Birmingham parking data	MAE, RMSE, R^2
(Rahman et al., 2022) [22]	Travel time	Deep learning and Temporal Convolutional Network	Simulation	Accuracy rate
(Yang et al., 2022) [23]	Parking space utilization	Neural network	China and USA real-world datasets	MAPE

4.1 Analysis and Discussion

In our comparative study, we found that the majority of works use smart cameras [6] and sensors [15, 18, 19] as IoT devices for parking management, while only a few studies explore the use of RFID technology [13].

Also, in smart parking systems, the overall works consider parking spaces as resources [6, 10, 17], while few of them such as [9] focus on energy resource management Furthermore, the majority of works consider a single smart parking [6, 16, 19], while few works concentrate on several smart parkings in a smart city [7, 21].

Additionally, we found that optimization of travel time to access a parking spot [7, 22] and maximization of parking space utilization [6, 9] is a central goal for most of the proposed works, in order to reduce urban congestion and pollution. Finally, we notice that relatively few works focused on optimizing distance [8] and user preferences [17]. Optimizing distance contributes to reducing traffic congestion and improves overall urban mobility, and ensuring satisfaction of driver's preferences enhances the overall user experience, promotes customer loyalty, and fosters a positive perception of smart parking systems.

By analyzing the table, we notice that machine learning are the most commonly used methods to predict available parking spaces [7, 22]; genetic algorithms [9, 16, 18] are crucial for optimal allocation with multiple parameters.

Moreover, we noticed that some works use real-world datasets such as China parking data [23], Istanbul parking data [7], and others to evaluate the performance of their algorithms. Other works are oriented towards simulation-based experiments using tools such as MATLAB [6, 17].

Furthermore, common algorithms such as FCFS, Priority (PR) algorithm, Study State Evolutionary Algorithm (SSEA), Dijkstra and Genetic Algorithm (GA) are used to improve the efficiency and profitability of smart parking,

In order to proof the effectiveness of the proposed solutions, evaluations were exhibited based on a set of performance metrics depending on the objectives and the technologies used in the solution. Some studies focus on resource utilization rate [17], while others consider factors like model accuracy [6] or multiple metrics such as MAE (Mean Absolute Error), MSE (mean Squared Error), MAPE (Mean Absolute Percentage Error) and RMSE (Root Mean Square Error) [7, 18, 23].

5 Proposed Solution

The objective of our work is to propose a solution that ensures efficient allocation of several parking spaces in a smart city by optimizing four criteria: the distance, the travel time to access a parking spot, driver's preferences and parking space utilization thus reducing traffic congestion, and improving the overall parking experience for drivers.

First, for a conceptual support of our approach, we propose a meta-model that brings out the main concepts of smart parking in IoT context. Then, we show the system architecture. Finally, we detail the allocation approach.

5.1 Metamodel of Concepts for Smart Parking in IoT

Figure 1 presents a metamodel inspired from our previous work [24] and extended to new concepts. An IoT node consists of a set of components including both IoT devices or services. Each IoT device (smart camera, sensor, actuator) generates and/or uses data. The user of the smart parking is attached to one vehicle that is assigned to a parking spot (place). Resources in a smart parking system can be either virtual like protocols, or physical such as parking places (spots), IoT devices and others (computational resources, energy resources, etc.)

A smart parking is managed by a smart parking management system (SPMS) that uses resource allocation techniques aiming to optimize one or more criteria: distance, travel time, parking space utilization, and drivers' preferences.

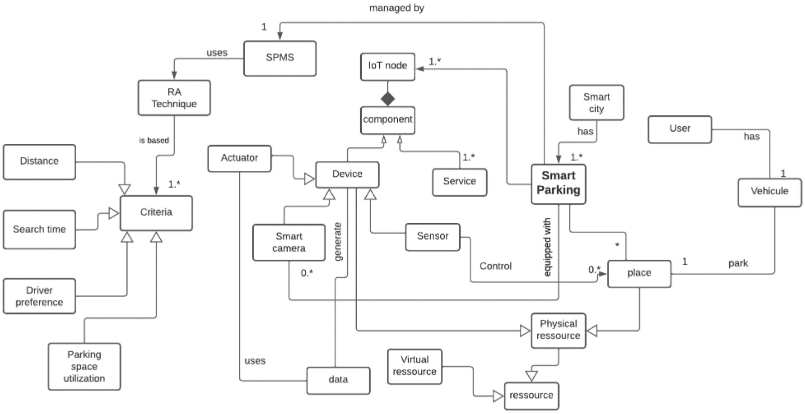


Fig. 1. Metamodel of concepts for smart parking in IoT.

5.2 System Architecture and Application Overview

Using advanced IoT technologies, the application dynamically allocates available parking spots, helping drivers to find the most suitable parking spots based on their preferences and needs. The application offers a range of features, including dynamic allocation of a parking space for a driver, providing real-time information on available/occupied parking spaces and providing the nearest available parking spot.

The system architecture of our smart parking application, shown in Fig. 2, follows a 4-layer design [14]. The Device Layer collects real-time data from parking sensors, cameras, and other devices. The Network Layer enables secure communication between the Device Layer and the Service Layer, using various communication technologies such as LANs, WANs, 4G/5G, and Bluetooth. The Service Layer processes and analyzes data to optimize parking and gain insights. The Application Layer provides user interfaces for web and mobile applications, facilitating user access and interaction with the Smart Parking System.

The architecture efficiently manages multiple smart parking spaces in a smart city, with sensors gathering data on parking spot availability. Real-time data is transmitted to the Cloud for storage and processing. Users receive up-to-date information on parking availability when they send a request to the system, which processes and responds with the best parking spot (meeting all optimization criteria), updating the database in real-time.

5.3 Genetic Algorithm for Optimization

Genetic Algorithm (GA) is an optimization method that draws inspiration from natural selection principles [20]. In our approach, we use the genetic algorithm when several users' requests occur at the same time (in case of heavy traffic). GA helps to optimize travel time, distance, parking space utilization and preferences involved in assigning parking spaces.

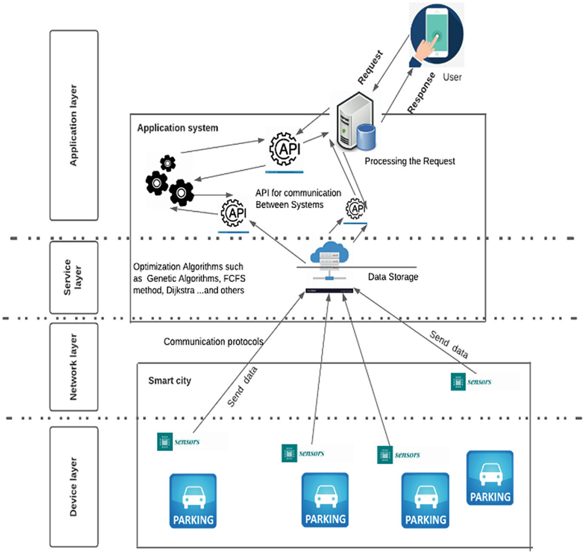


Fig. 2. Proposed system architecture of smart parking in smart city

GA Parameters. To obtain the best parameters, we employed sensitivity analysis by testing all possible combinations of values for various parameters. We have listed possible operator combinations for selection, mutation, and crossover. For selection, the types are “tournament”, “roulette_wheel”, “rank” and “random”. For mutation, the options include “random”, “swap”, “inversion,” and “scramble”. Lastly, for crossover, the types are “single_point”, “two_points” and “uniform”.

By evaluating the results of each combination using the genetic algorithm, we were able to determine which set of parameters performs the best solution. The results showed that the best combination for crossover, mutation, and selection is *[two_points, swap, roulette_wheel]* respectively.

Objective Function for Optimization. In our approach, we conducted a multi-objective optimization using four main criteria: distance, travel time, vehicle size, and driver preferences (such us parking spot near the entrance, covered spot, spot for disabled persons, spot near public transport). Our aim is to *minimize* the distance, and to *reduce* travel time, also to *maximize* the utilization of the parking space and to *meet as possible* all drivers’ preferences. By treating distance and time as separate criteria, we can give importance to minimize distance while recognizing that time is not just determined by distance but it is also influenced by various factors such as traffic conditions and speed limits. Considering the two criteria, separately, allows for a balanced approach that considers the importance of minimizing distance while considering the factors that affect travel time. Additionally, maximizing the use of the parking space is done by considering the size of the vehicle, ensuring that it corresponds to the assigned parking spot; any mismatch between the vehicle size and the parking spot results in a *penalty*. Moreover, maximizing driver’s satisfaction is done by ensuring that the parking spot characteristics align with the driver’s preferences. Any missing preference in relation to the parking spot's features incurred a *penalty*. By incorporating these penalties, we effectively optimize all four criteria simultaneously.

The objective function is formulated as follows:

$$f = 1/(wd_i * d_i) + (wp_i * p_i) + (wt_i * t_i) + (ws_i * s_i)) \tag{1}$$

f is the value of the objective function for the given solution. The weights assigned to all criteria play a crucial role in evaluating the solution; wd_i indicates the weight of distance criteria, wp_i is the weight of driver preferences criteria, wt_i is the weight of travel time criteria and ws_i reflects the weight of vehicle size criteria. The values d_i, p_i, t_i and s_i correspond to the distance criteria value, penalty based on driver preferences for mismatch, travel time, and penalty for vehicle size mismatch, respectively.

By using this objective function, we were able to achieve a comprehensive and coherent optimization strategy, addressing the needs of drivers in terms of proximity, efficiency, space utilization, and personal preferences.

GA Pseudocode

Algorithme 1 OptiPark: An Optimal Resource Allocation Algorithm for Smart Parking Systems

```
Input : Population size, maximum number of generations (MAX), Driver data, Parkings data
Output : Best solution (list of (parking spot ID, Driver ID))
Initialization : Random generation of solutions, number of iterations t=0, evaluate the objective function of each solution
While (t < MAX)
    Select the fittest solutions for reproduction
    Apply reproduction operators (crossover and mutation) to create a new population of solutions.
    Evaluate the fitness of each solution in the new population using the evaluation function.
    Replace the initial population with the new population of solutions
Return the best solution
```

5.4 FCFS for Affectation of Parking Spots

In our smart parking application, we can use the FCFS (First-Come, First-Served) algorithm to allocate parking spaces to drivers. Unlike the genetic algorithm (GA), the FCFS is used when requests come in one after another.

In this case, the algorithm simply assigns the next available and nearest spot to each incoming request in the order of its arriving (while minimizing the distance between the driver and the assigned parking place). This method ensures fairness in allocation but may not result in the most optimal overall assignment compared to using a genetic algorithm when many requests occur simultaneously.

6 Experimental Results

The following section presents the experimental results, including data simulation and the corresponding results.

6.1 Data Simulation

In our work, we simulated drivers' data, and parking spaces data. The goal is to generate representative data that can be used to perform the tests needed to validate and evaluate our approach.

Example of simulated Available Parking Spaces Data

parking_space = {"parking_id": 1, "location": (5, 10), "place_size": "medium", "availability": True, "place_characteristics": ["covered", "handicap_accessible"]}

Example of Simulated Driver Data

driver = {"driver_id": 1, "driver_location": (3, 7), "vehicle_size": "medium", "driver_preferences": ["covered", "near_transport"]}

With the simulated data, we developed a genetic algorithm using PyGAD library in Python and conducted multiple tests to obtain the best parameters and weights. Figure 3 presents the fitness function of the GA algorithm. It is a maximization function, meaning that the values of the fitness function increase over time. The data used in the GA algorithm to obtain the presented plot is randomly generated. It consists of 350 drivers and 35 parking where each parking contains 100 available parking spots.

The best solution is the one with the highest value. When the fitness value plateaus, indicating a lack of improvement over a certain number of generations, we stop and return the solution with the highest fitness value.

For the selection, mutation, and crossover functions, we employed the best combination previously mentioned.

Regarding the weights, we used the values 0.6, 0.2, 0.1 and 0.1, respectively for the criteria: preferences, distance, vehicle size and travel time to reflect the relative importance of each criterion. We assigned the weight of 0.1 to the criteria of vehicle size and travel time because although the size of the vehicle is a crucial factor in ensuring good use of the parking space, assigning it a higher weight could lead to potential constraints to accommodate larger vehicles. Similarly, travel time (weight = 0.1) is a big concern for drivers, but we wanted to avoid prioritizing it and potentially sacrificing other important criteria.

The weight of 0.6 assigned to the preference criterion indicates its predominant role in our optimization process. Driver satisfaction is highly dependent on satisfying their preferences. By assigning a higher weight to preferences, we aim to prioritize the customization of parking spaces to meet the diverse needs and preferences of drivers.

However, although distance remains an important factor, we gave it a weight of 0.2 which is relatively lower compared to preferences. We have considered that while it is desirable to minimize distance, it may be necessary to compromise on distance in favor of satisfying other preferences, such as the availability of parking spaces close to the entrance or close to public transport.

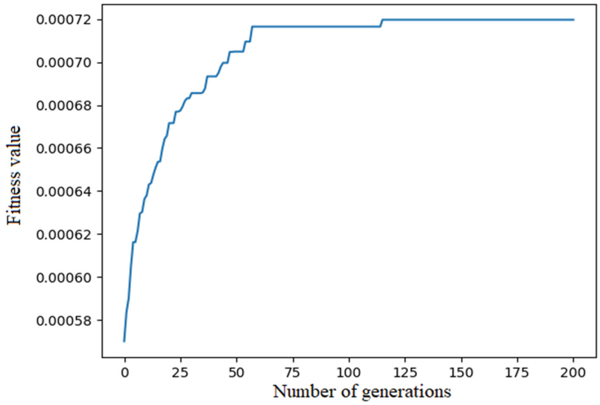


Fig. 3. The Fitness function of genetic algorithm

By examining the plot, we can observe a general upward trend in fitness values over generations, which is encouraging. This suggests that the genetic algorithm is making progress and effectively exploring the search space.

Comparison with the FCFS Algorithm. We developed FCFS algorithm (that can be used in case of light traffic, for example by night) with the same data, we simulated the arrival of drivers one by one and we conducted multiple tests, varying the number of drivers and available parking spaces in the smart city. In these tests, we evaluated the performance of both the GA the FCFS algorithms using the objective function. Figure 4 presents a comparative bar chart illustrating the performance of GA and FCFS in different scenarios.

These scenarios encompassed three possibilities: (1) the number of parking spots is equal to the number of drivers, (2) the number of parking spots is greater than the number of drivers, and (3) the number of parking spots is less than the number of drivers.

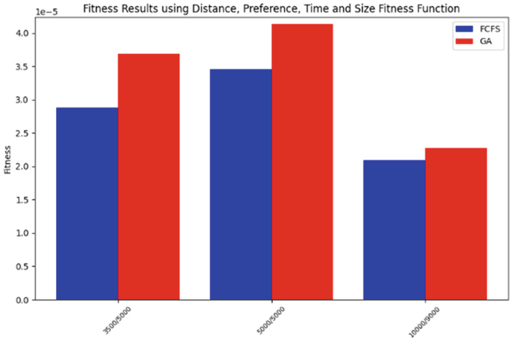


Fig. 4. Fitness Value based on Number of Drivers and Number of Available Parking Spots: Comparison between GA and FCFS

In all cases, the GA algorithm gives a better fitness than the FCFS. For example, when the number of parking spots is less than the number of drivers, with 35 parking lots, 100 parking spots in each lot and 5000 drivers, the GA algorithm achieved a fitness value of 3.68×10^{-5}, whereas the FCFS attained a fitness value of 2.87×10^{-5}.

It appears from Fig. 4 that GA outperforms FCFS, achieving higher fitness values, indicating its effectiveness in optimizing parking space allocation and improving overall system efficiency.

7 Conclusions and Future Works

In this paper, we have presented a real-time resource allocation solution for smart parking in a smart city context. In the existing literature, research works used the genetic algorithm to minimize the travel time and the use of parking spaces [9, 18] and the Dijkstra algorithm to minimize the distance [8]. However, only few studies looked at drivers' preferences where the authors used for example a Multiple Criteria Decision Analysis (MCDA) [17]. In our work, we considered the four criteria: distance, travel time, use of parking space and drivers' preferences. Indeed, our focus was on minimizing distance, reducing travel time to access a parking spot, maximizing parking space utilization, and satisfying user preferences. Based on a multi-objective function, we developed an optimization method using genetic algorithm that provides effective results.

Through our experiments and comparison with a naïve method (mainly the FCFS algorithm), we found that GA outperformed FCFS in terms of fitness value. The fitness value of GA was consistently greater than the FCFS one. This indicates that GA was able to minimize distance and travel time, maximize the use of parking spaces, and satisfy driver preferences.

Overall, our findings demonstrate that GA offers significant improvements over the traditional FCFS algorithm in terms of optimizing smart parking resources in a smart city environment. This research contributes to the ongoing efforts in developing efficient and intelligent solutions for urban mobility and resource management.

However, there are still important aspects to consider in future works. One of these aspects is the use of machine learning techniques to predict parking demands. Predictive models can significantly enhance the predictive capabilities of smart parking systems, allowing more accurate allocation of parking spaces and better management of parking resources. Another important aspect to consider is the integration of autonomous and connected vehicles [11]. The advancement of autonomous and connected vehicles presents an exciting opportunity to further improve smart parking. By integrating these technologies, we can enhance the efficiency of parking allocation and reduce traffic congestion.

References

1. Kotb, A.O., Shen, Y.-C., Huang, Y.: Smart parking guidance, monitoring and reservations: a review. IEEE Intell. Transp. Syst. Mag. **9**(2), 6–16 (2017)
2. Barriga, J.J., et al.: Smart parking: a literature review from the technological perspective. Appl. Sci. **9**, 4569 (2019)

3. Paidi, V., Fleyeh, H., Håkansson, J., Nyberg, R.G.: Smart parking sensors, technologies and applications for open parking lots: a review. IET Intell. Transp. Syst. **12**, 735–741 (2018)
4. Abrar, F., Mehedi, H., Muhtasim, C.: Smart parking systems: comprehensive review based on various aspects. Heliyon. **7**, e07050 (2021)
5. Geng, Y., Cassandras, C.G.: New smart parking system based on resource allocation and reservations. IEEE Trans. Intell. Transp. Syst. **14**(3), 1129–1139 (2013)
6. Varghese, A., Sreelekha, G.: An efficient algorithm for detection of vacant spaces in delimited and non-delimited parking lots. IEEE Trans. Intell. Transp. Syst. **21**(10), 4052–4062 (2020)
7. Canli, H., Toklu, S.: Deep learning-based mobile application design for smart parking. IEEE Access **9**, 61171–61183 (2021)
8. Zajam, A., Dholay, S.: Detecting efficient parking space using smart parking. In: 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Bengaluru, India, pp. 1–7 (2018)
9. Amini, M.H., Moghaddam, M., Karabasoglu, O.: Simultaneous allocation of electric vehicles' parking lots and distributed renewable resources in smart power distribution network. Sustain. Cities Soc. **28**, 332–342 (2017)
10. Raichura, K., Padhariya, N.: edPAS: event-based dynamic parking allocation system in vehicular networks. In: 2014 IEEE 15th International Conference on Mobile Data Management, Brisbane, QLD, Australia, pp. 79–84 (2014)
11. Balzano, W., Vitale, F.: DiG-Park: a smart parking availability searching method using V2V/V2I and DGP-class problem. In: 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA), Taipei, Taiwan, pp. 698–703 (2017)
12. Arellano-Verdejo, J., Alba, E.: Optimal allocation of public parking slots using evolutionary algorithms. In: 2016 International Conference on Intelligent Networking and Collaborative Systems (INCoS), Ostrava, Czech Republic, pp. 222–228 (2016)
13. Pham, T.N., Tsai, M.-F., Nguyen, D.B., Dow, C.-R., Deng, D.-J.: A cloud-based smart-parking system based on internet-of-things technologies. IEEE Access **3**, 1581–1591 (2015)
14. Ahmed, S., Rahman, M.S., Rahaman, M.S.: A blockchain-based architecture for integrated smart parking systems. In: 2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom workshops), pp. 177–182 (2019)
15. Nugraha, I.G.B.B., Tanamas, F.R.: Off-street parking space allocation and reservation system using event-driven algorithm. In; 2017 6th International Conference on Electrical Engineering and Informatics (ICEEI), pp. 1–5 (2017)
16. Shen, T., Hua, K., Liu, J.: Optimized public parking location modelling for green intelligent transportation system using genetic algorithms. IEEE Access **7**, 176870–176883 (2019)
17. Rehena, Z., Mondal, M.A., Janssen, M.: A multiple-criteria algorithm for smart parking: making fair and preferred parking reservations in smart cities. In: Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age, pp. 1–9 (2018)
18. Piccialli, F., Giampaolo, F., Prezioso, E., Crisci, D., Cuomo, S.: Predictive analytics for smart parking: a deep learning approach in forecasting of IoT data. ACM Trans. Internet Technol. (TOIT) **21**(3), 1–21 (2021)
19. Marcu, I. M., Ţigănuş, A., Drăgulinescu, A.M., Suciu Jr., G.: A new approach on smart-parking concept. In: Proceedings of the 6th Conference on the Engineering of Computer Based Systems, pp. 1–9 (2019)
20. Goldberg, D.E.: Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley Publishing Company (1989)
21. Tekouabou, S.C.K., Alaoui, E.A., Cherif, W., Silkan, H.: Improving parking availability prediction in smart cities with IoT and ensemble-based model. J. King Saud Univ. – Comput. Inf. Sci. **34**(3), 687–697 (2022)

22. Rahman, M.T., Zhang, Y., Arani, S.A., Shao, W.: MDLpark: available parking prediction for smart parking through mobile deep learning. In: Ma, H., Wang, X., Cheng, L., Cui, L., Liu, L., Zeng, A. (eds.) CWSN 2022. CCIS, vol. 1715, pp. 182–199. Springer, Singapore (2022). https://doi.org/10.1007/978-981-19-8350-4_15
23. Yang, H., Ke, R., Cui, Z., Wang, Y., Murthy, K.: Toward a real-time smart parking data management and prediction (SPDMP) system by attributes representation learning. Int. J. Intell. Syst. **37**, 4437–4470 (2022)
24. Tair, K., Boukhedouma, S.: Integration of internet of things in BPM lifecycle: concepts and comparison of approaches. In: 2022 First International Conference on Big Data, IoT, Web Intelligence and Applications (BIWA), pp. 71–76. IEEE (2022)

Marine Predatory Algorithm for Feature Selection in Speech Emotion Recognition

Osama Ahmad Alomari[2], Muhammad Al-Barham[2], and Ashraf Elnagar[1](✉)

[1] Department of Computer Science, University of Sharjah, Sharjah, UAE
ashraf@shrajah.ac.ae

[2] MLALP Research Group, University of Sharjah, Sharjah, UAE
{oalomari,malbarham}@shrajah.ac.ae

Abstract. In recent times, the recognition of human emotional states expressed through speech communication through speech communication has garnered significant interest among researchers in human-computer interaction. Various systems have been advanced to categorize states of speech emotion using features extracted from spoken utterances. Feature extraction plays a vital role in developing speech emotion recognition systems, as the performance of the learning model improves when the extracted features are reliable and capture the emotional characteristics of speech samples. However, some of the extracted features may be redundant, irrelevant, or noisy, which can diminish the classification performance of speech emotion recognizers. To address this issue and select efficient and precise speech emotion features, this paper introduces an effective feature selection method called MPA-KNN, which combines the marine predators algorithm with the KNN classifier. The proposed method's performance is evaluated using three distinct speech databases: the Surrey Audio-Visual Expressed Emotion (SAVEE), the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS), and the Arabic Emirati-accented speech database. The results showed that MPA-KNN improved the accuracy of speech emotion recognition compared to classical machine learning and optimization feature selection methods. Moreover, it overcomes several competitors in the literature that utilize the same datasets.

Keywords: Feature Selection · Emotion speech datasets · Marine Predators Algorithm

1 Introduction

Speech has become a fundamental means of extracting valuable information about a person's emotional state, establishing it as a standard form of human communication. However, the complexity of speech signal structure poses a significant challenge for computers in accurately discerning human emotions. The traditional approach to constructing an emotion identification system involves three main components: an emotion recognizer, feature extraction techniques,

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 279–291, 2024.
https://doi.org/10.1007/978-3-031-56728-5_24

and feature selection methods [36]. These components must address two critical factors that impact the performance of the emotion identification system: identifying the essential features that characterize emotions and creating a precise speech emotion recognition model [41]. The identification of predictive, discriminative, and meaningful features plays a crucial role in effectively classifying emotional states [34,39]. In the field of emotion identification and other domains, researchers have explored various metaheuristic search methods to tackle feature selection challenges [8]. For instance, in speech emotion recognition, a feature selection approach has been developed using two popular techniques: Cuckoo Search and genetic algorithm-II based on a nondominated sorting strategy (NSGA-II). These techniques modify the initial evolution of the population of the metaheuristic search algorithms to enhance feature selection [46]. Additionally, particle swarm optimization and a biogeography-based algorithm have been employed to create feature selection techniques for speech emotion recognition [43]. Another novel approach combines gradient-boosting decision trees with artificial bee colony optimization to improve the efficiency and informativeness of the selected features [43]. However, a significant challenge in optimization-based feature selection methods is premature convergence, which hinders the development of the best speech recognition system [10,16,44]. Consequently, searching for a new and robust optimization algorithm that can effectively address the feature selection problem in speech emotion recognition remains an dynamic and important area of research.

The Marine Predators Algorithm (MPA) is an innovative swarm-based optimization algorithm introduced by Faramarzi [13]. It emulates the social behavior of marine predators. The optimization process is formulated as follows: MPA utilizes two types of random walks, namely levy flight and the Brownian method, along with random population management rules. This algorithm is characterized by its flexibility in implementation, simplicity, easily adjustable parameters, and strong search capabilities. Moreover, MPA proved its quality in several applications such as ECG signals classification [17], dynamic clustering [42], photovoltaic array reconfiguration [45], fog computing [2], and COVID-19 medical images segmentation [1,3]. These successful applications highlight the algorithm's potential and its ability to deliver high-quality results in diverse fields.

This study introduces a novel approach for feature selection in emotion recognition systems, drawing inspiration from the appealing features and impressive performance of MPA. The research makes four main contributions, which can be outlined as follows:

1. An enhanced system for recognizing emotions in speech that utilizes an adapted MPA as a feature selector and the KNN classifier.
2. The efficiency of MPA-KNN was assessed on three datasets: a locally constructed real dataset, SAVEE database [21], the Ryerson Audio-Visual Database (RAVDESS) [28], and the Arabic Emirati-accent dataset [35].
3. Assessment of the performance of MPA against well-known optimization feature selection algorithms.

The remainder of the paper is structured as follows: Sect. 2 discusses the related work. Section 3 presents the research background. Section 4 describes the proposed method. Section 4.3 describes the results and discussion. Section 4.4 of the paper presents the concluding remarks and outlines future work.

2 Related Work

In human communication, language encompasses not only the syntactic and semantic aspects of a sentence but also the emotional state conveyed by the speaker. Speech Emotion Recognition (SER) is a field that aims to classify the emotions of a speaker based on their speech recordings, enabling the inference and prediction of their psychological states. This section provides a comprehensive overview of the key research conducted in the domain of SER, covering various aspects such as feature selection, feature extraction, machine learning techniques applied, the experimented datasets, and the performance achieved compared to avaliable benchmarks.

The study conducted by Gomathy et al. [15] introduced the Enhanced Cat Herd Optimization (ECSO) algorithm as a solution for selecting optimal features in order to minimize redundant features and computational costs. In this approach, the basic Cat Swarm Optimization (CSO) algorithm simulates a cat's behavior by considering its position, speed, fitness score, and search or tracking mode. To enhance population diversity and search capability, the authors proposed the opposite learning (OBL) method. Additionally, they employed an SVNN classifier in conjunction with the ESCO method, leading to a high detection rate.

The authors of [20] proposed an architecture designed to extract various features from speech signals, including sound net representations, chromatograms, and Mel-frequency cepstrum coefficients. Convolutional neural networks were utilized to identify emotions in samples obtained from the Ryerson Audio-Visual Database of Emotional Speech and Song. In another study, Shen et al. [37] utilized the Support Vector Machine algorithm to categorize extracted features, such as linear prediction cepstrum coefficients, mel-frequency cepstrum coefficients, pitch, and energy, from the Berlin Sentiment Database. The results demonstrated the strong capabilities of SVM in accurately recognizing emotional states.

Kanwal et al. [22] introduced a cluster-based genetic algorithm for optimizing feature selection. Their method incorporates fitness rating level clustering to identify outliers that can be eliminated in the subsequent phase. The reported accuracy using their approach was 89.6%.

In a different study by Daneshfar et al. [9], a modified quantum behavior particle swarm optimization (QPSO) method was introduced to address dimensionality reduction and the selection of machine learning parameters. Gaussian Mixture Models (GMM) were utilized for emotion classification. The authors also emphasized the strong correlations between human speech style and glottal waveform features, which led to improved recognition performance. However, it is important to note that the QPSO variant introduced higher complexity and slower convergence speed, resulting in a steeper training curve, despite the authors' claim of its potential for real-time applications.

3 Research Background

3.1 Emotion Speech Corpus

Benchmarked speech corpus such as Emirati speech database (ESD), RAVDESS corpus, and SAVEE database, are considered in this work. The ESD consists of a collection of eight unique sentences spoken by thirty-one different individuals. These speakers expressed six different emotions, including disgust, sad, happy, neutral, angry, and fear" emotions, and each sentence was repeated nine times. Consequently, there are a total of 13,392 audio files (31 × 8 × 9 × 6) in the dataset. Among these files, 90% were reserved for training purposes, while the remaining 10% were reserved for testing. In the RAVDESS dataset, there are a total of 7,356 files, but only audio files (1,440 in total) associated with eight specific emotions were considered for this research. From the overall files, 90% of the audio files were reserved for training, and the remaining 10% were reserved for testing. The SAVEE dataset contains the neutral emotion in addition to six other emotions. The neutral emotion comprises 120 utterances, whereas each of the other six emotions has 60 utterances. This leads to a total of 480 occurrences (6 × 60 + 120). In this study, 90% of the utterances were employed for training, while 10% were allocated for testing.

3.2 Feature Extraction

Numerous characteristics, or speech features, may be found in sound waves and speech signals. Accuracy of the suggested system is greatly impacted by the success of this level of the "speech signal processing" paradigm [26]. We have developed a framework for extracting Root Mean Square (RMS), Mel spectrogram, Chroma, Zero Crossing Rate (ZCR), and Mel Frequency Cepstral Coefficients (MFCC).

Mel-Frequency Cepstral Coefficient (MFCCs). MFFCs are widely utilized for feature extraction [26,31,33]. However, one major drawback of MFCCs is their susceptibility to distortion caused by their reliance on spectral structure. To address this issue, speech signals, which contain aperiodic content, can be leveraged to overcome the challenge [19]. The computation of MFCCs involves the use of two filters, which are set linearly for frequencies below 1,000 kHz and logarithmically set for frequencies above 1,000 Hz[29,32]. The formula used to compute MFCCs is provided by Alsabek et al. [5].

Zero Crossing Rate (zCR). The (ZCR) is a metric that quantifies the frequency at which the amplitude diminishes to zero during transmission. It represents the total count of transitions or oscillations from positive to negative and vice versa within a specific time interval. The ZCR provides an assessment of the primary frequency component of the signal [6,18,25]. The formula used to compute ZCR is provided by Staudinger and olikar [38].

Chroma. In music, the chroma feature is a term used to capture the tonal characteristics of an audio stream. It pertains to twelve distinct "pitch classes", often referred to as chroma features or chromograms. Employing chromatic features, such as pitch class profiles divided into twelve subdivisions, can be beneficial for evaluating music with meaningful pitch classifications. A notable advantage of chroma-related features is their capability to capture both the melodic and harmonic components of music while being resilient to changes in instrumentation and timbre. Consequently, chroma features are considered vital for conducting further semantic analysis, including tasks such as chord identification and estimation of harmonic similarity [23].

Root Mean Square. The Root Mean Square (RMS) is a measure that quantifies the volume or amplitude of an acoustic stream. To compute the RMS score, the square root of the sum of the mean squares of the sound sample amplitudes is taken. The formula used to compute RMS is provided by M. B. Er [12].

Mel Spectrogram. The Mel scale is a perceptual scale that represents pitches that are perceived as being equally distant from each other by listeners. It allows us to distinguish between different sounds even when they have different frequencies, but they are in proximity and under similar environmental conditions. For instance, a listener can discern the difference between sounds at 10,000 Hz and 15,000 Hz. The Mel spectrogram is generated by mapping frequencies to the Mel scale, which can be achieved using the Fourier transform [30,40]:

4 The Proposed MPA-KNN Feature Selection Method

4.1 Marine Predatory Algorithm

MPA is a nature-inspired optimization that emulates the hunting behavior of predators in the searching journey to their prey. When navigating through areas abundant in prey, MPA employs two motion strategies: Levy flight and Brownian motion. Typically, in metaheuristic algorithms, Levy flight is utilized as an exploitation operator, while Brownian motion serves as an exploration operator [11]. However, the creators of MPA integrated the strengths of both strategies to strike a proper balance between exploration and exploitation. The mathematical formulation of MPA can be formulated as follows:

Initialization. Like most population-based metaheuristic algorithms, MPA initiates the search process by generating individual solutions with a uniform distribution, as described in Eq. 1.

$$X_0 = X_{min} + rand(X_{max} - X_{min}) \tag{1}$$

Here, *rand* denotes a random number uniformly distributed between 0 and 1, while X_{min} and X_{max} refer to the minimum and maximum limits that constrain the values of decision variables during the search process.

Elite and Prey Matrix Construction. Based on the concept of "survival of the fittest", the predators are ranked, and the top predator is utilized to construct an Elite matrix (E) with dimensions $n \times d$, where n represents the number of predators and d represents the number of decision variables.

$$\boldsymbol{E} = \begin{bmatrix} X_{1,1}^1 & X_{1,2}^1 & \cdots & X_{1,d}^1 \\ X_{2,1}^1 & X_{2,2}^1 & \cdots & X_{2,d}^1 \\ \vdots & \vdots & \cdots & \vdots \\ X_{n,1}^1 & X_{n,2}^1 & \cdots & X_{n,d}^1 \end{bmatrix}, \tag{2}$$

where X denotes the predator with the highest fitness, which is replicated n times to form the Elite matrix E. Another matrix with equal dimensions as E is considered as the reference matrix for updating the positions of the prey and predators, as shown below:

$$\boldsymbol{Py} = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,4} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,d} \\ \vdots & \vdots & \cdots & \vdots \\ X_{n,1} & X_{1,2} & \cdots & X_{n,d} \end{bmatrix}, \tag{3}$$

These two matrices have a significant impact on steering the optimization process.

Optimization Process. In the first step, the MPA search process is initialized with a high-velocity ratio, assuming that predators move faster before approaching their prey. This initialization is similar to the common process in many metaheuristic algorithms at the initial stage of evolution. Mathematically, this step can be expressed as follows:

$$\text{while} \quad t < \frac{1}{3} \cdot t_{\max} \quad \text{do} \quad s\vec{z}_i = \vec{R_l} \otimes (\vec{E_i} - \vec{R_L} \otimes p\vec{y}_i) \tag{4}$$

$$\vec{Py}_i = \vec{Py}_i + P \cdot R \otimes sz_i \tag{5}$$

Here, $\vec{R_L}$ is a vector of values randomly assigned between 0 and 1, following a normal distribution that represents Brownian motion. The prey movement is formulated by multiplying $\vec{R_L}$ with the prey. The constant $P = 0.5$ denotes a fixed value, and sz represents the step size. $Iter$ refers to the first iteration, while Max_{iter} represents the maximum iteration. This optimization scenario, applied only during the initial one-third of the iterations, promotes high exploration since the initialization values for step size and velocity are in a wide range.

In the intermediate step, both prey and predator move at an equal speed. Within the optimization context, this step acts as an intermediary phase between exploration and exploitation, where the transition from exploitation to exploration occurs temporarily. MPA divides the population into two halves, where half of the population engages in exploratory search using the Brownian method, while the second half utilizes the exploitative search operator employing the Levy method. Mathematically, this step can be described as follows:

$$\frac{1}{3} \cdot t_{\max} < t < \frac{2}{3} \cdot t_{\max} \tag{6}$$

In the case of the initial half of the population:

$$\vec{sz} = \vec{R_l} \otimes (\vec{E_i} - \vec{R_{Levy}} \otimes \vec{py_i}) \tag{7}$$

$$\vec{py_i} = \vec{py_i} + P \cdot \vec{R} \otimes \vec{sz} \tag{8}$$

The vector R_{Levy} is generated with random values between 0 and 1, which follows the Levy distribution. The motion behaviour of predators in the Levy strategy is formulated by multiplying R_{levy} with py_i, while the prey's movement is influenced by adding the step size to the current prey position. The second half of the population is updated as follows:

$$\vec{sz} = \vec{R_B} \otimes (\vec{R_B} \otimes \vec{E_i} - \vec{P_i}) \tag{9}$$

$$\vec{py_i} = \vec{E_i} + P \cdot CF \otimes \vec{sz} \tag{10}$$

where CF is a parameter that is iteratively updated to regulate the size of each step. It is defined by:

$$CF = \left(1 - \frac{t}{t_{\max}}\right)^{(2\frac{t}{t_{\max}})} \tag{11}$$

CF affects the step size through its influence on the variation in prey movement.

In the last step of the optimization process, the predator is assumed to have a faster movement than the prey. It utilizes the Levy strategy, which allows for movement in a low-velocity range and is known for its high exploitation capability. This step can be summarized as follows:

$$\text{while} \quad t > \frac{2}{3} \cdot t_{\max}$$

$$\vec{sz} = \vec{R_L} \otimes (\vec{R_L} \otimes \vec{E_i} - \vec{Py_i}) \tag{12}$$

$$\vec{py_i} = \vec{E_i} + P \cdot CF \otimes \vec{sz} \tag{13}$$

In this step, the motion of the predator in the Levy strategy is represented by $\vec{R_L} \otimes \vec{E_i}$, which mimics its behavior. Meanwhile, the prey's location is adjusted by adding the step size to the position of the elite individual. Another factor that influences the behavior of the marine predator is environmental issues like eddy formation or Fish Aggregating Devices (FADS), which are devices used to lure fish for various purposes [14]. In the presence of FADS, the predators randomly move to different positions in search of food, introducing variation into candidate solutions. This is captured by the following equation:

$$\vec{py_1} \leftarrow \begin{cases} \vec{py_1} + CF[\vec{X_L} + R \otimes (\vec{X_U} - \vec{X_L})] \otimes \vec{U} & \text{if } r < FADS \\ \vec{py_i} + [FADS(1-r) + r](\vec{py_{r1}} - \vec{py_{r2}}) & \text{if } r > FADS \end{cases} \tag{14}$$

r represents a random number ranging from 0 to 1. The value FADS = 0.2 indicates the probability of incorporating FADS into the optimization process. Additionally, $\vec{U}$ is a binary vector consisting of zeros and ones.

4.2 MPA-KNN Implementation Process

The proposed method consists of two main stages, including feature extraction and feature selection. Feature extraction plays a crucial role in the development of speech emotion recognition systems, as it involves analyzing the emotions expressed in speech audio data by extracting a set of relevant features. In this study, the 'Emirati-accented' dataset employs the extraction of 40 MFCCs per frame, resulting in a total of 40 features. For the 'RAVDESS' dataset, features are extracted using Mel, Chroma, and the concatenation of "delta-delta, MFCCs-delta, and MFCCs", yielding a total of 260 features. In the case of the SAVEE dataset, MEL, ZCR, Chroma, RMS, and MFCCs are used, resulting in a total of 182 output features. In Feature selection, the extracted features from each dataset are passed for further optimization using MPA as feature subset generation and KNN to classify each candidate feature subset produced by MPA. The proposed MPA-KNN steps is illustrated below:

1. Step1: In general, the MPA starts by a set of random solutions (i.e., candidate feature subsets) drawn from the binary search space of the feature selection problem. Each solution is a binary vector with a size corresponding to the number of the raw features present in speech dataset. The 1's in the binary vector imply that this feature is selected; otherwise, it is ignored.

2. Step 2: The fitness function score is assessed for each solution in the population, representing candidate feature subsets. The features marked as ‘1’ are considered as input features for the reduced-dimensional dataset. The dataset is then split into training and testing data sets, with a 90%:10% ratio for training and testing, respectively. A classification model is built using KNN method based on the training data. The model is then used to classify the speech samples in the testing data. The construction of the elite and prey matrices, as described in Eqs. (2) and (3), respectively, is an important step in the process.
3. Step 3: During the initial one-third of the iterations, the population solutions are updated utilizing Eqs. (4) and (5).
4. Step 4: For the second third, the solutions/predators are divided into two halves. For the first half, the levy flight is applied as introduced in Eqs. (7) and (8). For the second half, the solutions are passed to the Brownian method as introduced in Eqs. (9) and (10).
5. Step 5: For the last third of the iterations, the solutions in the population are updated using Eq. (12) and (13).
6. Step 6: The new solutions are evaluated and the top predator is updated.
7. Step 7: Implement the effect of FADs for each predator using Eq. (14) and save the high quality solutions in the memory.
8. Step 8: The iterations may continue till termination condition is met.
9. Step 9: The best emotion features are produced.

4.3 Results and Discussion

In this section, we evaluated the efficacy of our suggested technique by contrasting it with a few widely-used and up-to-date optimization algorithms. These include the Slime Mould Algorithm (SMA) [27], Arithmetic Optimization Algorithm (AOA) [4], White Shark Optimizer (WSH) [7], and Particle Swarm Optimization (PSO) [24]. Three speech emotion datasets were used in the evaluation including SAVEE, RAVDESS, and Emirati-accented. All of the above algorithms are stochastic in nature and utilized as feature selection methods. To ensure unbiased results, each algorithm was conducted with 10 separate runs, and the mean results were calculated. The parameter settings associated with each optimization method are provided in Table 1. The results of MPA-KNN algorithm and other optimization feature selection algorithms are assessed based on Accuracy, Precision, Recall, F1 Score, and Wilcoxon signed-rank statistical test, as shown in Table 2. It should be noted that the programming language used in this research is Matlab version 9.12.0 (R2022a).

According to the results presented in Table 2, HMPA-KNN outranks other feature selector techniques on all performance measurements. MPA-KNN exhibited statistically significant outcomes in comparison to alternative feature selection approaches.

Table 1. Parameter settings of optimization-based feature selection methods.

Approach	Parameters
AOA	$\alpha = 5$, $\mu = 0.5$
WSO	$f_{min} = 0.07$, $f_{max} = 0.75$, $a_0 = 6.25$, $a_1 = 100$, $a_2 = 0.0005$, $\tau = 4.125$
PSO	C1 = 2, C2 = 2, W = 0.9
SMA	z = 0.03
MPA	FADs = 0.2, P = 0.5

Table 2. Results MPA-KNN and other optimization-based feature selection approaches

Datasets	Measurments	KNN	MPA-KNN	SMA-KNN	PSO-KNN	WSH-KNN	AOA-KNN
EM	Acc	83.25	**91.73**	89.17	91.37	88.40	87.97
	Precision	83.16	**91.91**	89.22	91.37	88.38	87.92
	Recall	83.26	**91.90**	89.17	91.37	88.42	87.95
	F1score	83.19	**91.88**	89.16	91.35	88.38	87.92
	T-Sig.			*	*	*	*
RAVDESS	Acc	56.64	**82.01**	73.4	80.69	78.54	77.01
	Precision	56.59	**80.82**	72.91	80.33	77.85	76.99
	Recall	57.44	**81.58**	73.57	80.64	78.62	77.17
	F1score	55.99	**80.96**	72.73	80.11	77.84	76.49
	T-Sig.			*	*	*	*
SAVEE	Acc	65.95	**84.58**	78.12	84.17	80.41	82.50
	Precision	63.29	83.62	77.04	**84.76**	81.12	82.56
	Recall	57.99	**80.44**	74.33	80.10	76.48	78.71
	F1score	58.72	80.80	73.99	**80.91**	76.73	78.65
	T-Sig.			*	*	*	*

4.4 Conclusion and Future Research

This paper presents an automated speech emotion recognition system that combines MPA-based feature selection with a KNN classifier. The proposed method aims to accurately analyze the emotional states expressed in human speech by selecting relevant features. The method's performance is evaluated on three speech emotion datasets: SAVEE, RAVDESS, and Emirati-accented. Evaluation metrics such as classification accuracy, F1-score, recall, precision, and recall are used to assess the efficiency of the proposed method. A comparison is made with other feature selection methods, including SMA-KNN, GA-KNN, PSO-KNN, AOA-KNN, and WSH-KNN. The findings show that the HMPA-KNN method surpasses the other methods. Future work involves enhancing the MPA algorithm by incorporating different intelligent feature selection schemes, introducing new search operators, or hybridizing it with other population-based algorithms.

References

1. Abd Elaziz, M., et al.: An improved marine predators algorithm with fuzzy entropy for multi-level thresholding: real world example of COVID-19 CT image segmentation. IEEE Access **8**, 125306–125330 (2020)
2. Abdel-Basset, M., Mohamed, R., Elhoseny, M., Bashir, A.K., Jolfaei, A., Kumar, N.: Energy-aware marine predators algorithm for task scheduling in IoT-based fog computing applications. IEEE Trans. Industr. Inf. **17**(7), 5068–5076 (2020)
3. Abdel-Basset, M., Mohamed, R., Elhoseny, M., Chakrabortty, R.K., Ryan, M.: A hybrid COVID-19 detection model using an improved marine predators algorithm and a ranking-based diversity reduction strategy. IEEE Access **8**, 79521–79540 (2020)
4. Abualigah, L., Diabat, A., Mirjalili, S., Abd Elaziz, M., Gandomi, A.H.: The arithmetic optimization algorithm. Comput. Methods Appl. Mech. Eng. **376**, 113609 (2021)
5. Alsabek, M.B., Shahin, I., Hassan, A.: Studying the similarity of COVID-19 sounds based on correlation analysis of MFCC. In: 2020 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI), pp. 1–5. IEEE (2020)
6. Bachu, R., Kopparthi, S., Adapa, B., Barkana, B.D.: Voiced/unvoiced decision for speech signals based on zero-crossing rate and energy. In: Elleithy, K. (ed.) Advanced Techniques in Computing Sciences and Software Engineering, pp. 279–282. Springer, Dordrecht (2010). https://doi.org/10.1007/978-90-481-3660-5_47
7. Braik, M., Hammouri, A., Atwan, J., Al-Betar, M.A., Awadallah, M.A.: White shark optimizer: a novel bio-inspired meta-heuristic algorithm for global optimization problems. Knowl.-Based Syst. **243**, 108457 (2022)
8. Brezočnik, L., Fister, I., Jr., Podgorelec, V.: Swarm intelligence algorithms for feature selection: a review. Appl. Sci. **8**(9), 1521 (2018)
9. Daneshfar, F., Kabudian, S.J.: Speech emotion recognition using discriminative dimension reduction by employing a modified quantum-behaved particle swarm optimization algorithm. Multimed. Tools Appl. **79**(1), 1261–1289 (2020)
10. Darekar, R.V., Dhande, A.P.: Emotion recognition from Marathi speech database using adaptive artificial neural network. Biol. Inspir. Cogn. Archit. **23**, 35–42 (2018)
11. Dokeroglu, T., Sevinc, E., Kucukyilmaz, T., Cosar, A.: A survey on new generation metaheuristic algorithms. Comput. Ind. Eng. **137**, 106040 (2019)
12. Er, M.B.: A novel approach for classification of speech emotions based on deep and acoustic features. IEEE Access **8**, 221640–221653 (2020)
13. Faramarzi, A., Heidarinejad, M., Mirjalili, S., Gandomi, A.H.: Marine predators algorithm: a nature-inspired metaheuristic. Expert Syst. Appl. **152**, 113377 (2020)
14. Filmalter, J.D., Dagorn, L., Cowley, P.D., Taquet, M.: First descriptions of the behavior of silky sharks, Carcharhinus falciformis, around drifting fish aggregating devices in the Indian ocean. Bull. Mar. Sci. **87**(3), 325–337 (2011)
15. Gomathy, M.: Optimal feature selection for speech emotion recognition using enhanced cat swarm optimization algorithm. Int. J. Speech Technol. **24**(1), 155–163 (2021)
16. He, H., Tan, Y., Ying, J., Zhang, W.: Strengthen EEG-based emotion recognition using firefly integrated optimization algorithm. Appl. Soft Comput. **94**, 106426 (2020)

17. Houssein, E.H., Hassaballah, M., Ibrahim, I.E., AbdElminaam, D.S., Wazery, Y.M.: An automatic arrhythmia classification model based on improved marine predators algorithm and convolutions neural networks. Expert Syst. Appl. **187**, 115936 (2022)
18. Ibrahim, Y.A., Odiketa, J.C., Ibiyemi, T.S.: Preprocessing technique in automatic speech recognition for human computer interaction: an overview. Ann. Comput. Sci. Ser. **15**(1), 186–191 (2017)
19. Ishizuka, K., Nakatani, T., Minami, Y., Miyazaki, N.: Speech feature extraction method using subband-based periodicity and nonperiodicity decomposition. J. Acoust. Soc. Am. **120**(1), 443–452 (2006)
20. Issa, D., Demirci, M.F., Yazici, A.: Speech emotion recognition with deep convolutional neural networks. Biomed. Signal Process. Control **59**, 101894 (2020)
21. Jackson, P., Haq, S.: Surrey audio-visual expressed emotion (SAVEE) database. University of Surrey, Guildford (2014)
22. Kanwal, S., Asghar, S.: Speech emotion recognition using clustering based GA-optimized feature set. IEEE Access **9**, 125830–125842 (2021)
23. Kattel, M., Nepal, A., Shah, A., Shrestha, D.: Chroma feature extraction. In: Conference: Chroma Feature Extraction using Fourier Transform, no. 20 (2019)
24. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proceedings of ICNN 1995-International Conference on Neural Networks, vol. 4, pp. 1942–1948. IEEE (1995)
25. Kos, M., Kačič, Z., Vlaj, D.: Acoustic classification and segmentation using modified spectral roll-off and variance-based features. Digit. Signal Process. **23**(2), 659–674 (2013)
26. Kurzekar, P.K., Deshmukh, R.R., Waghmare, V.B., Shrishrimal, P.P.: A comparative study of feature extraction techniques for speech recognition system. Int. J. Innov. Res. Sci. Eng. Technol. **3**(12), 18006–18016 (2014)
27. Li, S., Chen, H., Wang, M., Heidari, A.A., Mirjalili, S.: Slime mould algorithm: a new method for stochastic optimization. Futur. Gener. Comput. Syst. **111**, 300–323 (2020)
28. Livingstone, S.R., Russo, F.A.: The Ryerson audio-visual database of emotional speech and song (RAVDESS): a dynamic, multimodal set of facial and vocal expressions in north American English. PLoS ONE **13**(5), e0196391 (2018)
29. Muda, L., Begam, M., Elamvazuthi, I.: Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. arXiv preprint arXiv:1003.4083 (2010)
30. Panda, S.K., Jena, A.K., Panda, M.R., Panda, S.: Speech emotion recognition using multimodal feature fusion with machine learning approach. Multimed. Tools Appl. **82**, 1–19 (2023)
31. Shahin, I.: Identifying speakers using their emotion cues. Int. J. Speech Technol. **14**(2), 89–98 (2011)
32. Shahin, I.: Studying and enhancing talking condition recognition in stressful and emotional talking environments based on HMMs, CHMM2s and SPHMMs. J. Multimodal User Interfaces **6**(1), 59–71 (2012)
33. Shahin, I.: Novel third-order hidden Markov models for speaker identification in shouted talking environments. Eng. Appl. Artif. Intell. **35**, 316–323 (2014)
34. Shahin, I., Alomari, O.A., Nassif, A.B., Afyouni, I., Hashem, I.A., Elnagar, A.: An efficient feature selection method for Arabic and English speech emotion recognition using grey wolf optimizer. Appl. Acoust. **205**, 109279 (2023)
35. Shahin, I., Nassif, A.B., Hamsa, S.: Emotion recognition using hybrid gaussian mixture model and deep neural network. IEEE Access **7**, 26777–26787 (2019)

36. Sheikhan, M., Bejani, M., Gharavian, D.: Modular neural-SVM scheme for speech emotion recognition using ANOVA feature selection method. Neural Comput. Appl. **23**(1), 215–227 (2013)
37. Shen, P., Changjun, Z., Chen, X.: Automatic speech emotion recognition using support vector machine. In: Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology, vol. 2, pp. 621–625. IEEE (2011)
38. Staudinger, T., Polikar, R.: Analysis of complexity based EEG features for the diagnosis of Alzheimer's disease. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 2033–2036. IEEE (2011)
39. Sun, L., Fu, S., Wang, F.: Decision tree SVM model with fisher feature selection for speech emotion recognition. EURASIP J. Audio Speech Music Process. **2019**(1), 1–14 (2019)
40. Thornton, B.: Audio recognition using mel spectrograms and convolution neural networks (2019)
41. Wang, F., Verhelst, W., Sahli, H.: Relevance vector machine based speech emotion recognition. In: D'Mello, S., Graesser, A., Schuller, B., Martin, J.-C. (eds.) ACII 2011. LNCS, vol. 6975, pp. 111–120. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-24571-8_12
42. Wang, N., Wang, J.S., Zhu, L., Wang, H., Wang, G.: A novel dynamic clustering method by integrating marine predators algorithm and particle swarm optimization algorithm. IEEE Access **9**, 3557–3569 (2020)
43. Yogesh, C., et al.: A new hybrid PSO assisted biogeography-based optimization for emotion and stress recognition from speech signal. Expert Syst. Appl. **69**, 149–158 (2017)
44. Yogesh, C., Hariharan, M., Ngadiran, R., Adom, A.H., Yaacob, S., Polat, K.: Hybrid BBO_PSO and higher order spectral features for emotion and stress recognition from natural speech. Appl. Soft Comput. **56**, 217–232 (2017)
45. Yousri, D., Babu, T.S., Beshr, E., Eteiba, M.B., Allam, D.: A robust strategy based on marine predators algorithm for large scale photovoltaic array reconfiguration to mitigate the partial shading effect on the performance of PV system. IEEE Access **8**, 112407–112426 (2020)
46. Zhang, Z.: Speech feature selection and emotion recognition based on weighted binary cuckoo search. Alex. Eng. J. **60**(1), 1499–1507 (2021)

Machine Learning Algorithms are Used for Fake Review Detection

Wesam Hameed Asaad(✉), Ragheed Allami, and Yossra Hussain Ali

University of Technology, Baghdad, Iraq
cs.20.48@grad.uotechnology.edu.iq, {110046,
Yossra.H.Ali}@uotechnology.edu.iq

Abstract. An internet business's revenue can be significantly impacted by user reviews. Online users read reviews before choosing any products or services. Because of this, a company's profitability and reputation are directly impacted by the reliability of internet reviews. Due to this, some businesses pay spammers to post false reviews. These false reviews abuse the purchase decisions of customers. The approaches for feature extraction currently in use are examined. We will see an injustice in these ratings and reviews. The text analysis method used in this study was sentiment analysis (SA); currently, the area of text analysis attracting the greatest interest. One of the main issues SA is presently experiencing is how to differentiate between negative, neutral, and positive opinion reviews. In this article, We contrast supervised and unsupervised machine learning particularly. Our research demonstrates that supervised machine learning is more accurate and effective than unsupervised learning.

Keywords: Fake review · Opinion Mining · Feature extraction · Machine learning

1 Introduction

Sentiment analysis has developed into a well-liked and interesting area of research recently. It is used to analyze and assess the opinions of various persons [1]. Customers can post their opinions or reviews on various websites in the modern internet era. Those reviews could be helpful to businesses as well as potential customers who wish to learn more regarding services or products before deciding [2]. Consumer reviews have become significantly more prevalent recently, it was reported. Customer reviews affect the decisions of prospective buyers. Put differently, consumers make buying decisions depending on product reviews they read on social media. Consumer reviews, therefore, offer people a vital service. Positive reviews typically yield substantial financial benefits, while negative reviews frequently have the reverse effect [3, 4]. Thus, there is an increasing propensity to depend on customer feedback to change firms by enhancing their services, marketing, and goods [5]. Therefore, customers are becoming more significant in the marketplace. The producer of the Acer laptop has been inspired to make a higher-resolution version of the device after reading reviews that lamented its low

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 292–302, 2024.
https://doi.org/10.1007/978-3-031-56728-5_25

display quality. Due to how individuals freely express and use their remarks, customer review websites have run into issues. On social media (Facebook, Twitter, etc.), everyone can submit criticisms of any firm at any time and with no obligations. Due to a lack of restrictions, a few companies utilize social media to unfairly promote their goods, stores, or brands while disparaging those of their rivals. Imagine that some customers who bought a particular digital camera posted unfavorable image quality reviews.

These reviews negatively impact the public's perception of digital cameras. Consequently, the camera manufacturer can employ a person or organization to fabricate favorable reviews of the camera. Comparable to this, the producer might give the paid individuals instructions to publish negative reviews of the products of competing companies to advance the business. It is considered dishonest to submit reviews for things that you haven't used yourself [6]. Thus, a spammer is someone who submits fake reviews. Spammers who cooperate to achieve a shared goal are indicated as "groups of spammers" [7]. Various investigations have been conducted into the problem of fake review detection. Identifying whether a review is authentic or fraudulent is the primary task in fake review identification. In this study, we apply two machine learning approaches, supervised and unsupervised, for the detection of false reviews. Research Contributions based on the following:

1. Using supervised machine learning and unsupervised, classifying texts into binary classifications (spam and non-spam).
2. Evaluating how well the proposed approach for separating text into non-spam and spam categories works.
3. A comparison of the suggested method with contemporary spam detection techniques

1.1 Problem Statement

Due to the multiple characteristics of text spam, the detection of fraudulent reviews (spam) from textual material using a supervised ML system is a growing issue. The present work addresses the problem of classifying spam (fake reviews) from text using supervised ML. The goal is to create a prediction model that gives a text review's spam class 0, 1, using a set of text reviews (r1, r2, r3, … rn) as input. ri, where 0 represents ham (not spam), and 1 is spam.

2 Related Work

How to infer people's thoughts from text is a topic that is now being studied by several researchers who have been working on algorithms to extract and analyze massive amounts of user-generated data from Twitter and Facebook. This research gathers many studies on the subject as a result.

1 - Elmogy et al. (2021). LR, NB, KNN, and SVM algorithms were used to work on a real data set for restaurants without features generated from user behaviors in order to detect bogus reviews and extract features from reviews [8].

2 - Bansode and Birajdar (2021). Used algorithms to process user evaluations, whether they were positive or negative, that might be used to gauge customer sentiment and neutralize a product when processing hotel ratings [9].

3 - Hussein et al. (2022). This study investigates the use of sentiment analysis to categorize tweets from Twitter users. This strategy can assist in breaking down the material into positive, neutral, and negative viewpoints for analysis. Prior to constructing feature vectors, this data is first pre-processed. The classification techniques were based on ML. The techniques that have been utilized in the research for classifying documents as negative or positive include Naïve Bayes, SVMs, and Maximum Entropy. The dataset for this work uses the Twitter API and consists of subscribing tweets. After pre-processing, ML methods have been used in order to determine if the tweets are negative or positive. [10].

Common terms for phony reviews include "spam reviews", "deceptive opinions", and "spam opinions", and those who create them are sometimes referred to as "spammers". There are three sorts of fake reviews, also referred to as spam opinions:

- Fake review: characterize customers who leave evaluations on products or businesses to promote them or harm their reputation. These reviews are referred to as false or deceitful, and it can be difficult to tell the difference by reading on your own.
- Reviews of a brand exclusively characterize those who are making brand-related comments about the products.
- Non-reviews: those that lack context, express no true perspective, or are merely ads. The final two varieties, disruptive spam opinions, pose little threat and are easily distinguishable by anyone reading them [11, 12].

Enhancing the performance of a pattern recognition system or a ML system is the aim of feature extraction. By decreasing data to its key elements, feature extraction produces more valuable data that may be put into the machine and deep learning models. It is crucial to eliminate any extraneous elements from the data that can reduce the model's accuracy [8]. Numerous techniques for extracting features for fake review detection were explored in the literature. Utilizing textual elements is one such approach [16]. It has a sentiment classification section [17] that depends on the percentage of negative and positive terms in the review, like "weak" and "good" [19].

3 Proposed Model Methodology

In this part, we mimic our proposed spam review detection model, as shown in Fig. 1. Our suggested model is broken down into four phases. Data Preprocessing and. Natural Language Processing (NLP) techniques including word vector. The feature selection process, which includes Word Embeddings (sentence2Vec) approaches, is covered in the third phase for Executing the supervised and unsupervised machine learning.

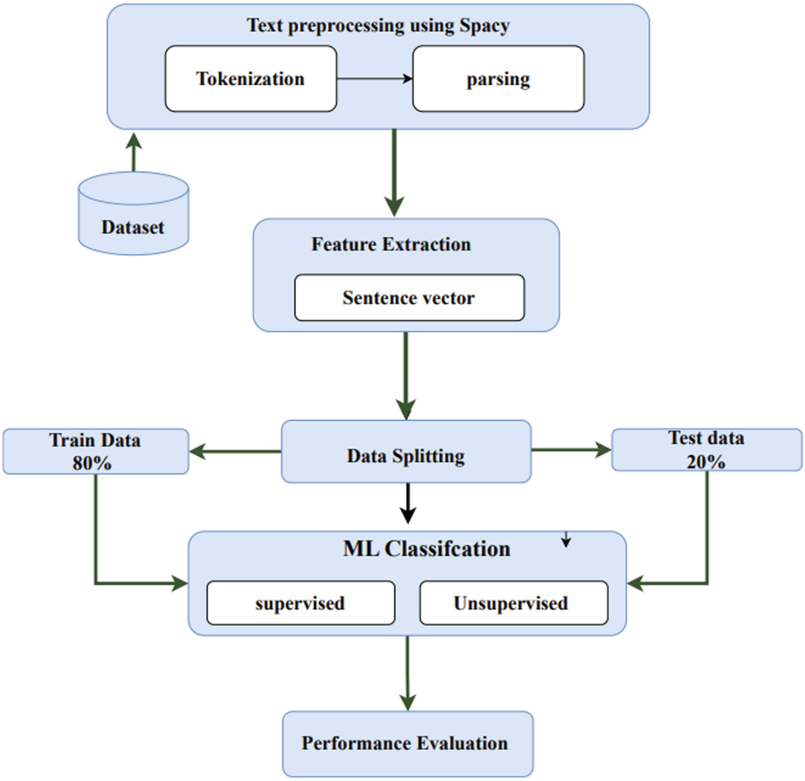


Fig. 1. The proposed Mode.

3.1 Dataset

The dataset is Yalep Dataset, which is publicly available at https://www.kaggle.com/datasets/yelp-dataset/yelp-dataset [9]. The dataset is displayed in Table 1. In our model, the dataset is divided in half, with 20% designated for testing and 80% for training as fake spam and as real, non-spam content.

Table 1. Dataset statistics

Dataset	Description	No. of reviews	Numbers of reviews in labeled classes	
			spam	Not spam
D1	Fake review	355210	36133	319077

Data for training. As was already said, training data makes up 80% of the dataset. The training dataset samples are shown in Table 2.

Table 2. Contains a collection of example entries drawn from training data.

Review	Label Spam (1) and Not Spam (0)
food snack selects popular reek dish appet tra…	1
littl place soho wonder lamb sandwich glass wi…	1
aircondit make much nois hard sleep night	0
backyard hotel total mess happen hotel star	0

The model is tested using the testing dataset once it is trained to utilize the training dataset. Put differently; the classifier receives test data to see how well it works. The entire dataset is divided by 20% for testing.

3.2 Fake Review Detection Model Phases

Phase I: The Preprocessing Phase of our suggested paradigm discusses Data Acquisition and Data Preprocessing. Below is a discussion of each of them in more detail. 1) Data gathering: datasets contain honest and dishonest real-world reviews from the Yelp dataset. They can be utilized as the unlabeled dataset and are open to the public. Yelp Dataset is available at [17]. Data preparation It is necessary to pre-process the classified instances from the Ott Dataset and the unlabeled examples obtained from Yelp.com and afterward labeled. Natural Language Processing (NLP) is used for light preprocessing, including word vectors:

Phase II: Feature Extraction: The suggested model's second stage includes selection. Our proposed model uses sentence vector embeddings. Feature selection for conventional machine learning classifiers uses sentences 2Vec word embeddings procedure. In this instance, feature selection is done using the average of all the word vectors in the text. A method called word embedding converts words into real-valued vectors in a high-dimensional space, where the similarity in meaning between words corresponds to closeness in the vector space. A technique for creating such an embedding is Word2Vec. Grouping related word vectors in vector space is Word2vec's goal and main use. Word2vec generates vectors, which are dispersed numerical representations of word properties like the context of specific words.

Phase III: Proposed Model:

There are two machine learning techniques we can use to identify spam reviews (supervised and unsupervised). Also, traditional machine learning classifiers such as (XGBoost [18] and LightGBM [20]) for supervised and Kmeans [17], minibatch Kmeans [21]) for unsupervised.

Traditional machine learning is applied in our research for the Yelp dataset, and significant results are obtained that outperform the performance of conventional classifiers. In the most straightforward scenario, the entire ML process can be explained for a single review. The review text is initially pre-processed using NLP algorithms. Next, a matrix is used to represent the input texts. The word embeddings (low-dimensional representations) of each row of the matrix are sentence vectors that each represent a single word.

A feature map was employed. Each site does matrix multiplication, and the total output is shown onto the feature map.

4 Proposed System Functionality: Testing and Evaluation

SpaCy is a free open-source natural language processing library in Python programming languages. Used primarily in production software development, spaCy also supports machine learning workflows via PyTorch and TensorFlow statistical models. SpaCy provides fast and accurate parsing, named entity recognition, and easy access to word vectors.

Our experiments have been split into two sections. The experiment I focuses on the experimental outcome of the experiment II, which also covers the experimental outcome of "Yelp Dataset" over-supervised machine learning.

4.1 Data Set Visualization

Figure 2 shows the unbalanced dataset with labeling spam and unspam (0,1). Our work based on a balanced dataset as shown in Fig. 3.

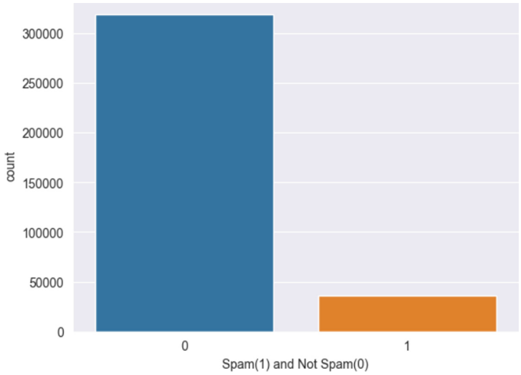


Fig. 2. Unbalanced dataset.

1) Experiment I: In this experiment, various train test ratio and embedding dimension values were used to test the accuracy. With test dimensions 40, 30, 20, and embedding dimensions 60, 70, 80, we have attempted every conceivable scenario.

   In Table 3, we listed the highest accuracy reached by (XGBoost and LightGBM), and in Table 4, we listed the best accuracy achieved with corresponding ratio and dimensions. From Table 3, the highest accuracy achieved for XGBoos is 80.8% with ratio 20:80 embedding.
2) Experiment II: In this experiment, we evaluated the performance of the "Yelp Dataset" using unsupervised ML (kmeans and minibatch K-means) and hidden dimensions with the same values of train test ratio. We have tested every conceivable scenario, similar to Experiment I. Figures 4 and 5 show distribution for the Kmeans cluster and Minibatch Kmeans.

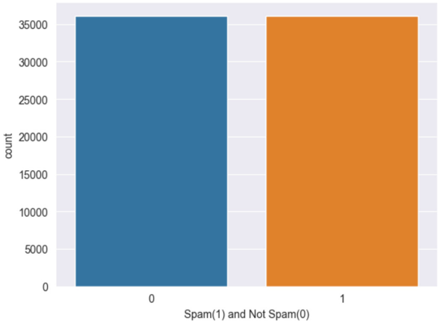


Fig. 3. Balanced dataset.

Table 3. The accuracy of supervised ML (XGBoost,LightGBM)

Algorithm	XGBoost	Light GBM
Accuracy	80.8%	79.7%

Table 4. The accuracy of unsupervised ML (Kmeans, MiniKmenas)

Algorithm	Kmeans	Minibatch kmeans
Accuracy	50.8%	48.7%

5 Comparing the Performance of Classifiers

Additional machine learning classifiers are used to assess the proposed system's classification performance, and the comparative analysis produces a qualitative assessment of the proposed LightGBM and XGboost classifiers for predicting legitimate and fraudulent reviews from textual material. We assess the performance of various classifiers concerning several evaluation metrics, including recall, precision, F1, and accuracy measures, to look at the classification performance outcome. The comparison analysis yields a qualitative evaluation of suggested XGBoost classifier's capability to distinguish between fake and real reviews depending on textual content. Table 5 show the Evaluation of XGBoost and Light GBM and Table 6 show the Evaluation of Kmeans and Minibatch Kmeans.

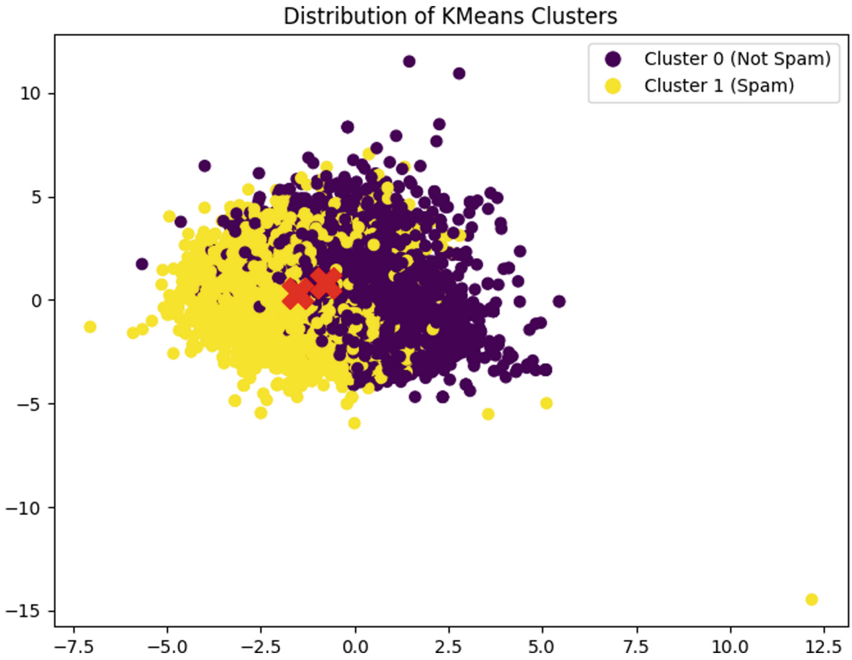


Fig. 4. Distribution of K-means clusters.

6 Conclusion and Future Work

The sentiment analysis of reviews can readily be done using these techniques. The feature vector is then created by combining the features that were retrieved. Reviews are classified using several machine-learning classifiers. Sentiment analysis is necessary to comprehend the emotions sent in any type of communication, including posts, tweets, social media, products, and other topics. Sentiment analysis can be done using machine learning techniques. On the other side, machine learning is easier and more effective, but it needs tagged data.

This work classifies the text into fake (spam) and real (non-spam) evaluations using supervised machine learning techniques like LightGBM, XGboost, and Unsupervised machine learning naming Kmeans and mini Kmeans a specified set of parameters. Prior to submitting the text to the ML classifier, noise is reduced using a variety of preprocessing techniques. Compared to the other research, the model has the highest accuracy, according to the experimental findings of XGboost and LightGBM (80). Since our work suffers from the following limitations:

1. Use limited feature extraction techniques.
2. The random splitting technique divides the data set into testing and training groups. In future research, we strongly advise using hybrid techniques in the features selection phase and cutting-edge machine learning strategies like deep learning techniques for better results.

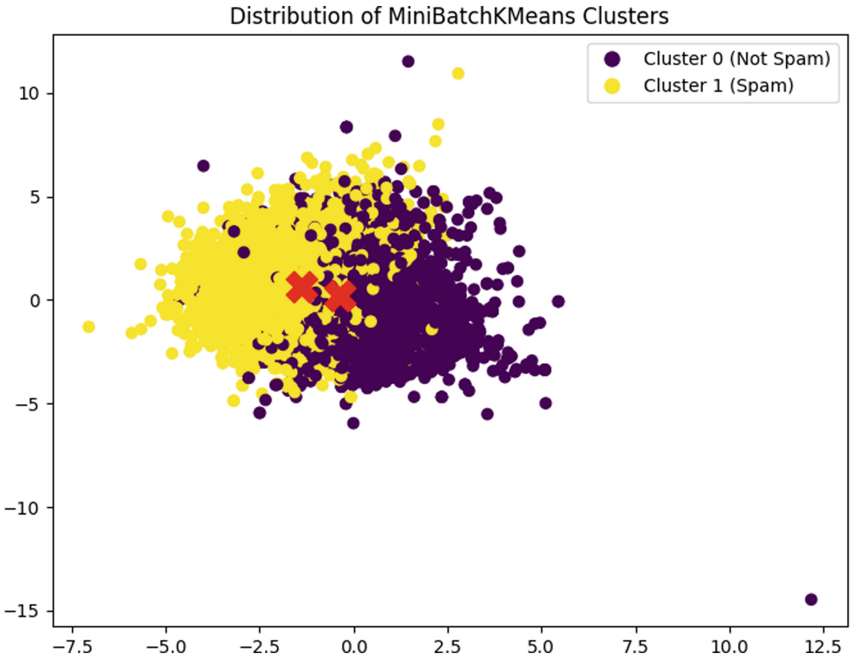


Fig. 5. Distribution of Minibtach K-means clusters.

Table 5. Evaluation of XGBoost and Light GBM.

confusion matrix	precision	recall	f1-score	support
0	%71	%70	%70	7264
1	%70	%71	%70	7190
	The Evaluation of Light GBM			
0	%68	%64	%66	7264
1	%66	%70	%68	7190

Table 6. Evaluation of Kmeans and Minibatch Kmeans.

confusion matrix	precision	recall	f1-score	support
0	%90	%100	%95	95754
1	%20	%00	%1	10809
	The Evaluation of Minibatch Kmeans			
0	%49	%46	%47	7264
1	%48	%51	%50	7190

References

1. Rezapour, M.: Sentiment classification of skewed shoppers' reviews using machine learning techniques, examining the textual features. Eng. Rep. **3**(1), e12280 (2021)
2. Liu, B.: Sentiment analysis and opinion mining. Synthesis Lect. Hum. Lang. Technol. **5**(1), 1–167 (2012)
3. Gillan, S.L., Koch, A., Starks, L.T.: Firms and social responsibility: a review of ESG and CSR research in corporate finance. J. Corp. Finan. **66**, 101889 (2021)
4. Barbado, R., Araque, O., Iglesias, C.A.: A framework for fake review detection in online consumer electronics retailers. Inf. Process. Manage. **56**(4), 1234–1244 (2019)
5. Gupta, R., Pathak, S., Sharma, M., Poornalatha, G.: Feature based opinion mining for mobile reviews. In 2022 First International Conference on Artificial Intelligence Trends and Pattern Recognition (ICAITPR), pp. 1–4. IEEE (2022)
6. Heydari, A., ali Tavakoli, M., Salim, N., Heydari, Z.: Detection of review spam: a survey. Expert Syst. Appl. **42**(7), 3634–3642 (2015)
7. Elmogy, A.M., Tariq, U., Ammar, M., Ibrahim, A.: Fake reviews detection using supervised machine learning. Int. J. Adv. Comput. Sci. Appl. **12**(1), 601–606 (2021)
8. Bansode, M., Birajdar, A.: Fake review prediction and review analysis. Int. J. Innov. Technol. Explor. Eng. **10**(7), 143–151 (2021)
9. Hussein, D.J., Rashad, M.N., Mirza, K.I., Hussein, D.L.: Machine learning approach to sentiment analysis in data mining. Passer J. Basic Appl. Sci. **4**(1), 71–77 (2022)
10. Ren, Y., Ji, D.: Learning to detect deceptive opinion spam: a survey. IEEE Access **7**, 42934–42945 (2019)
11. Krommyda, M., Rigos, A., Bouklas, K., Amditis, A.: An experimental analysis of data annotation methodologies for emotion detection in short text posted on social media. In: Informatics, vol. 8, no. 1, p. 19. MDPI (2021)
12. Rathor, A.S., Agarwal, A., Dimri, P.: Comparative study of machine learning approaches for Amazon reviews. Procedia Comput. Sci. **132**, 1552–1561 (2018)
13. Noori, B.: Classification of customer reviews using machine learning algorithms. Appl. Artif. Intell. **35**(8), 567–588 (2021)
14. Hasan, A., Moin, S., Karim, A., Shamshirband, S.: Machine learning-based sentiment analysis for twitter accounts. Math. Comput. Appl. **23**(1), 11 (2018)
15. Lucini, F.R., Tonetto, L.M., Fogliatto, F.S., Anzanello, M.J.: Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews. J. Air Transp. Manag. **83**, 101760 (2020)
16. Kim Amplayo, R., Brazinskas, A., Suhara, Y., Wang, X., Liu, B.: Beyond opinion mining: Summarizing opinions of customer reviews. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 3447–3450 (2022)

17. Alagrash, Y., Mohan, N., Gollapalli, S.R., Rrushi, J.: Machine learning and recognition of user tasks for malware detection. In: 2019 First IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA), pp. 73–81. IEEE (2019)
18. Li, J., et al.: Application of XGBoost algorithm in the optimization of pollutant concentration. Atmos. Res. **276**, 106238 (2022)
19. Alabdullah, A.A., Iqbal, M., Zahid, M., Khan, K., Amin, M.N., Jalal, F.E.: Prediction of rapid chloride penetration resistance of metakaolin based high strength concrete using light GBM and XGBoost models by incorporating SHAP analysis. Constr. Build. Mater. **345**, 128296 (2022)
20. Hicks, S.C., Liu, R., Ni, Y., Purdom, E., Risso, D.: Mbkmeans: fast clustering for single cell data using mini-batch k-means. PLoS Comput. Biol. **17**(1), e1008625 (2021)
21. Zhu, X., Sun, J., He, Z., Jiang, J., Wang, Z.: Staleness-reduction mini-batch *K*-means. IEEE Trans. Neural Netw. Learn. Syst. (2023)

Development and Research of Models for Optimization Information Flow in Interactive Analysis Big Data in Geographic Information Systems

Ali Abdulkarem Habib Alrammahi[1](✉), Farah Abbas Obaid Sari[1],
and Bushra Kamil Hilal[2]

[1] Department of Computer Sciences, Faculty of Computer Science and Mathematics, University of Kufa, Najaf, Iraq
{alia.alramahi,faraha.altaee}@uokufa.edu.iq

[2] Department of Computer Information Systems, Faculty of Computer Science and Information Technology, University of Qadisiya, Qadisiya, Iraq
bushra.k.h@qu.edu.iq

Abstract. A large number of areas of application of geographic information systems (GIS) involves the continuous accumulation data. The need to record the state of an observed object or phenomenon generates an intense flow of heterogeneous data over long-time intervals. The period of storage of the received information in many cases is not limited and is limited solely technical capabilities of archiving information. For this reason, GIS databases correspond to modern ideas about big data (big data), since they have a significant volume that is constantly growing, are heterogeneous in presentation formats, and have a value determined by a reliable display of objects, phenomena and events of the real world.

A new model of representation of the analysis workspace by two informational components, skeleton and environment, is proposed. The skeleton is formed by deterministic queries to the GIS database, the environment is built by intelligent procedures as a shell of the skeleton. The difference of the proposed concept is the use of a utility function, the parameters of which are the complexity of the skeleton and the environment. The application of the proposed concept allows optimizing the quality of decision-making by maximizing the utility function of the workspace.

Keywords: Skeleton Information · Environment Information · GIS database · Analytical information model

1 Introduction

1.1 Related Background

Modern GIS systems are equipped with a variety of software tools for statistical, spatial, topological, morphological, cartographic and other types of analysis, which make it possible to build new cartographic objects by placing them on the original cartographic

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 303–317, 2024.
https://doi.org/10.1007/978-3-031-56728-5_26

base. Therefore, the used parser forms an add-on independently data stream capable of improving the quality of decisions made. However, in practice, quality improvement does not always occur. The reason is that as the total volume of GIS data increases [1, 2], the complexity of selecting and arranging cartographic objects in the workspace increases.

Cartographic images become more complex, the volume of visual information increases, and the glut of user-analyst information flows from the GIS reduces the quality of perception and, as a result, worsens the value of solutions to applied problems. Therefore, special care must be taken to adapt the GIS to the behavior of the user-analyst in order to reduce the complexity of the workspace under study and ensure its integrity. A promising way to solve this problem is the development of new models for interactive analysis based on knowledge [3–5]. As the study of scientific publications has shown, this issue remains insufficiently studied. Government research programs in the field of science and technology emphasize the importance of research in the field of big data processing. This direction is appropriate from a scientific and practical point of view, as evidenced by a large number of local and scientific publications [6–8].

1.2 Aim of Study

It consists in the development and research of information models to improve the information process in the interactive analysis of big data of a geospatial nature. Based on the developed models of the analysis workspace, actions to improve the quality of information.

Decision support based on spatial data. To achieve this goal, the following tasks are being solved:

1. Analysis of modern means of storing and processing big data, and models for organizing dialogue in geographic information systems that use big data.
2. Develop and research a new concept of workspace representation in interactive analysis, explicitly using the iteration-dependent utility function.
3. Development and research of an adaptive model of interactive dialogue based on knowledge arising in the process of collective work of users with GIS.
4. An experimental study of the proposed models for improving information flow in the interactive analysis of large geospatial data.

2 Methodology

2.1 The Study Problem

Is devoted to a review of existing methods and means interactive analysis of big GIS data. Models of representation and visualization of data in geoinformation systems are considered [9]. The main cause of visualization problems in GIS is redundancy.

The information base of real systems contains much more information than is required to solve the applied problem [10]. Visualization of all available data in a given area of space is meaningless, since human perception of the flow of visual information has a natural limit. Because of this, the analyst does not receive necessary information and is forced to take measures to reduce redundancy [11, 12].

$$G = <S, T, C>,$$

$$V = F(G),$$

$$|V| < V^{*},$$

where G is a three-component model of the GIS information base, which includes a set of cartographic (C), temporal (T) and semantic objects (C). The render operator (()) generates an image (V) of limited complexity ($V*$). As follows from this model, the selection of the necessary data from G is not included in the GIS function, i.e. implemented manually by the user.

The static visualization model looks like:

$$V = F_1(<S, T>) \cup F_2(C)$$

$$|V| < V^{*}, R(V) < R^{*}.$$

Unlike the previous model, here two independent visualization operators for spatiotemporal and semantic data $F1$ and $F2$ are used. A set of variously rendered objects form a cartographic image. Image fidelity (V) in the case under consideration plays an important role, since any project of the cartographic area of analysis is based on the need to reliably represent the picture of the real world. All manipulations with info base objects are performed by the user manually.

Visualization model for geoinformation mapping:

$$S = \bigcup_i L_{si}, T = \bigcup_i L_{ti}, C = \bigcup_i L_{ci},$$

$$\omega R = <sR, tR, cR>, sR \subseteq S, tR \subseteq T, cR \subseteq C,$$

$$V = F(\varphi(<S, T, C>))$$

$$|V| < V^{*}, R(V) < R^{*}.$$

Here Φ is the operator for selecting objects corresponding to the task, F is the operator for visualizing objects. The operator Φ essentially interprets the description of the mapping project, i.e. implements part of the useful data selection functions.

The visualization model in interactive analysis is represented as

$$S = \bigcup_i L_{si}, T = \bigcup_i L_{ti}, C = \bigcup_i L_{ci},$$

$$\omega R = <sR, tR, cR>, sR \subseteq S, tR \subseteq T, cR \subseteq C,$$

$$Z = zk,$$

$$V = F(\varphi(<S, T, C>, zk))$$

$$|V| < V^*, R(V) < R^*.$$

$$I(V, zk) \rightarrow max.$$

Here (V, k) is the utility function of the sequence of cartographic images generated to solve the applied problem.

The characteristics of big GIS data are considered, which in the traditional sense reflect the problems and opportunities for processing large amounts of data: volume (volume), speed (velocity) and diversity (variety), reliability (veracity) to describe the integrity and quality data. Additional characteristics are analyzed, such as variability, validity, volatility, visibility, value, and visualization. It is concluded that any multiparametric assessment of geodata classifies them as big data [13].

An analysis of approaches to building analytical systems for big data led to the conclusion that hardware and software support includes a significant proportion of universal components that bear the main functional load for access and transmission. Cartographic data. In fact, more intensive information flows are generated, but the problems of effectively presenting flow elements for visual analysis are not solved. Since the ultimate goal of using geospatial big data is to be in useful content, the task of constructing "intelligent" filters of information flows arises. Their purpose is to select the necessary information based on knowledge. It is concluded that the latter problem remains little studied in relation to interactive analysis for the purpose of decision making. The goals and objectives of adapting the GIS information base to the course of interactive analysis are studied [14, 15]. It is noted that as the amount of data grows, direct manipulation of the workspace becomes less and less effective and the role of dialogue management grows. The subsystem of the dialogue with the user must adapt the workspace in order to create the user has a holistic mental image of the task. The first task of adaptation is to set the necessary boundaries for the flow of cartographic information to the user who solves the applied problem. The second task of adaptation is related to the use of useful information in the context of modifying the GIS information structure. This task is also a consequence of the use of big geodata. The essence of the problem making the best use of map data is to reuse knowledge about the important details of space and time: instances and classes of features and relationships. This is hindered by the size of the information base, which makes it difficult not only to search of these elements, but also a lack of knowledge about them. In the latter case, hidden dependencies between data are implied, which are discovered as a result of the collective use of geoinformation services. Knowledge should not only be discovered as a result of collective work, but also spread in online communities.

2.2 Proposed Method

This section is dedicated to the development of an information visualization model during interactive analysis of the workspace.

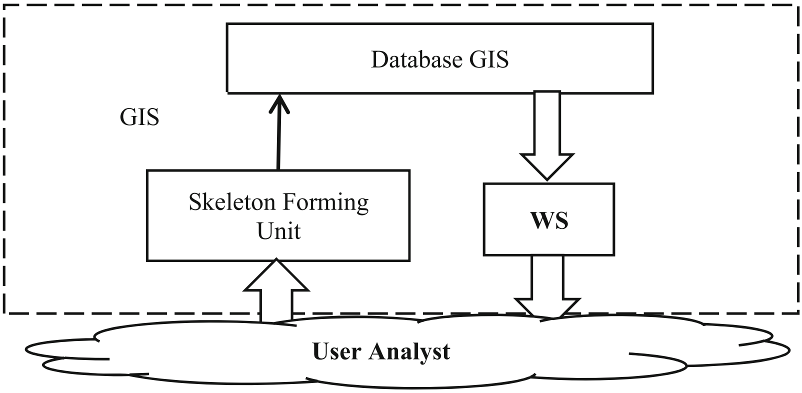


Fig. 1. GIS Information Model with a Workspace.

On Fig. 1 shows a GIS information model illustrating the information flow whose optimization is the subject of this paper. The workspace (WS) is the main an information element that allows the user-analyst to solve applied problems. Information flow results from execution of a sequence of queries to the GIS database (DBGIS). Represented by the workspace, the stream essentially carries cartographic data to the analyst user. The flow intensity (t) is variable and is determined by the speed of perception of the cartographic information and its visual analysis by the user. It should be noted that the only tool for optimizing the quality of the information flow is the manual formation of queries to the GIS database.

A generalized algorithm for interactive search for a solution based on geospatial data is given (Fig. 2):

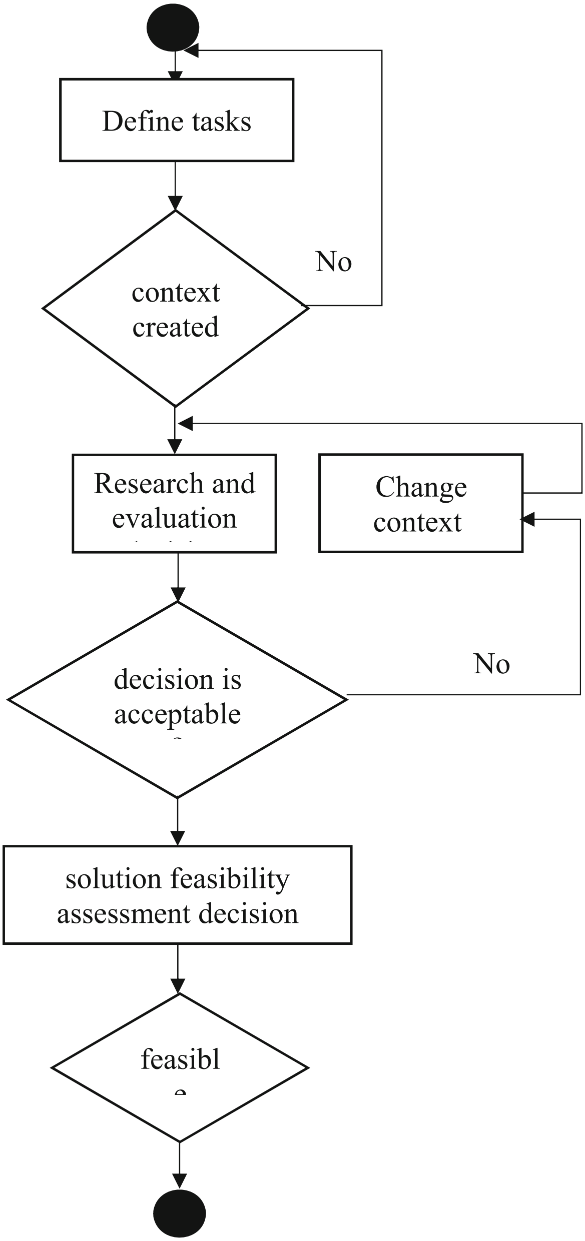


Fig. 2. General solution search algorithm.

According to the algorithm, the statement of the problem is formulated, which should be implemented by the GIS visualization subsystem:

$$\left\{\begin{array}{c} I(\omega_i) \rightarrow max, \\ \omega_j = K\Big(\Omega, c_m, \omega_j', \omega_j - 1, \omega_j - 2, \ldots, \omega_0\Big), \\ \omega_j' = q_i\big(\omega_{j-1}, \Omega\big), \\ R\big(\omega_j\big) < R^*, \\ \omega_j, \omega_j' \subset \Omega \\ j = \overline{0, |j|}, q_i, \in Q, c_m \in C. \end{array}\right. \tag{1}$$

Through $\Omega = \omega_1, \omega_2, \ldots \omega_n$ indicates the set of features available to the GIS from external data sources. The set $Q = q_i$ is a set of queries, each of which either changes the set of workspace objects (ω) or changes its visual representation:

$$\omega_j' = q_i\big(\omega_{j-1}, \Omega\big),$$

where ω' is the modified workspace instance. $C = c_k$ denotes the set of contexts in which semantic invariants are supported. I (ω) denotes the workspace utility function w.

The value of the function is greater, the higher the level of professional perception of the workspace by the user. R (ω) denotes the computing resources required to visualize the workspace w by the user equipment. A measure of the volume of resources can be considered the number of cartographic elements of the workspace. Since the workspace after the execution of each request $q_i \in Q$ changes its state, through $\omega_{\ j}, j = 0{,}1{,}2\ldots$ $|J|$ the sequence of these states is indicated. Index j plays the role of discrete time in a visual analysis session. Then the sequence of states w_j is discrete process in which the current state is a consequence of all previous ones. Thus, a GIS should provide maximum usefulness for representing the workspace in each of its states.

K is an operator for mapping an intermediate state of the workspace in a given context to a new state based on available GIS data and the current history of visual analysis. An intermediate state is the one into which the workspace passes after the execution of the next user request. Taking into account the features of interactive interaction, it was concluded that the best perception can be achieved under the following conditions:

- the number of workspace objects is close to N^*.
- workspace has minimal redundancy.
- the workspace satisfies the integrity conditions.

Then the piecewise linear utility function of visualization control for $j = \overline{a, |J|}$ can be set as follows:

$$I\big(\omega_j\big) = \begin{cases} 1 - (N^* - |wj|)/N^*, & |\omega_j| < N^*, \\ 1, & \big|\omega_j\big| = N^* \\ 1 - \frac{|\omega_j| - N*}{|\omega_{complex}|}, & \big|\omega_j\big| > N^*\omega_{complex} \end{cases} \tag{2}$$

It can be seen that the maximum utility value $I\big(\omega_j\big) = 1$ is reached at the point N^*, which is the only maximum point. Deviation from N^* up or down leads to a decrease in utility and $\rightarrow I\big(\omega_j\big)$ 0.

To implement the display operator, a representation is introduced
$\omega = B \cup E$
$B \subseteq \Omega$: $\forall_{\omega i} \in B \Rightarrow \exists_{qj}$ $(X_S, X_T, X_C, X_E) = true, j = 1,Q, i = 1,|m_W|$
$B \cap E = \emptyset, E \subseteq \Omega,$

Where B is the set of cartographic objects selected by queries Q to the info base Ω. Let's call the set B a skeleton. The skeleton is formed by objects that have been explicitly requested by the user through the GIS dialog menu system. The E set is a skeleton environment designed to improve the perception of the workspace. The objects of the environment were not explicitly requested by the user, however, they represent important features of the area that affect the assessment of the situation. The mapping is implemented by the operator K and takes the form.

$$E = \overline{K}(B, c_m) \tag{3}$$

Taking into account (3) and (4), problem (1) is formulated as follows:

$$\begin{cases} I(w) \to max \\ w = B \cup E \\ B = \cup_j qj^{(X_s, X_T, X_c, X_E)}, \\ E = \overline{K}(B, c_m), \\ |w| < |w_{complex|}, \\ qj \in Q, c_m \in C \end{cases} \tag{4}$$

In accordance with (4), determining the environment (E) for a given skeleton (B) according to the context $cm \in C$ becomes the main computational operation in the dialog control problem. The environment E is formed by applying expert rules (B, Ω) for constructing a visual image for a given skeleton B.

$$\omega_i \in \boldsymbol{E} \Rightarrow \mathrm{K}(\mathrm{B}, \Omega) = true, i = \overline{1, |mw|}.$$

The expert rules K (B, Ω) represent knowledge about how the workspace boundaries (w) are constructed for a given skeleton B and how the resulting area is filled with environment primitives E. The paper presents the structure of rules corresponding to one of two classes:

1. rules for determining the boundaries of the work area;
2. rules for selection of environment primitives.

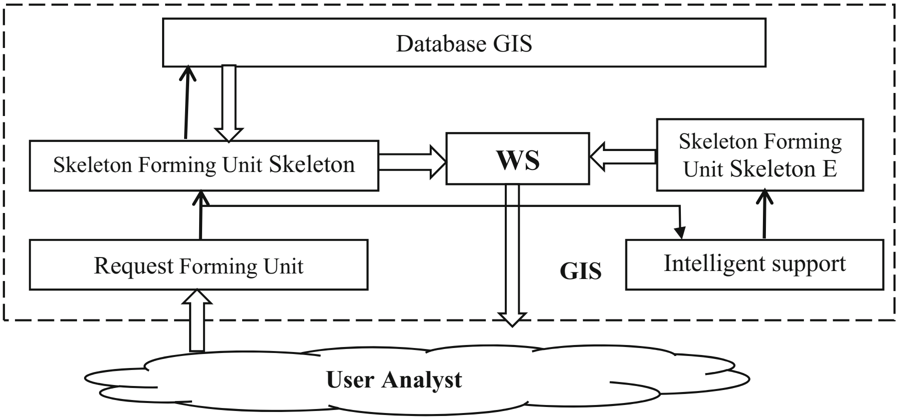


Fig. 3. GIS information model using the new model

On Fig. 3 shows the information structure of a GIS using the representation of the WS in the form of a skeleton and an environment. It can be seen that, thanks to the introduced components, the information flow is decomposed.

into two components. The skeleton formation block (*B*) is a GIS element designed to form queries in the GIS database that generate an information flow of the most significant data.

The environment generation block (*E*) generates a data stream that provides a certain semantic content of the skeleton. Requests to the GIS database for skeleton objects are built by the intellectual support block based on knowledge about the usefulness of the generated cartographic images. By introducing the WS representation by the skeleton and the environment, it is thus possible to split the information flow and, acting on the environment, maximize the usefulness of the workspace as a whole.

Next, the representation of the workspace with indeterminate elements is examined. The concept of u-objects has been introduced, which generate defects in the visual display of the workspace. A workspace with u-objects can be viewed as a set of objects with the following structure:

$$\omega = B \cup E \cup B^{+} \cup B^{-} \tag{5}$$

where $B+$ is the set of u-objects that do not generate image defects, the set $B-$ are u-objects that cause defects. Analysis (5) allows us to conclude the following. First, the use of u-object does not change the nature of the rendering control. As before, perception is limited by the number of visualized objects, and the semantic content is provided by the selection of the environment. Concept of objects, representing defects, makes it possible to extend the visualization control method to maps containing objects with uncertainty.

Table 1. Workspace views with u-objects

VIEW SKELETON (B^*)	VIEW ENVIRONMENT (E^*)	ASSIGNING A VIEW
$B \cup B- \cup B+$	E	Generation of alternative solutions
$B \cup B+$	E	Risk assessment
$B- \cup B+$	E	Assessment of the quality of additionally involved data
B	$B- \cup B+ \cup B$	Strengthening the semantic details of the context

Secondly, representation (5) creates a new opportunity to diversify visual analysis by changing the "semantic perspective" of the display. The essence of such an operation is as follows: the sets from (5) can be combined into the skeleton and the environment in different ways, which gives rise to visual images of various semantic orientations. Workspace representations with u-objects are summarized in Table 1.

3 Results and Analysis

A user classification was implemented based on historical structured data by applying a machine learning algorithm and forming a "stereotypical" skeleton [16, 17] (Fig. 4).

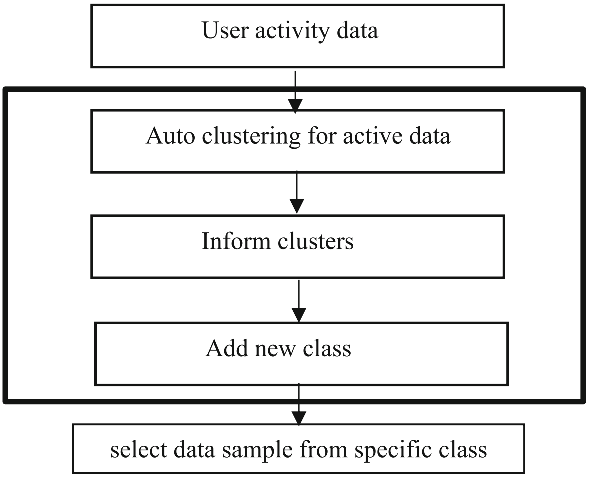

User activity data
Auto clustering for active data
Inform clusters
Add new class
select data sample from specific class

Fig. 4. Data scaling algorithm for adaptive service from an applied point of view

To assess the redundancy of the flow of cartographic information, a probabilistic model for selecting map fragments into the analysis workspace was used [18–20]. The essence of the model is as follows: it is known that maps are stored as rectangular components. The minimum amount of data that can be obtained from a GIS request cannot be less than the size of one component. As the analysis showed, a well-perceived working area approximately corresponds to this size, i.e. the number of objects. Because the real workspace spans multiple tiles, there is redundancy. Therefore, it is of interest how fast the number of tiles grows with an increase in the area of analysis. To do this, we used a probabilistic model of random coverage of lattice cells of a closed curve of an arbitrary shape, known from integral geometry. In particular, this relation is for estimating the upper bound on the number of cells covered by a random closed curve of known length. The lattice consists of regions bounded by closed piecewise smooth curves. The expression for the average number of covered cells is:

$$E(v) = \frac{2\pi(a_0 + F_1) + L_0 L_1}{2\pi a_0} \tag{6}$$

where v is the number of cells covered by the area randomly thrown onto the grid, a_0 is the area of the grid cell, L_0 is its perimeter, $F1$ is the area of the randomly generated area, 1 is the perimeter of this area.

Table 2. Redundancy coefficients and utility estimates for the "Administrator"

Redundancy coefficient (*k*), Utility score (*R*)																
User1	2,1	1,5	2,5	1,9	2,0	1,8	1,6	2,3	1,8	1,8	1,8	1,5	1,5	1,8	2,0	2,3
	6	6	7	6	7	5	6	8	5	6	5	4	4	6	6	8
User2	1,2	1,0	2,8	2,1	4,0	1,2	2,5	1,3	2,4	2,5	1,5	1,9	3,0	1,9	1,5	1,1
	4	4	6	7	7	4	7	5	7	7	5	6	8	6	6	5
User3	2,4	1,5	2,3	1,7	2,4	2,4	1,5	1,5	2,1	1,5	2,2	1,4	1,4	2,0	2,2	1,0
	8	4	9	4	8	8	4	3	6	4	6	2	2	6	7	5
User4	2,0	2,0	2,0	1,0	1,0	2,5	1,4	1,0	1,0	2,0	2,0	1,9	1,9	1,9	2,3	2,2
	5	6	6	2	5	8	3	3	4	6	7	8	6	5	7	7

Table 2, 3, 4 show experimentally obtained samples of values of the redundancy coefficient (*k*) and expert estimates of utility (*R*). Usefulness ratings are obtained by a user survey. The rating scale was 10-point.

Table 3. Redundancy factors and utility estimates for the context "Security Department"

Redundancy coefficient (*k*), Utility score (*R*)																
User1	1,8	1.6	2.2	2.5	2.3	2.0	2.3	2.5	1.8	1.8	2.4	2.0	2.2	2.4	1.8	2.1
	8	7	8	9	7	6	8	10	5	4	9	4	9	9	8	9
User2	2.3	2.5	1.8	1.8	1.6	2.2	1.9	2.0	2.2	2.5	2.3	1.8	2.6	1.8	1.9	2.3
	10	9	5	6	6	9	6	6	6	9	8	7	10	7	7	9

Table 4. Redundancy ratios and utility estimates for the context "Energy Supply Department"

Redundancy coefficient (*k*), Utility score (*R*)																
User1	2,2	3,3	1,0	6,5	3,5	2,0	4,0	4,8	3,0	3,4	1,0	3,1	3,5	5,1	3,1	1,0
	6	6	6	8	5	5	8	7	6	6	5	7	6	7	4	5
User2	3,0	5,5	3,1	4,3	3,3	1,0	3,1	3,4	1,0	5,2	1,0	3,3	3,0	5,6	1,5	1,6
	6	10	7	9	6	6	5	7	5	8	4	7	7	9	4	5
User3	3,4	3,3	4,1	3,1	3,4	3,7	4,0	3,7	3,1	3,4	3,4	3,7	3,7	3,1	3,4	3,5
	7	6	10	6	6	6	9	6	6	9	6	7	7	6	8	8
User4	3,1	4,0	3,1	6,0	6,0	3,4	1,0	1,0	1,0	3,0	1,0	3,1	7,2	4,9	2,8	5,6
	6	8	6	9	9	6	5	6	5	6	6	6	10	9	6	10
User5	5,0	6,4	1,0	1,0	1,0	4,0	3,1	5,0	3,3	6,2	2,0	2,0	3,5	3,3	3,4	3,5
	10	10	5	5	5	8	7	10	6	10	7	7	8	8	7	8
User6	5,0	1,0	1,0	3,3	3,4	3,1	1,0	1,0	2,8	3,5	5,6	1,0	1,0	5,6	5,2	6,5
	8	5	5	6	6	6	5	5	5	5	8	5	5	9	8	9

In Table 5 shows the names of the contexts and the average values of the redundancy reduction coefficients for each of the contexts. Analyzing the obtained data, it should be noted that the use of the visualization control subsystem ensured the reduction of redundancy in all contexts. The difference in values testifies to the different quality of knowledge embedded in the experimental sample of the system.

Table 5. Names of contexts and average scores

Name context	Administrator	Security Service	Department of Energy Supply	Department of Engineering communications	Logistics Department
K	1,9	2,1	3,3	4	10
R	6	8	7	9	8

Table 6 shows the changes in various system parameters as a percentage relative to the initial metrics before and after the introduction of the intelligent assistance system with the adaptation system. The options are:

Table 6. Changes in parameters before and after the implementation of the geoservices intelligent assistance system in percent.

	Requests per second (less is better)	Satisfaction geoservices, accuracy adaptation (more - better)	Time user's work with geoservices (less - better)	Load on the DB CPU geoservice (less better)	Accuracy classification user system (more - better)
Before the implementation of the system intellectual support	100%	Unknown	100%	100%	Not accepted
After the introduction of rhenium systems intellectual support	73,3% (−26,7%)	81,7%	89% (−11%)	92% (−8%)	94,5%

4 Conclusion

In the context of the paper that was completed, the following scientific and practical conclusions were obtained:

1. Classification methods for analyzing big data by means of geographic information systems, the information base of which belongs to the class of big data. The classification differs in that it is based on the principle of technical implementation of the process of analyzing the accumulated data. Use of classification allows you to develop a rational strategy for improving the quality of information support for decision-making.
2. The concept of representing the analysis workspace with two components - the skeleton and the environment. The skeleton is formed by deterministic queries to the GIS database, the environment is built by intelligent procedures as a shell of the skeleton. The difference of the proposed concept is the use of a utility function, the parameters of which are the complexity skeleton and environment. The application of the proposed concept allows optimizing the quality of decision making by maximizing the utility function of the workspace.
3. A method for adapting the user's dialogue with the GIS, subject to changes in the structure of its information base. The method is based on extracting knowledge about objects useful for visual analysis by generating rules. Rules are considered as hypotheses that can be confirmed or refuted during specified time by members of a professional group of analysts who use the service. The application of the proposed method will reduce the redundancy of the analysis workspace and increase its integrity.

4. Results of experimental analysis of the effectiveness of the proposed methods. A distinctive feature of the results is the use of modern modeling software based on machine learning algorithms. The results of processing experimental data showed that the efficiency dialogue interaction increased by 10–50%.

References

1. Thornton, L.E., Pearce, J.R., Kavanagh, A.M.: Using geographic information systems (GIS) to assess the role of the built environment in influencing obesity: a glossary. Int. J. Behav. Nutr. Phys. Act. **8**(71) (2011)
2. Research on the application of geographic information system in tourism management. Procedia Environ. Sci. **12**(B), 1104–1109 (2012)
3. Wei, H., Qing-xin, X., Tang, X.-S.: A knowledge-based problem solving method in GIS application. Knowl.-Based Syst. **24**(4), 542–553 (2011)
4. Paul, K., Jha, V.C.: Paradigm shifts in geographical research and geospatial applications. Sociedade & Natureza **33** (2021)
5. Senouci, R., Taibi, N.-E., Teodoro, A.C., Duarte, L., Mansour, H., Yahia Meddah, R.: GIS-based expert knowledge for landslide susceptibility mapping (LSM): case of Mostaganem Coast District ,West of Algeria. Sustainability **13**(2), 630 (2021)
6. Müller, O., Junglas, I., vom Brocke, J., Debortoli, S.: Utilizing big data analytics for information systems research: challenges, promises and guidelines. Eur. J. Inf. Syst. **25**(4), 289–302 (2016)
7. Li, S., Yang, H., Huang, Y., Zhou, Q.: Geo-spatial big data storage based on NoSQL database. Sci. Wuhan Univ. **42**(2), 163–169 (2017)
8. Dempsey, C.: Where is the phrase "80% of data is geographic" (2012). https://www.gislounge.com/80-percent-data-is-geographic/
9. Kakkar, D., Lewis, B., Guan, W.: Interactive analysis of big geospatial data with high-performance computing: a case study of partisan segregation in the United States, vol. 26 (2022)
10. Ma, M., Wu, Y., Chen, L., Li, J., Jing, N.: Interactive and online buffer-overlay analytics of large-scale spatial data. ISPRS Int. J. Geo Inf. **8**(1), 21 (2019)
11. Cybulski, P., Medyńska-Gulij, B.: Cartographic redundancy in reducing change blindness in detecting extreme values in spatio-temporal maps. ISPRS Int. J. Geo Inf. **7**(1), 8 (2018)
12. Vicentiy, A.V., Shishaev, M.G.: Reducing digital geographic images to solve problems of regional management information support. In: Silhavy, R. (ed.) CSOC 2020. AISC, vol. 1225, pp. 461–469. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-51971-1_38
13. Shen, Z., Li, M.: Big data support of urban planning and management, China: the experience (2018)
14. Bill, R., et al.: Geospatial information research: state of the art, case studies and future perspectives. PFG – J. Photogram. Remote Sens. Geoinf. Sci. **90**, 349–389 (2022)
15. Sternad Zabukovšek, S., Tominc, P., Deželak, Z., Nalbandyan, G., Bobek, S.: Acceptance of GIS within ERP system: research study in higher education. ISPRS Int. J. Geo-Inf. **11**(83) (2022)
16. Upreti, A.: Machine learning application in G.I.S. and remote sensing: an overview. Preprints (2022)
17. Sari, F.A.O., et al.: Networks cyber security model by using machine learning techniques. Int. J. Intell. Syst. Appl. Eng. **10**(1), 257–263 (2022)
18. Jansuwan, S., Chen, A., Xu, X., Yang, C.: Assessing redundancy of freight transportation networks. In: Annual Meeting of the Transportation Research Board, Washington (2013)

19. Chan, J., Teknomo, K.: Road reconstruction and redundancy analysis on the road network: a case study of the ateneo de manila university network. Procedia Soc. Behav. Sci. **218**, 56–75 (2016)
20. Zhang, Z., Erqi, X., Zhang, H.: Complex network and redundancy analysis of spatial–temporal dynamic changes and driving forces behind changes in oases within the Tarim Basin in northwestern China. CATENA **201**, 105–216 (2021)

Comparison of Text Summarization Methods in Turkish Texts

Semih Marangoz[1](✉) and Ahmet Sayar[2]

[1] School of Computer Engineering, University of Kocaeli University, Kocaeli, Turkey
semihmgoz@gmail.com

[2] Department of Computer Engineering, University of Kocaeli University, Kocaeli, Turkey
ahmet.sayar@kocaeli.edu.tr

Abstract. In this study focusing on extractive automatic text summarization, popular text summarization algorithms commonly used in other languages were compared. Due to their suffix-based structure, the impact of these algorithms on Turkish may not be as effective as English and Chinese, which have been extensively studied. Accordingly, the most commonly used extractive text summarization approaches were investigated, and some of them were tested and compared on Turkish texts. In line with the study, five summaries were generated using the TextRank, LexRank, Luhn algorithms, and two word frequency-based summarization algorithms that we developed, based on a dataset of 130 news texts summarized by three individuals. The similarity metrics were calculated using the Rouge Metric algorithm by comparing the output summaries with the reference summaries. The selected summarization algorithms were chosen among the most commonly used extractive text summarization algorithms, and they are all extractive text summarization algorithms. As a result of the comparison, it was observed that the algorithm developed based on sentence selection using the frequency of word stems had the highest similarity value. The study's outcome will involve the identification of the most suitable automatic summarization algorithm for Turkish. In this context, conclusions can be drawn regarding the applicability of various methods, the potential for achieving more advantageous results when approached from specific angles, and the aspects requiring reinforcement. This way, the aim is to facilitate the attainment of proficient outcomes in Turkish-specific summarization, thus ensuring a professional culmination.

Keywords: NLP · Natural Language Processing · Turkish Text Summarization · TextRank · LexRank · Luhn

1 Introduction

In today's time, which is highly valuable, although accessing information has become easier, obtaining the necessary information still takes a considerable amount of time. The increasing amount of data in online and print sources, coupled with its rapid growth, highlights the necessity of text summarization in terms of both time and performance efficiency. As this trend continues, it becomes increasingly difficult to determine whether the

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 318–332, 2024.
https://doi.org/10.1007/978-3-031-56728-5_27

increasing information is clean data, requiring the need to spend a significant amount of time scanning different sources for verification. In this regard, text summarization techniques geared towards fast data analysis have been developed. These techniques require initially working on the basic structure of the language to establish rules and focus on the building blocks that ensure semantic coherence. Since each language contains numerous similar or independent rules, language-based studies necessitate specific work for each language. These studies encompass extensive research conducted from the 1950s to the present. While the conducted studies primarily focused on the two most widely spoken languages worldwide, Chinese and English, even for these languages, achieving summarizations on par with human translation level is not possible. Although a 100% accuracy rate has not been achieved, an examination of the scanned sources reveals that the progress of technology, increased research efforts, and advancements in artificial intelligence and deep learning have significantly improved the rate of accuracy compared to the past. The fact that the developed systems mostly focus on only a few languages increases the possibility of other languages lagging behind in this technology. This study aimed to determine the degree of success of inferential summarization models developed for other languages in the Turkish language. Additionally, aims to examine recently shared works and provide guidance on which system is more suitable for ensuring that the Turkish language does not lag behind in automatic text summarization, following the necessary efforts to adapt to Turkish, and identifying areas that need to be strengthened and focused on. In the study, inference-based text summarization algorithms were tested on a dataset with reference summaries, and detailed comparisons of the results were conducted to analyze and make observations. The dataset consists of 130 news texts, and three separate summaries created by individuals for these news texts are used as references. The most commonly used text summarization algorithms in the literature were determined by conducting a detailed search in the field of extractive text summarization when selecting the algorithms for comparison. The examined studies also reveal that each work has taken automatic text summarization one step further, and with the passage of time, each developed method has contributed to improving accuracy [1]. In the early stages, studies were conducted within a single document, which later progressed with the use of multiple documents [2], and then domain-specific summarization approaches [3] further increased the accuracy rate. In addition, it was observed that in query-based techniques [4], the accuracy of the summary extracted without a query was higher compared to the summary extracted with a query. Research also indicated that abstractive text summarization techniques yielded more successful results compared to extractive techniques [1, 5–7]. However, the abstractive technique requires a more complex development method compared to the extractive technique. Instead of performing these tasks with a single system, hybrid systems can be used to generate closer summaries. Furthermore, utilizing auxiliary tools to clean dirty data also yields beneficial results. In addition to the frequently used summarization algorithms mentioned in the literature related to the study, an algorithm was developed based on word frequency value, and two different outputs were produced using this algorithm. The system was fed with 130 news texts, resulting in the generation of 650 summary texts. However, before generating the outputs, erroneous words in the news texts and summaries were edited, and the cleaning of dirty data was ensured. Subsequently, these generated summary texts were compared with 390 reference summaries using the Rouge (Recall-Oriented Understudy for Gisting Evaluation) Metric algorithm. Among the Rouge outputs, a comparison was made with the three most suitable extractive text summarization methods, and the averages of these values were used in the study. The methodology, testing, progress, results, and comparison processes of the study will be explained in detail in the subsequent stages.

2 Text Summarization

Text summarization is fundamentally based on linguistics. The better the linguistic structure of the text to be summarized is analyzed, the more accurate result can be obtained. Automatic text summarization can be implemented through three different methods based on the type of input, purpose, and output (Fig. 1). In terms of input type, single-document and multi-document methods are used in the summarization process. In terms of purpose, general, domain-specific, and query-based methods are employed in generating summaries. In general summarization, the summary of the text is directly extracted without relying on any specific focus. In domain-specific summarization, the summary is extracted with a specific domain focus, resulting in a higher accuracy rate. Query-based summarization is tailored to a given query and is primarily used in search engines. Regarding the output type, extractive and abstractive approaches are applied in summarization. In the extractive approach, the original text is fragmented and the summary is generated using the correct keywords, without any loss of information or disruption of semantic coherence. In the abstractive approach, not only the parts of the text are used, but new sentences are constructed with different words to preserve the meaning without distortion [8, 9].

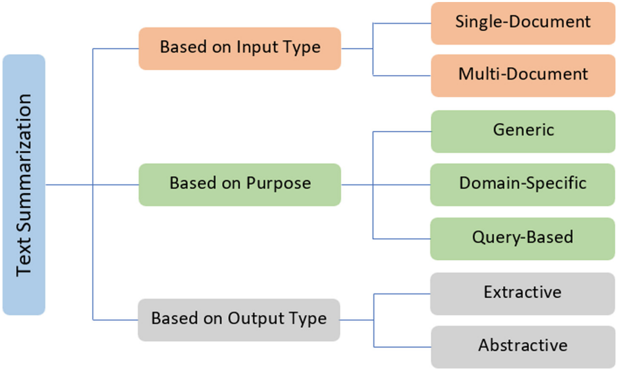


Fig. 1. Methods of Text Summarizations [9].

Latent Semantic Analysis (LSA) and Natural Language Processing (NLP) are commonly used techniques for text summarization [10]. NLP, which is utilized in this study, involves the analysis of the structure of natural languages to generate meaningful new output. When examining the conducted studies, it can be observed that NLP methods are predominantly employed in the automatic text summarization process. NLP is applied in various areas such as spell checking, translation processes, information extraction, language learning, text generation, text summarization, sentiment analysis, speech recognition, word and spelling correction.

In cases where the text content is not written in accordance with the rules of the language, it is necessary to identify the dirty data and make the text as compliant with the rules as possible before processing the summarization process. Adjustments should be made for incorrect or missing punctuation usage, informal expressions of words, or the absence of certain letters in order to ensure adherence to the rules.

In this study, the focus was placed on the technique of inferential text summarization within the context of text summarization approaches, and an attempt was made to present our findings by examining its achievements on Turkish news texts.

3 Literature Review

While there is a growing need to summarize the increasing amount of information on the internet, and the most accurate way to extract summaries is through human intervention, relying solely on human factors to generate summaries in this pool of information is quite challenging. Through a literature review, studies, comparisons, and research on this subject are summarized in order to find solutions.

In their study, Uzundere et al. [11] provided a method for summarizing Turkish news texts by scoring the sentences of the texts based on certain features. By using the reference summaries obtained from users, high-scoring sentences were selected to summarize the text. For the text to be summarized, the sentences were placed and sorted according to their word frequencies in the code, and the top 10% of sentences were selected based on the resulting sentence frequency index. The Zemberek [12] library was utilized to extract headlines and word stems and detect proper names. As a result, 10 news texts consisting of 7 to 42 sentences were summarized by 15 users, and the summaries generated by the application were compared, yielding a system performance of 55%.

In Erhandı et al.'s [13] study on deep learning-based text summarization, artificial neural networks were utilized. The study was prepared in two different comparative approaches. The first approach involved a Turkish-English comparison while keeping the dataset size constant, and the second approach focused on observing the output by working with a larger dataset exclusively for the Turkish language. News articles were provided as the dataset in the study, and single-sentence summaries were obtained from the system. The Keras library and PyCharm IDE were used, and three submodels, namely, encoder, decoder, and embedding, were employed for model construction. Accordingly, based on the test and training data provided as input, the system produces a one-sentence headline summary as its output. As the number of test data increased, slight improvements were observed. In the Turkish-English comparison, it was observed that as the number of epochs and test data increased, the English translation proximity was slightly higher than Turkish.

In Gündoğdu et al.'s [14] study conducted on the Turkish language, studies related to systems extracting summaries from textual content in the digital environment were examined to investigate the difficulties encountered and the methods used to overcome them. The reviewed studies mainly employed scoring sentences and words of the text to determine frequency values and utilized this approach for summarization. In this method, the text is divided into paragraphs, sentences, and words, which are evaluated and summarized accordingly.

In Hark et al.'s [15] study, which was prepared using an extractive text summarization method, the graph entropy method was employed. The Shannon Entropy [16] method was used to interpret the structural information content and stability represented by graphs. The Document Understanding Conference (DUC-2002) texts, which are publicly accessible and have summaries, were used as the dataset, and calculations were performed using Rouge metrics on 200-word and 400-word summaries. The text to be summarized was initially represented by graphs, and graph entropy was then applied to examine the structure of the texts. Subsequently, the sentences generated in the graph were compared with each other to establish a weighted sentence ranking. This method transfers the relationship between sentences to the graph, creating a general structure by associating nodes and edges with all combinations. The results showed that 400-word summaries exhibited better outcomes compared to 200-word summaries. Furthermore, when the achieved values were examined, LSA, TextRank, and LexRank text summarization systems generally achieved higher Rouge scores.

Boorugu et al.'s [17] study focuses on comparing NLP-based summarization studies. This study was conducted to meet the need for a system that can analyze all comments to reach a consensus by converting them into specific patterns with the same meaning. SVM [9] and Naive Bayes hybrid summarization techniques have shown noteworthy results in the examined studies. For summarizing the comments in this study, the seq2seq model is proposed, which is an advanced version of the Long Short-Term Memory approach used to increase the accuracy of the developed summary, accompanied by the attention mechanism.

In Güran et al.'s [18] study conducted a new Turkish text summarization system that combines structural and semantic features using a hybrid approach has been developed. This system, obtained from Turkish Wikipedia links, utilizes a total of five structural features, including three semantic features and two newly proposed feature. The weights calculated with two new approaches consist of an analytical hierarchical process based on expert evaluations derived from pairwise comparisons of the features, and an artificial bee colony algorithm used for automatically determining the feature weights. Experiments are conducted on a Turkish compilation consisting of 110 documents, which includes summary documents created by three human evaluators, in order to evaluate the performance of the hybrid system. The results indicate that combining all the features rather than using them individually yields better performance. The article also advocates for the integration of structural and semantic sentence features, claiming that it leads to the generation of more consistent summaries. Moreover, by extracting semantically related consecutive words using Turkish Wikipedia, the article contributes to the development of semantic document features based on Latent Semantic Analysis (LSA).

In Erdağı's [19] study, in addition to the summarization methods commonly used in the literature, hybrid method incorporating vowel harmony, consonant harmony, and both were proposed. News articles were used as the dataset, and these texts were initially summarized by three individuals. Subsequently, these three summaries were combined into a single summary using statistical methods. A total of 215 news articles were independently summarized by three teachers, which served as the reference data for control purposes. The number of summary sentences was determined based on the original text. Prior to summarization, the texts were examined and cleaned from noisy data. Eleven methods were applied in the study, eight of which are Term Frequency, Keyword, Title, Sentence Length, Sentence Position (Last Sentence), Sentence Position (First Sentence), Named Entity, and Singular Plural methods mentioned in the literature. The other three methods are the proposed Vowel Harmony, Consonant Harmony, and Hybrid methods that combine both. The scoring process was performed for each sentence using the MinMax Scaling method, generating values between 0 and 1. The relevant summarization methodologies were categorized, and their arithmetic averages were calculated and ranked based on similarity. The methods compared using the Rouge and BLEU metrics resulted in a similar ranking. In this single-document and extractive-based study, the inclusion of the proposed methods alongside the ones mentioned in the literature resulted in increased similarity values according to the Rouge-1, Rouge-2, and Rouge-L results. Similarly, the positive impact of the proposed methods was observed in the BLEU metric as well.

Additionally, in the decision mechanism regarding whether a sentence should be included in the summary or not, the K-Nearest Neighbor Classifier, Decision Tree Classifier, Random Forest Classifier, and Naive Bayes Classifier were used, with the K-Nearest Neighbor and Decision Tree structures demonstrating the highest performance. A 75% training and 25% test data split was applied in this case.

In the study conducted by Ülker et al. [20], a new Turkish dataset consisting of summaries created by individuals is presented, which is quite rare and abstract and designed for summarization studies. In this study, which is most similar to our work, the LexRank, TextRank, Luhn, LSA, and Random methods are compared. It is observed that all of the employed methods yield successful results in the dataset. In this study, which is compared using the Rouge metric algorithm, when examining the F (f-measure) value, it is mostly observed that TextRank algorithm has the highest value, and Random algorithm has the lowest value, particularly in terms of Rouge-1, Rouge-2, and Rouge-L. When examining this study for the R (recall) value that we focused on in our research, it is observed that the Rouge-1, Rouge-2, and Rouge-L values for LexRank, TexRank, and Luhn algorithms are quite close to each other, and there is no clear winner in terms of the metric. The TextRank algorithm, which ranks at the end in terms of similarity in our study, is ranked closely with LexRank and Luhn algorithms in terms of high similarity values in this study, and the reasons for this are discussed in the conclusion section. Therewithal it is observed that the algorithm with the lowest value in the recall output is mostly the LSA algorithm.

The related studies in the literature are as follows, whereas the details regarding our study are being elaborated under the sections of methodology and dataset.

4 Methodology and Dataset

The generated works were developed in Python language, and summarization and comparisons were compiled using Visual Studio Code.

In the study, for the TextRank algorithm, sentences and words were first tokenized and transferred to lists using the sent_tokenize and word_tokenize methods of the nltk.tokenize library. The nltk.corpus library's stopwords method was used to exclude irrelevant words. During these processes, the 'language' parameter was set to 'turkish'. For extracting word stems and finding their frequencies, the nltk.stem and PorterStemmer were used. The summarization process was performed separately for each text using the necessary methods, and sentence scores were determined. Finally, the highest-frequency sentences, amounting to one-third of the total number of sentences in the text, were selected to create the summary.

For the LexRank algorithm, sentences were tokenized using the sent_tokenize method of the nltk.tokenize library and transferred to lists. The nltk.corpus library's stopwords method was used to exclude irrelevant words, and the 'language' parameter was set to 'turkish'. Then, using the sklearn.feature_extraction.text module and the CountVectorizer with the fit_transform method, the sentences in the text were transformed into numerical vectors, creating a matrix. This matrix represents the foundation of the LexRank algorithm and is scored by calculating cosine similarity using the cosine_similarity method from the sklearn.metrics.pairwise library. This method calculates the similarity values of the text vectors in the matrix. Similarly, the highest-frequency sentences, amounting to one-third of the total number of sentences in the text, were selected to create the summary.

For the Luhn algorithm, the words, converted to lowercase, were tokenized using the sent_tokenize and word_tokenize methods of the nltk.tokenize library to create sentence and word lists. The 'language' parameter was set to 'turkish' in this case as well. The number of occurrences of words in the text was collected and stored in a list using the collections Counter. Based on the frequency of occurrence, sentence scores were determined. The sentences with the highest scores, amounting to one-third of the total number of sentences in the text, were selected to create the summary.

TextRank, LexRank, and Luhn algorithms were utilized with minor revisions, while preserving their structures, by employing the relevant libraries. The study, which resulted in the generation of two summary outputs based on frequency, consists of several sections. In this study, firstly, sentences are parsed through the sent_tokenize method of the nltk.tokenize library. Then, empty sentences or sentences containing only punctuation marks within the sentences are cleaned without changing the order, and then the roots of the words are determined. The widely used Zemberek library, which has comprehensive studies on the Turkish language structure, was utilized for the determination of word roots. The roots of all the words in the text are accessed using Zemberek, and these roots are collected into a list while keeping track of the total number of occurrences of a word root in the text. At this stage, it was observed in the tests that the roots of some words could not be determined, some words could not be addressed due to punctuation marks, or some proper names were also addressed ambiguously as "Unknown" (UNK). Adjustments were made for these cases by adding the words correctly to the library either through the code or by ensuring their correct spelling in the texts and all summaries in order to prevent incorrect frequency values and incorrect summarization. Pointless word roots such as conjunctions and suffixes that should not be included in the calculation were removed from the root list. The nltk.corpus library's stopwords method was used for the removal process. However, it was observed that some words or suffixes that should not be included in the calculation were missing, so they were added to this list as well. In addition, to prevent incorrect determination of sentence weights, the frequency values of word roots addressed as UNK were assigned as zero as a precaution (although punctuation marks and spaces are cleaned, these values can still be added to the list as UNK due to reasons that may arise from spelling errors in the text). After determining the word root values and the sentences to be scored, the sentences are scored based on the frequency values of the word roots they contain.

The calculation process is performed as follows:

Let's assume we have a sentence consisting of three words as an example, and the score of this sentence is being determined. In this case, let's say the root of the first word appears once in the text, the root of the second word appears six times, and the root of the third word appears four times. In this case, the score of this sentence is determined as $1 + 6 + 4 = 11$. This way, all sentences are scored, and it is crucial to accurately identify the word roots. Furthermore, in order to prevent erroneous calculations, adjustments have been made for special names, dates, numerical values, and detailed examinations have been conducted manually in texts and summaries. Finally, the sentences are sorted in descending order based on their scores, and two different types of summary outputs are generated. In the first type, a sentence with the highest frequency value, which is equal to one-third of the total number of sentences in the text, is brought in the correct order and the inferential summarization process is performed. This method is abbreviated as '1in3' in the study. In the second type, the total scores of the sentences in the text are divided by the number of sentences in the text to find an average value, and the sentences with scores above this average value are arranged in the correct order to create an inferential summary text. This method is abbreviated as 'Avg' in the study.

Each generated summary is categorized and transferred to folders for analysis according to the Rouge metric method. The summarized texts obtained through the algorithms are compared with reference summaries to determine similarity values, and the averages are calculated. Two average values are calculated for the outputs. The first one is the arithmetic average of the resulting metrics, and the second one is the metric averages obtained based on the number of sentences in the summary texts.

The Average of Arithmetic Averages (A.A.A.) (Table 3): Here, the summary texts generated by the system using the TextRank, LexRank, Luhn, Avg, and 1in3 summarization methods were compared with the 3 summaries provided by real individuals for 130 texts, and the r1, r2, and rl rouge values were obtained. The arithmetic averages of the rouge values obtained for each text and summarization method were calculated, and then these averages were grouped based on the summarization method, summed up, and divided by the total number of texts, which is 130, to obtain the arithmetic rouge metric average of the dataset.

The Average of Sentence-based Averages (A.S.B.A.) (Table 3): The aim here is to obtain an average based on the number of sentences in the summarized text. Again, the system summaries obtained from five applied algorithms were compared with the reference summaries of the dataset to calculate the r1, r2, and rl values. However, in this case, after dividing the number of sentences in the reference summary by the number of sentences in the system summary, the result values were obtained by multiplying them with the rouge metric. This process was performed separately for each text, summarization algorithm, reference summary, and rouge value. The goal here is to take into account the disadvantage of the work, which aims to have a minimum summary with maximum efficiency, for summaries with a high number of sentences. Thus, by evaluating the rouge value at the sentence level for summaries with a high number of sentences, a fair average is aimed to be achieved. The arithmetic averages of the resulting values were calculated for each text, and the outputs of all texts were summed up and divided by the total number of texts, which is 130, to obtain the sentence-based rouge metric average of the dataset.

The dataset used in this study is the text dataset consisting of 130 news texts from various newspapers was utilized in the "Otomatik Metin Özetleme Sistemi" [21] doctoral thesis. In the texts, while the minimum number of sentences in a text is 7, the text with the maximum number of sentences consists of 63 sentences. The reference summaries were obtained by taking the summaries separately generated by three individuals for the texts within the same study. In the relevant study, it is stated that an interface was created for the summarization method and individuals created the summaries by selecting sentences. In addition, no sentence restriction was imposed on the summarizers regarding the texts. In this regard, the first summarizer provided summarization with a rate of 34%, the second summarizer with a rate of 37%, and the third summarizer with a rate of 26%.

The study was conducted and tested on a computer with an Intel(R) Core(TM) i7-3610QM CPU, 2.30 GHz processor, Intel(R) HD Graphics 4000 graphics card, 8 GB RAM, and a 64-bit Windows operating system. The resulting values were examined and detailed in the results section.

5 Tests and Evaluation of Summarization Techniques

A comprehensive review was conducted by scanning the literature on automatic text summarization methods and approaches, examining studies conducted at an international level and specific to different languages. Open-source projects published in this field were also investigated, and experiments were conducted and outputs were evaluated based on these approaches. Progress was made by conducting research and studies on commonly used Python language and widely used libraries specific to this field. During the research process, GitHub projects were downloaded and executed in Visual Studio Code to examine the studies prepared for summarization, and the logic of these studies was investigated to analyze the approaches applied in different languages. It was observed that these studies, mostly focused on English, were predominantly developed in Python with the utilization of specific libraries. The aim of the study is to contribute to the literature by comparing the accuracy rates of commonly used extractive text summarization methodologies specifically for the Turkish language. It has been observed in the examined open-source projects and literature review that certain extractive text summarization algorithms are frequently preferred due to their higher accuracy in producing results. As a result of the investigations, the three most preferred algorithms were determined, and detailed examinations were conducted on these three algorithms. The identified algorithms are TextRank, LexRank, and Luhn algorithms. Therefore, in this study, summaries generated using the most commonly used TextRank, LexRank, and Luhn algorithms, as well as two different outputs of the frequency-based text summarization algorithm developed in Python, were compared with reference outputs summarized by three individuals using the Rouge metric algorithm.

Among the algorithms used, TextRank is based on the creation of nodes by dividing the text into sentences or groups of words and weighting these nodes based on their similarities in a graph structure. This way, a summary text is generated based on the ranking of weights [22]. In the LexRank algorithm, the weights of sentences in the text are determined, and graph-based nodes are created. Cosine similarity, which is used to measure the similarity between vectors, is employed to determine node similarity. Important sentences are identified and a summary text is generated by taking into account sentence weights and similarity degrees [23]. In the Luhn algorithm, calculations are made by considering the length, importance level, and word frequency of sentences. It is based on the logic of dividing the text into sentences according to the importance score, and when determining the importance score, the sentence length and frequency of important words in sentences are taken into account. The summary is created by ordering sentences with high importance scores [24].

In addition to these three summarization algorithms, to strengthen the study, an algorithm based on word frequency, which has been emphasized in some open-source works and relies on extractive text summarization, has been developed. This algorithm is expected to provide more accurate outputs as it is tailored specifically for this study and revised based on Turkish sentence structures. As detailed in the methodology section, various revisions have been made to achieve results closer to accuracy in this algorithm, which selects sentences based on the frequency of word stems in the text. In addition to the stopwords present in the nltk.corpus library for Turkish, certain words and affixes have been added to this list. Furthermore, in order to find stems, the Zemberek library has been enhanced with the addition of unidentified proper nouns (e.g., YAY-KUR, Kızılçukur, Memur-Sen) and words with undetectable stems (e.g., eşitlik, rinoplasti, tüf). The text and sentences have been encoded as utf-8-sig to ensure proper recognition of the Turkish alphabet and the removal of empty or excessively separated sentences has been ensured.

As an example, the original text of a news in the data set, summaries made by individuals and summaries made by the system are presented below (Table 1).

Table 1. A news text and its corresponding summaries.

Original	Aslanlı Köşk dizi seti oldu TMSF’nin Halis Toprak’ın ödenmemiş borçlarından dolayı el koyduğu Aslanlı Köşk, “Kuzey Güney” ekibinin yeni mekanı oldu Halis Toprak’ın, 21 yıl yaşadıktan geçtiğimiz şubat ayında boşaltmak zorunda kaldığı Sarıyer’deki meşhur Aslanlı Köşk, artık bir dizi seti Köşk, haftalardır Kanal D’nin sevilen dizisi “Kuzey Güney” dizisinin ekibini ağırlıyor Dizinin zengin ve şımarık karakteri Banu Sinaner’in ailesiyle yaşadığı ev olarak işte bu köşk kullanılıyor Aylardır boş duran köşk, bu dizi için yapım firması tarafından kiralanıp baştan aşağı restore edildi Köşkün yenilenme sürecinde kalabalık bir mimar ve işçi grubu çalıştı Aslanlı Köşk’ün iç dekorasyonu ve bahçe peyzajı, tam iki ayda yenilendi Ve yenilenen köşk, dizinin çarşamba günün yayınlanan bölümde ilk kez gösterildi
1. Person Summ	TMSF’nin Halis Toprak’ın ödenmemiş borçlarından dolayı el koyduğu Aslanlı Köşk, “Kuzey Güney” ekibinin yeni mekanı oldu. Dizinin zengin ve şımarık karakteri Banu Sinaner’in ailesiyle yaşadığı ev olarak işte bu köşk kullanılıyor. Aylardır boş duran köşk, bu dizi için yapım firması tarafından kiralanıp baştan aşağı restore edildi
2. Person Summ	TMSF’nin Halis Toprak’ın ödenmemiş borçlarından dolayı el koyduğu Aslanlı Köşk, “Kuzey Güney” ekibinin yeni mekanı oldu. Halis Toprak’ın, 21 yıl yaşadıktan geçtiğimiz şubat ayında boşaltmak zorunda kaldığı Sarıyer’deki meşhur Aslanlı Köşk, artık bir dizi seti. Aylardır boş duran köşk, bu dizi için yapım firması tarafından kiralanıp baştan aşağı restore edildi
3. Person Summ	Halis Toprak’ın, 21 yıl yaşadıktan geçtiğimiz şubat ayında boşaltmak zorunda kaldığı Sarıyer’deki meşhur Aslanlı Köşk, artık bir dizi seti. Köşk, haftalardır Kanal D’nin sevilen dizisi “Kuzey Güney” dizisinin ekibini ağırlıyor. Aslanlı Köşk’ün iç dekorasyonu ve bahçe peyzajı, tam iki ayda yenilendi
Avg	TMSF’nin Halis Toprak’ın ödenmemiş borçlarından dolayı el koyduğu Aslanlı Köşk, “Kuzey Güney” ekibinin yeni mekanı oldu. Halis Toprak’ın, 21 yıl yaşadıktan geçtiğimiz şubat ayında boşaltmak zorunda kaldığı Sarıyer’deki meşhur Aslanlı Köşk, artık bir dizi seti. Köşk, haftalardır Kanal D’nin sevilen dizisi “Kuzey Güney” dizisinin ekibini ağırlıyor. Dizinin zengin ve şımarık karakteri Banu Sinaner’in ailesiyle yaşadığı ev olarak işte bu köşk kullanılıyor
1in3	TMSF’nin Halis Toprak’ın ödenmemiş borçlarından dolayı el koyduğu Aslanlı Köşk, “Kuzey Güney” ekibinin yeni mekanı oldu. Halis Toprak’ın, 21 yıl yaşadıktan geçtiğimiz şubat ayında boşaltmak zorunda kaldığı Sarıyer’deki meşhur Aslanlı Köşk, artık bir dizi seti. Köşk, haftalardır Kanal D’nin sevilen dizisi “Kuzey Güney” dizisinin ekibini ağırlıyor

(*continued*)

Table 1. (*continued*)

TextR	Aslanlı Köşk dizi seti oldu. TMSF'nin Halis Toprak'ın ödenmemiş borçlarından dolayı el koyduğu Aslanlı Köşk, "Kuzey Güney" ekibinin yeni mekanı oldu. Halis Toprak'ın, 21 yıl yaşadıktan geçtiğimiz şubat ayında boşaltmak zorunda kaldığı Sarıyer'deki meşhur Aslanlı Köşk, artık bir dizi seti
LexR	Aslanlı Köşk dizi seti oldu. TMSF'nin Halis Toprak'ın ödenmemiş borçlarından dolayı el koyduğu Aslanlı Köşk, "Kuzey Güney" ekibinin yeni mekanı oldu. Halis Toprak'ın, 21 yıl yaşadıktan geçtiğimiz şubat ayında boşaltmak zorunda kaldığı Sarıyer'deki meşhur Aslanlı Köşk, artık bir dizi seti
Luhn	Aylardır boş duran köşk, bu dizi için yapım firması tarafından kiralanıp baştan aşağı restore edildi. Halis Toprak'ın, 21 yıl yaşadıktan geçtiğimiz şubat ayında boşaltmak zorunda kaldığı Sarıyer'deki meşhur Aslanlı Köşk, artık bir dizi seti. Ve yenilenen köşk, dizinin çarşamba günün yayınlanan bölümde ilk kez gösterildi

In the conducted study, the comparison process of the generated outputs with the reference texts was performed using the Rouge Metric method. This metric, which is widely used in the field of NLP, is used to determine how similar and inclusive the generated summary is compared to the reference summary [25]. Different calculations are performed during the comparison, and these calculations are expressed with different Rouge outputs. It has also been observed in the research and investigations that text summarization comparisons are most commonly evaluated using Rouge-1 (r1), Rouge-2 (r2), and Rouge-L (rl) scores. Among these outputs, r1 calculates at the word level, r2 calculates at the level of word pairs (bigrams), and rl calculates at the level of the longest common subsequence.

Due to literature reviews, examinations of previously conducted studies, and its development in the field of NLP, the rouge metric method stands out as one of the most reliable approaches for conducting summarization comparisons. In particularly, sub-metrics such as r1, r2, and rl are used to objectively compare text similarity in such studies. This approach, which facilitates the comparison between reference text and text automatically summarized by the system, is widely embraced as an industry-standard practice. Thus, the utilization of the rouge metric holds significant importance in this study for the analysis of results and accurate statistical guidance.

When calculating r1, the number of matching words between the reference summary and the system-generated summary is divided by the total number of words in the reference to generate a score. When calculating r2, the number of matching bigrams is divided by the total number of bigrams in the reference to obtain a value. This method provides a slightly more comprehensive comparison of summary similarity. When calculating rl, the length of the longest common subsequence between two texts is divided by the total number of words in the reference summary. In this calculation, word order is also taken into account to derive a score. The comparison results in a value between zero and one, where zero represents the lowest similarity and one represents the highest similarity (Table 3). In the study, the averages of the most commonly used r1, r2 and rl values of the recall (r) outputs were compared. Each of these outputs provides information regarding similarity; however, for a better observation of accuracy, it is more appropriate to examine each value in order to achieve more consistent results. In order to achieve more consistent similarity values and higher accuracy rates, instead of using a single metric, the study performed a comparison by taking the averages of r1, r2, and rl metrics.

Table 2. The division of the total number of sentences in the dataset by the amount of data.

Arts	Ref1	Ref2	Ref3	Avg	1in3	TextR	LexR	Luhn
20.4	6.7	7.2	5.1	9.1	7.3	7.2	7.2	7.2

Table 3. Rouge metric averages of approaches.

1–130	A.A.A	A.S.B.A
Avg	0,744455478026611	0,549012936417251
1in3	0,661578194026025	0,588113788089480
TextR	0,352319233708779	0,313637309182563
LexR	0,489789005423816	0,436648350243383
Luhn	0,574737435229811	0,523319599562985

6 Discussion

As a result of the conducted study, when the Rouge Metric values and their averages for r1, r2, and rl are examined, the algorithm with the highest similarity in terms of arithmetic average (A.A.A.) among the five extractive summarization algorithms is the algorithm called Avg. This algorithm determines the weight of a sentence based on the frequency of all words and selects sentences that have an average frequency value above a certain threshold. However, the disadvantage of this algorithm is that the number of sentences is slightly higher compared to the other sentences. Furthermore, when the sentence averages of the remaining summary texts are examined (Table 2), it is observed that the average number of sentences is 7.2. When the summarization methodologies with the average number of sentences are considered, the algorithm with the highest metric value is the algorithm called 1in3, which selects sentences that have the highest frequency value among the determined sentences and brings one-third of the number of sentences in the text with the highest frequency. The algorithm with the lowest average value in the ranking is the TexRank (TextR) algorithm with a metric value of 0.35 (Table 3).

When we examine the Rouge metric averages with a formula that takes into account the number of sentences, the values come out in a similar ranking (A.S.B.A.). Here, the Avg algorithm appears lower than the arithmetic average as expected due to its disadvantage. This demonstrates that this approach, which focuses on a sentence basis, provides consistent results.

Another result that can be derived from the table is that the TextRank algorithm, which is one of the most commonly used extractive text summarization algorithms, does not perform well specifically for the Turkish language.

When reviewing the literature, it is observed that there is a study comparing these three algorithms or any two of them specifically for the Turkish language. When examining this study, which is also emphasized in the literature review section, it is observed that in the comparison of the recall (R) value used in our study, the TextRank, LexRank, and Luhn values are generally very close to each other and each algorithm has instances where it performs better. However, in our study, it is observed that the TextRank value is noticeably lower. Here, auxiliary processes such as sentence and word tokenization, stemming, and graph construction, which are performed outside the algorithm, can affect the summarization process. It is considered that the deviation of the TextRank algorithm in this study may be related to these factors.

7 Conclusion

The conclusion to be drawn from the evaluations in the discussion section and the values of work outputs is that extractive summarization methods based on word-level sentence frequency are more successful for Turkish.

It is anticipated that this study will contribute to the comparison and decision-making processes regarding extractive text summarization algorithms that are scarce in the literature specifically for Turkish.

In the future, to improve the study, the synthesis of different two or three algorithms can be performed to obtain higher accuracy outputs. The result comparison can be detailed by using a larger dataset. Additionally, improvements can be made to the generated summarization algorithm (Avg and 1in3) by detecting and automatically correcting misspelled words, identifying incorrect punctuation marks, improving sentence segmentation, and parsing proper names, abbreviations, or unknown words based on inference.

References

1. Chatterjee, N., Mittal, A., Goyal, S.: Single document extractive text summarization using genetic algorithms. In: 2012 Third International Conference on Emerging Applications of Information Technology, pp. 19–23. IEEE (2012). https://doi.org/10.1109/EAIT.2012.6407852
2. Tandel, A., Modi, B., Gupta, P., Wagle, S., Khedkar, S.: Multi-document text summarization-a survey. In: 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE), pp. 331–334. IEEE (2016). https://doi.org/10.1109/SAPIENCE.2016.7684115
3. Shi, Z.: The design and implementation of domain-specific text summarization system based on co-reference resolution algorithm. In: 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery, vol. 5, pp. 2390–2394. IEEE (2010). https://doi.org/10.1109/FSKD.2010.5569529
4. Wei, Y., Zhizhuo, Y.: Query based Summarization using topic background knowledge. In: 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), pp. 2569–2572. IEEE (2017). https://doi.org/10.1109/FSKD.2017.8393180

5. Gigioli, P., Sagar, N., Rao, A., Voyles, J.: Domain-aware abstractive text summarization for medical documents. In: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 2338–2343. IEEE (2018). https://doi.org/10.1109/BIBM.2018.8621539
6. Raphal, N., Duwarah, H., Daniel, P.: Survey on abstractive text summarization. In: 2018 International Conference on Communication and Signal Processing (ICCSP), pp. 0513–0517. IEEE (2018). https://doi.org/10.1109/ICCSP.2018.8524532
7. Jain, A., Bhatia, D., Thakur, M.K.: Extractive text summarization using word vector embedding. In: 2017 International Conference on machine learning and data science (MLDS), pp. 51–55. IEEE (2017. https://doi.org/10.1109/MLDS.2017.12
8. Güran, A., Arslan, S.N., Kılıç, E., Diri, B.: Metin özetleme için cümle seçim metotları (2014). https://doi.org/10.1109/SIU.2014.6830198
9. Ramesh, G., Somasekar, J., Madhavi, K., Ramu, G.: Best keyword set recommendations for building service-based systems. Int. J. Sci. Technol. Res. **8**(10) (2019)
10. Altintaş, V., Topal, K., Albayrak, M.: Sosyal medya platformu üzerinde gizli anlam analizi. Avrupa Bilim Teknol. Dergisi **16**, 863–869 (2019). https://doi.org/10.31590/ejosat.590521
11. Uzundere, E., Dedja, E., Diri, B., Amasyalı, M.F.: Türkçe haber metinleri için otomatik özetleme. Akıllı Sistemlerde Yenilikler Uygulamaları Sempozyumu 1–4 (2008)
12. Zemberek: Natural Language Processing Library for Turkish. (n.d.). https://code.google.com/archive/p/zemberek/. Accessed 19 July 2023
13. Erhandi, B., Çalli, Ö.Ü.F.: Derin Özetleme ile Metin Özetleme (2020)
14. Gündoğdu, Ö.E., Duru, N.: Türkçe Metin Özetlemede Kullanılan Yöntemler. 18. Akademik Bilişim Konferansı-AB 2016 (2016)
15. Hark, C., Uçkan, T., Seyyarer, E., Karci, A.: Extractive text summarization via graph entropy çizge entropi ile çikarici metin özetleme. In: 2019 International Artificial Intelligence and Data Processing Symposium (IDAP), pp. 1–5. IEEE (2019). https://doi.org/10.1109/IDAP.2019.8875936
16. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**(3), 379–423 (1948). https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
17. Boorugu, R., Ramesh, G.: A survey on NLP based text summarization for summarizing product reviews. In: 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 352–356. IEEE (2020). https://doi.org/10.1109/ICIRCA48905.2020.9183355
18. Güran, A., Bayazit, N.G., Gürbüz, M.Z.: Efficient feature integration with Wikipedia-based semantic feature extraction for Turkish text summarization. Turk. J. Electr. Eng. Comput. Sci. **21**(5), 1411–1425 (2013). https://doi.org/10.3906/elk-1201-15
19. Erdağı, E.: Türkçe metinlerde çıkarım tabanlı otomatik metin özetleme (2023)
20. Ülker, M., Özer, A.B.: TTSD: a novel dataset for Turkish text summarization. In: 2021 9th International Symposium on Digital Forensics and Security (ISDFS), pp. 1–6. IEEE (2021). https://doi.org/10.1109/ISDFS52919.2021.9486337
21. Güran, A.: Otomatik metin özetleme sistemi (2013)
22. Mihalcea, R., Tarau, P.: TextRank: bringing order into text. In: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, pp. 404–411 (2004)
23. Erkan, G., Radev, D.R.: LexRank: graph-based lexical centrality as salience in text summarization. J. Artif. Intell. Res. **22**, 457–479 (2004). https://doi.org/10.1613/jair.1523
24. Luhn, H.P.: The automatic creation of literature abstracts. IBM J. Res. Dev. **2**(2), 159–165 (1958). https://doi.org/10.1147/rd.22.0159
25. Lin, C.Y.: ROUGE: a package for automatic evaluation of summaries. In: Text Summarization Branches Out, pp. 74–81 (2004)

Formation of a Speech Database in the Karakalpak Language for Speech Synthesis Systems

N. S. Mamatov[1](✉), K. M. Jalelov[1], B. N. Samijonov[2], A. N. Samijonov[3], and A. D. Madaminjonov[1]

[1] Digital Technologies and Artificial Intelligence, Tashkent Institute of Irrigation and Agricultural Mechanization Engineers" National Research University, Tashkent, Uzbekistan
m_narzullo@mail.ru
[2] Sejong University, Seoul, South Korea
[3] Tashkent University of Information Technologies Named After Muhammad Al-Khwarizmi, Tashkent, Uzbekistan

Abstract. The article deals with speech synthesis, the conversion of an arbitrary text given in a natural language into spoken form in that language, and this issue has been of interest to mankind since ancient times. Also, the paper described mechanical, electrical, articulating, format, linear prediction, concatenative, selective, and statistical parametric synthesis methods. The paper describes how a high-quality speech database is essential for creating speech synthesis systems. Also, the paper notes that such databases have been created for many languages, but no speech database has been created to translate text information in the Karakalpak language. Therefore, creating a speech database for converting Karakalpak text into speech is urgent. Furthermore, this article depicts the problems and their solutions related to the creation of a speech database in the Karakalpak language.

Keywords: text-to-speech (TTS) · speech database · synthesis system · digital signal processing (DSP)

1 Introduction

The Karakalpak language, a member of the Turkic language family, holds the official status in the Republic of Karakalpakstan, an autonomous region within the Republic of Uzbekistan. About a million people speak this language, which combined with Kazakh and Nogai, two other Turkic languages that are a part of the Kipchak language group, makes up the Kipchak language family. The Khorezm, Navoi, and Bukhara regions, as well as nearby Kazakhstan and Turkmenistan, the Russian Federation, and Afghanistan, are the main distribution areas for Karakalpak language speakers. The two north-eastern and two south-western dialects of the Karakalpak language are primarily separated from one another phonetically. The literary Karakalpak language evolved during the initial decades of the 20th century, drawing inspiration from the northeastern dialect. It boasts

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 333–343, 2024.
https://doi.org/10.1007/978-3-031-56728-5_28

a remarkable history and culture, making it a fascinating linguistic development. The Karakalpak people's traditional music, dance, and literature are all expressed in this language, which is regarded as their mother tongue. Speaking of several Turkic languages, it is widely utilized and of strategic importance in Central Asia. Communication and understanding amongst the many ethnic and linguistic groups in the area are made easier by knowing and comprehending the Karakalpak language.

The Karakalpak language is also important for scientific research. Linguists and anthropologists can learn a lot about the history and culture of the Karakalpak people by studying their language. The development of a text synthesis system in the Karakalpak language is expected to contribute to the advancement of language technologies and the preservation of this language for future generations.

The creation of a text-to-speech (TTS) synthesis system in any natural language, including the Karakalpak language, requires a speech database. It is employed to instruct the system on how to produce the proper intonation, rhythm, and sounds of the language. This necessitates using a wide variety of speech, including official and informal speech as well as speech with various emotions, intonations, and accents.

The ability to produce high-quality synthetic speech will help make the language more approachable to a wider audience, including non-fluent speakers, as speech technologies advance. As a result, the Karakalpak language will be encouraged to be used in a variety of contexts, including business, entertainment, and education. This will also assist to conserve the language for future generations.

This article discusses the technical and linguistic considerations of speech data recording and the importance and sources of speech data recording for native speakers.

TTS is a cutting-edge technology that transforms written text into spoken language, utilizing algorithms and advanced techniques to analyze and comprehend the written material to produce precise speech output.

The fundamental components of TTS systems include text analysis, linguistic processing, and speech synthesis. The text analysis module scrutinizes the written content. This breaks the text down into manageable chunks, like words, phrases, and sentences. The language processing component then applies grammar, pronunciation, and intonation rules to provide a phonetic representation of the text. The phonetically represented text is then translated into speech that is sufficient for hearing through the speech synthesis component [1].

TTS systems can produce speech that sounds robotic or natural-sounding and mimics human voice and intonation. TTS technology has many different uses. These include speech recognition technology, automated voice assistants, and aids for people who have trouble speaking [2].

2 Methodology

In general, a TTS synthesizer consists of a text analysis module and a digital signal processing (DSP) module (Fig. 1). The text analysis module serves to create a phonetic transcription of the read text with the necessary intonation and rhythm. The DSP module enables the production of synthetic speech that is suitable for transcription by the text analysis module (Fig. 2).

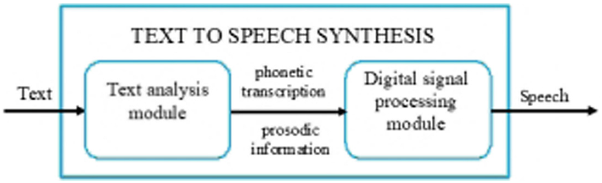


Fig. 1. General functional scheme of TTS synthesis

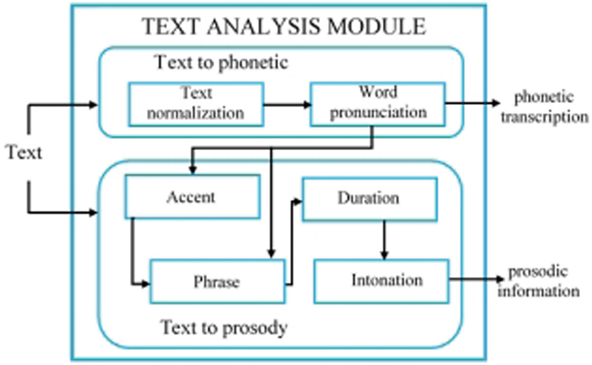


Fig. 2. TTS system text analysis module

The text analysis step is very complex because this step only needs to generate all the information required by the DSP module (for speech generation) from the text. However, plain text does not contain all the information needed to create a speech. The first block of the text analysis module converts the entered text into phonetic transcription (Text-to-Phonetics (T2P)), and the second block produces prosodic information from the text (Text-to-Prosody).

The text-to-phonetic transcription block can be divided into such modules as text normalization and word pronunciation. Below is a brief description of these two modules:

The supplied text is organized by the text normalization module using predetermined word lists. It stretches any numbers, acronyms, or idioms it finds into full text. Usually, basic grammar is used for this.

Word pronunciation The pronunciation of a string of words is established once they have been generated using the text normalization module. The simple letter-to-speech (LTS) rule is used when words are spoken exactly as they are written. When the contrary situation arises, a morphosyntactic analyzer becomes necessary. This tool is designed to group words within sentences based on their syntactic relationships, including nouns, verbs, and adjectives, while also classifying speech elements like prepositions, stems, and adverbs according to their specific properties. A lexicon is then used to establish how they should be pronounced.

Turning from text to prosody. The term prosody refers to certain features of the speech signal, such as loudness in pitch, i.e., intonation, or vocal changes in tempo, duration, stress, and rhythm. The naturalness of speech is mainly described in terms of prosody. Prosodic events are also called suprasegmental events because these events do not coincide with segments (sounds, phonemes), but with syllables or groups of syllables.

How speech sounds are pronounced, known as prosody, plays a significant role in communicating the meaning of sentences. To generate the appropriate prosody for a given text, a text-to-prosody block is utilized. This block utilizes both the text itself and the output of the word pronunciation module to produce prosodic information. The text-to-prosody block can be further divided into several subprocesses, each of which is responsible for determining stress, phrasing, duration, and intonation for every sentence. Below is a brief description of these four processes:

The accent is based on word order. The placement of words in a sentence determines where the emphasis is placed, and this can greatly impact the way the sentence is interpreted. When constructing a sentence, it is important to consider the role of different types of words. For instance, content words such as nouns, adjectives, and verbs are usually given greater stress compared to function words such as articles, prepositions, and conjunctions. Understanding these patterns of emphasis can aid in predicting the intonation and duration of the sentence. By considering these factors, one can ensure that the intended meaning and emphasis of the sentence are accurately conveyed to the listener or reader.

Phrases are divided into phrasal verbs, and phrase boundaries are set that match the text, and these boundaries make it possible to restore pauses and intonation contours.

Intonation determines the type and meaning of a sentence. The sentence can be neutral, a command, or a question. In addition, intonation can be used to determine speaker features such as the speaker's gender, age, and emotions. The intonation module creates pitch contours for sentences. For example, "Esikti ash" and "Esikti ashıń" have different prosody. In terms of intonation contour, the first sentence is imperative and has a relatively flat pitch contour, while the second is interrogative, signifying a tone that ascends towards the sentence's conclusion.

Duration Segmental continuity is an important aspect of prosody and affects the general rhythm of speech, stress, syntactic structure of the sentence, and speech rate. Many other factors affect the length of a speech segment.

Digital signal processing (DSP) module. This module uses phonetic transcription and prosodic data generated based on the text analysis module for speech formation. This can be done in two ways, namely:

- using a set of rules describing how one phoneme affects another (called the coarticulation effect)
- storing different copies of each speech sound unit and using them as the final acoustic units.

The two primary categories of TTS systems, synthesis by rule and synthesis by combination, have been developed based on the methods described above. The DSP module's overall structure is illustrated in the accompanying figure (Fig. 3).

A TTS system is a sophisticated piece of computer software that combines several parts to create synthesized speech from written text. The following will be the TTS system's primary elements:

Text analysis. The TTS system's text analysis component analyzes the incoming text to separate and categorize individual text components including words, phrases,

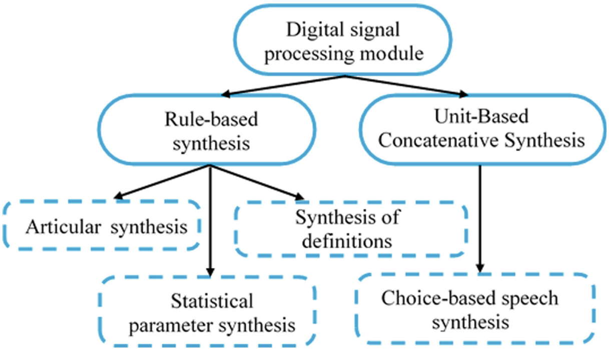

Digital signal processing module
Rule-based synthesis
Unit-Based Concatenative Synthesis
Articular synthesis
Synthesis of definitions
Statistical parameter synthesis
Choice-based speech synthesis

Fig. 3. Schematic of a digital signal processing (DSP) module

and sentences. To analyze the grammatical structure of the text, it carries out tasks like tokenization, assigning part-of-speech labels, and parsing [3].

Linguistic processing. The TTS system's linguistic processing section uses linguistic models and rules to translate a text's syntactic expression into a phonetic representation. Prosody, stress, and intonation, as well as pronunciation modeling, fall under this category.

Speech synthesis. Using a speech synthesis component, the TTS system converts the phonetic representation of the text into audible speech. To create understandable and natural-sounding speech sounds, it makes use of a variety of signal-processing techniques, such as filtering, amplification, and modulation (Fig. 4).

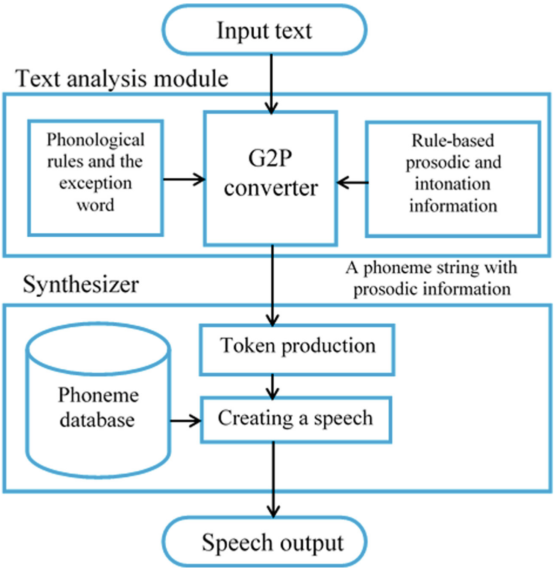

Input text
Text analysis module
Phonological rules and the exception word
G2P converter
Rule-based prosodic and intonation information
Synthesizer
A phoneme string with prosodic information
Token production
Phoneme database
Creating a speech
Speech output

Fig. 4. Speech synthesis system

Sound database. This is a collection of speech samples written by native speakers. Samples are used to train a TTS system to produce speech that closely resembles natural speech. The database employs phonetic transcriptions of speech samples for the training of linguistic and speech synthesis models.

User interface. It is the component of a TTS system that allows users to enter text and interact with the system, and it can be a command-line interface, a web interface, or an API.

Audio Output: The audio output component of a TTS system is used to convert synthesized speech into an audio format that can be played on a loudspeaker or other audio device [4].

The database typically necessitates speech samples from a native speaker proficient in the desired language, along with phonetic transcriptions and linguistic annotations [5]. Some of the ways in which a speech database is important to a TTS system include:

Teaching the speech synthesis model. The TTS system makes extensive use of machine-learning techniques to generate natural-sounding speech. The speech database serves to train the speech synthesis model of the system, enabling it to learn and compare the written text with the corresponding speech sounds. The more high-quality speech data a system has access to, the better it will be at generating natural-sounding speech [6].

Improve pronunciation accuracy. The speech database can be used to improve the system's accuracy in pronouncing words and phrases. By providing the system with multiple examples of how a native speaker would pronounce words, it can be taught to pronounce them correctly and consistently.

Support for multiple voices. Typically, a TTS system uses speech databases to generate multiple voices, each with its features. By training the system on speech data from different speakers, it can be trained to produce different voices with different pitches, tones, and other features.

Strengthening intonation and prosody. A high-quality speech database helps the TTS system produce speech with natural-sounding intonation and prosody [7]. This speech output can be represented as a human, and its diagram is shown in the figure below (Fig. 5).

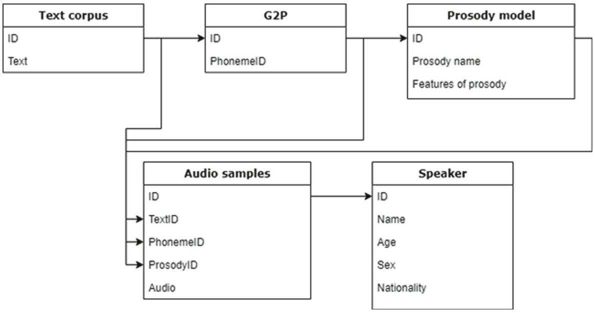


Fig. 5. The scheme of expressing speech like a human

In the absence of a speech database, a TTS system relies solely on rule-based algorithms to generate speech, resulting in less clear and less natural-sounding speech.

The creation of a speech database for the Karakalpak text-to-speech (TTS) synthesis system was carried out in the following main steps:

Choosing a speaker. When creating a speech database, it is very important to select a Karakalpak speaker, that is, a speaker when recording speech samples. It requires the speaker to have a clear, natural, and beautiful pronunciation and to read the text coherently and expressively. It is also important that the speaker have a popular and pleasant voice.

Recording speech samples. A chosen speaker is tasked with reading an extensive range of sentences in a dedicated room equipped with high-quality recording devices. The speech samples need to encompass diverse speech sounds, comprising vowels, consonants, and diphthongs, while also capturing various prosodic elements like stress, intonation, and rhythm. Speech samples should cover a variety of contexts, including different sentence types, word positions, and phonetic environments.

Transcription of speech samples. It is desirable to phonetically transcribe the recorded speech samples by a linguist. Phonetic transcriptions must represent the exact pronunciation of each speech sound in the sample.

Explanation. Speech samples must be annotated with additional linguistic information. For example, part-of-speech tags, prosodic cues, and contextual information can be interpreted. This information helps the TTS system generate more natural-sounding speech.

Initial processing. Pre-processing of recorded speech samples can also be performed to remove noise, normalize pitch, and adjust pitch and tempo. This helps to ensure that the speech samples in the database are of high quality and consistency [8–12].

TTS system training. The final step is to use the speech database to train the TTS system's speech synthesis model. Based on the patterns and features of the speech samples in the model database, it learns to match the corresponding speech sounds to the written text [7]. Then, through the TTS system, it can form high-quality, natural-sounding speech in the Karakalpak language. Building a speech database for the Karakalpak TTS synthesis system presents certain challenges. Some of these issues are outlined below.

Dialect and phonological diversity. There are several dialects of the Karakalpak language, each of which has its own phonological features. This can make it difficult to choose a single vowel that represents the entire language. In this case, it is required to record speech samples from several speakers representing different dialects and phonological features.

Limited resources for linguistic annotations. Linguistic annotations of speech samples are important in building a high-quality speech database. However, resources for this task may be limited, such as linguists or software tools.

Technical limitations. The recording equipment and software used to create a speech database must be of high quality and capture the subtle features of speech sounds. In addition, the TTS system itself may have technical limitations that affect the quality of speech output [18].

For the following reasons, it is necessary for the speaker who records the speech samples for the database to be fluent in his native language.

Naturalness. A native speaker of a language can produce speech that is natural and unique. They become relatively more familiar with important features of the language, including intonation, stress, and rhythm. This has a strong impact on the quality of speech generated by the TTS system.

Compatibility. Only a native speaker can provide coherent speech patterns that represent the language as a whole. They may also have consistent pronunciation and intonation. This allows the TTS system to provide natural-sounding speech output that is representative of the language.

Phonological accuracy. A native speaker is more likely to have a thorough understanding of the phonology of the language needed to correctly transcribe speech patterns. This TTS system serves to generate speech output that accurately reflects the pronunciation of the language.

Some of the important technical aspects of speech database recording depend on the equipment used, the recording environment, and the features of the recording technique used.

Equipment. The quality of the recording equipment used can significantly affect the quality of the speech samples obtained. High-quality microphones and recording software are required to capture high-definition recordings. For example, a condenser microphone with a flat frequency response can record a wide frequency range and create clear and detailed recordings.

Environment. The recording environment must be chosen wisely to minimize various surrounding noises and interferences. A soundproof booth or studio is usually used to achieve this. It is also desirable that the recording environment be free of echoes that may affect the quality of speech samples. For example, acoustic foam panels can be used to reduce reflections and echoes.

Techniques. Nowadays, many techniques can be used to obtain high-quality speech samples. For example, the speaker may be required to speak clearly and consistently with appropriate pauses between words and sentences [19].

Recording format. The format of the recorded speech samples also affects the quality of the speech database. Common formats used for speech databases include WAV, MP3, and FLAC. These formats have different relationships between file size and sound quality. For example, uncompressed formats such as WAV provide the highest resolution.

In addition to the technical aspects of creating a speech database for a TTS system, there are also linguistic aspects.

Speech style. To cover the speech styles used in regular conversation, a speech database must have a variety of styles, such as formal, informal, and casual. Diverse speech tempos and accents should be captured on audio.

Dialect. A speech database is also required to take into account the different dialects and accents used in natural language. To ensure the speech database encompasses various dialects, recordings can be collected from individuals residing in different regions.

Context. It is essential to gather speech samples from diverse settings such as phone conversations, public speaking, and group discussions. By incorporating a wide array of speech contexts into the database, the synthesized speech can be adapted to suit various scenarios effectively.

Feeling. A speech database must contain speech samples representing various emotions, including joy, anger, sadness, and fear. It allows natural expression to express different emotions in synthesized speech.

Dictionary. A speech database must also take into account the different vocabulary and sentence structures used in the target language. This can include technical jargon, slang, and colloquialisms. In the Karakalpak language, it is often necessary to use an explanatory dictionary. Because linguistic meanings, examples, and comments are given in explanatory dictionaries [20].

3 Results

The speakers read the texts at their natural pace and style in a quiet and closed environment, as well as in a noisy environment, and they were recorded. In this case, the speakers strictly followed the orthographic rules, the audio recordings of the female singers were sampled at 44.1 kHz and stored at 16 bits/sample. Male vocalists were recorded in a home studio, sampled at 48 kHz, and saved at 24bit/sample.

The resulting database consists of 19 h of audio material consisting of more than 34,000 segments. It took half a year to create the entire database, and the uncompressed data is more than 12 GB in size.

The database consists of information such as age, gender, work experience, and recording device of the speakers, and a Tacotron 2 system was used to demonstrate the use of the database. The subjective demonstration showed that the trained moles are suitable for practical use. For all listeners, the rating recordings were presented in the same order and one at a time. In this case, records were randomly selected at each stage (Table 1).

Table 1. Mean opinion score (MOS) results

System	Male	Female
Tacotron 2	4,15 ± 0,05	4,18 ± 0,08

Each listener was allowed to hear each audio recording only once, and the system was open to all listeners. Each recording was re-evaluated up to 10 times for female and male speakers. It was found that female speakers made more mistakes than male speakers.

4 Conclusions

The purpose of this article on building a speech database for a TTS system in Karakalpak is to demonstrate how to get around the challenges involved in building a voice database for TTS. The significance of a speech database in producing natural-sounding speech output is emphasized, and linguistic and technical factors including accommodating various speech contexts and styles and capturing distinctive phonemes are explored when recording a speech database.

Playing a pivotal role in developing a TTS synthesis system that meets the needs and expectations of the language's users and speakers would significantly aid in the preservation and promotion of the Karakalpak language.

The Karakalpak TTS synthesis system's speech database is a crucial part of it. Ensures that the output of the speech synthesizer is natural-sounding, clear, and appropriate for a range of situations. It is challenging to build a speech database for Karakalpak text-to-speech synthesis because there are few sources available and linguistic and technical considerations must be made. However, a high-quality speech database can be produced with careful planning and execution. As a result, more Karakalpak language speakers will be able to employ sophisticated synthesis systems.

There are numerous difficulties in developing a speech database for the Karakalpak text-to-speech technology. Since it contains several distinctive phonemes that are not found in other languages. Building a speech database with native speakers is crucial. Because only native speakers are capable of efficiently producing the necessary phonemes.

For the creation of a high-quality voice database, technical considerations including the choice of recording equipment, atmosphere, and data storage are equally crucial. The instruments utilized must be of the highest caliber and completely and precisely cover the linguistic quirks. An environment that is peaceful and devoid of outside noise is ideal to guarantee that speech samples are clear and useful.

Linguistic aspects such as ensuring the diversity of speech styles, contexts, dialects, and emotions are important in creating a speech database that accurately represents the Karakalpak language. It is natural and necessary to form a synthesized speech suitable for different scenarios.

References

1. Taylor, P.: Text-to-Speech Synthesis. Cambridge University Press, Cambridge (2009)
2. Nwakanma, I., Oluigbo, I., Izunna, O.: Text–to–speech synthesis (TTS). Int. J. Res. Inf. Technol. **2**(5), 154–163 (2014)
3. Onaolapo, J., Idachaba, F.E., Badejo, J.A., Odu, T., Adu, O.I.: A Simplified Overview of Text-To-Speech Synthesis (2014)
4. Oliveira, L.C., Paulo, S., Figueira, L., Mendes, C., Nunes, A., Godinho, J.: Methodologies for designing and recording speech databases for corpus based synthesis. In: International Conference on Language Resources and Evaluation (2008)
5. Chalamandaris, A., Karabetsos, S., Tsiakoulis, P., Raptis, S.: A unit selection text-to-speech synthesis system optimized for use with screen readers. IEEE Trans. Consum. Electron. **56**(3), 1890–1897 (2010). https://doi.org/10.1109/TCE.2010.5606343
6. Gahlawat, M., Malik, A., Bansal, P.: Natural speech synthesizer for blind persons using hybrid approach. Procedia Comput. Sci. **41**, 83–88 (2014). ISSN 1877–0509, https://doi.org/10.1016/j.procs.2014.11.088
7. Nagy, P., Németh, G.: Improving HMM speech synthesis of interrogative sentences by pitch track transformations. Speech Commun. **82**, 97–112 (2016). C (September 2016), https://doi.org/10.1016/j.specom.2016.06.005
8. Mamatov, N.S., Niyozmatova, N.A., Abdullaev, S.S., Samijonov, A.N., Erejepov, K.K.: Speech recognition based on transformer neural networks. In: 2021 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, pp. 1–5 (2021). https://doi.org/10.1109/ICISCT52966.2021.9670093

9. Mamatov, N., Niyozmatova, N., Samijonov, A.: Software for preprocessing voice signals. Int. J. Appl. Sci. Eng. **18**, 2020163 (2021). https://doi.org/10.6703/IJASE.202103_18(1).006
10. Narzillo, M., Abdurashid, S., Parakhat, N., Nilufar, N.: Automatic speaker identification by voice based on vector quantization method. Int. J. Innov. Technol. Explor. Eng. **8**(10), 2443–2445 (2019). https://doi.org/10.35940/ijitee.J9523.0881019
11. Wiedecke, B., Narzillo, M., Payazov, M., Abdurashid, S.: Acoustic signal analysis and identification. Int. J. Innov. Technol. Explor. Eng. **8**(10), 2440–2442 (2019). https://doi.org/10.35940/ijitee.J9522.0881019
12. Narzillo, M., Abdurashid, S., Parakhat, N., Nilufar, N.: Karakalpak speech recognition with CMU sphinx. Int. J. Innov. Technol. Explor. Eng. **8**(10), 2446–2448 (2019). https://doi.org/10.35940/ijitee.J9524.0881019
13. Niyozmatova, N.A., Mamatov, N.S., Tulyaganova, Sh. A., Samijonov, A.N., Samijonov, B. N: Methods for determining speech activity of Uzbek speech in recognition systems. In: AIP Conference Proceedings, vol. 2789(1), p. 050019, 23 June 2023. https://doi.org/10.1063/5.0145438
14. Mamatov, N.S., Niyozmatova, N.A., Yuldoshev, Y.S., Abdullaev, S.S., Samijonov, A.N.: Automatic speech recognition on the neutral network based on attention mechanism. In: Zaynidinov, H., Singh, M., Tiwary, U.S., Singh, D. (eds.) Intelligent Human Computer Interaction. IHCI 2022. Lecture Notes in Computer Science, vol. 13741, pp. 100–108. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-27199-1_11
15. Mamatov, N.S., Niyozmatova, N.A., Samijonov, A.N., Samijonov, B.N.: Construction of language models for Uzbek language. In: 2022 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, pp. 1–4 (2022). https://doi.org/10.1109/ICISCT55600.2022.10146788
16. Niyozmatova, N.A., Mamatov, N.S., Otaxonova, B.I., Samijonov, A.N., Erejepov, K.K.: Classification based on decision trees and neural networks. In: 2021 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, pp. 01–04 (2021). https://doi.org/10.1109/ICISCT52966.2021.9670345
17. Babomuradov Ozod, O.J., Mamatov, N.S., Boboev, L.B., Otaxonova, B.: Text documents classification in Uzbek language. Int. J. Recent Technol. Eng. **8**(2 Special Issue 11), 3787–3789(2019). https://doi.org/10.35940/ijrte.B1493.0982S1119
18. Saratxaga, I., Navas, E., Hernáez, I., Luengo, I.: Designing and recording an emotional speech database for corpus based synthesis in Basque. In: Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006), Genoa, Italy. European Language Resources Association (ELRA) (2006)
19. Iida, A., Campbell, N.: Speech database design for a concatenative text-to-speech synthesis system for individuals with communication disorders. Int. J. Speech Technol. **6**, 379–392 (2003). https://doi.org/10.1023/A:1025761017833
20. Esemuratova, R., Kalendarova, M., Japaqova, R.: Qaraqalpaq tiliniń túsindirme sózligi. Qaraqalpaqstan baspası. Nókis (1982)

Approaches to Solving Problems of Markov Modeling Training in Speech Recognition

D. T. Muxamediyeva[1](✉), N. A. Niyozmatova[1], R. A. Sobirov[1], B. N. Samijonov[2], and E. Kh. Khamidov[1]

[1] Digital Technologies and Artificial Intelligence, "Tashkent Institute of Irrigation and Agricultural Mechanization Engineers" National Research University, Tashkent, Uzbekistan
dilnoz134@rambler.ru

[2] Sejong University, Seoul, South Korea

Abstract. The article discusses approaches to solving problems of learning Markov modeling in speech recognition. A Markov process is a stochastic process consisting of a sequence of random states, where the probability of transition from one state to another depends only on the current state and does not depend on previous states. The result of observing such a process is a sequence of states that the system goes through during the observation period. The task of model training is considered the most difficult when using Markov models in recognition systems, since there is no known unique and universal way to solve it, and the quality of recognition depends on the result of model training. Therefore, special attention must be paid to training the model.

Keywords: genetic algorithm · training · automatic speech recognition · speech · Hidden Markov Models · Mel-cepstral coefficients · ant colony algorithm · immune algorithm

1 Introduction

At present, speech recognition technologies are intensively developing. There are many areas where speech technologies can find effective practical applications, such as medicine, banking, telephony, robotics, the automotive industry, forensics, etc. [1].

Currently, special attention is paid to the following promising areas of speech technologies:

1. Speech recognition for human-machine interfaces: such technologies are used in smart homes, voice assistants, automotive systems, mobile and other devices where voice control is required.
2. Speech synthesis technologies: These technologies are used to create an artificial voice, for example, for people with a speech impairment.
3. Systems that determine the personality of a speaker can be used for two main purposes: verification and identification.

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 344–357, 2024.
https://doi.org/10.1007/978-3-031-56728-5_29

Speaker verification systems are used to verify a person's identity. Typically, such systems are used to authorize users to access computer and information systems, as well as to control access. In this case, the system checks whether there is a voice in a pre-created voice database that the system expects to hear. Speaker identification systems, on the other hand, allow the identification of a person. Such systems are used in areas such as military affairs, communications, forensics, etc. The above-mentioned systems use methods and algorithms for speech recognition and signal processing, which make it possible to extract unique characteristics of the voice, such as frequencies, intonations, and rhythms [2–11]. One such method is Hidden Markov Models (HMMs).

HMMs are a mathematical tool for modeling data sequences where the next state depends only on the current state and is independent of previous states. HMMs can be used for speech recognition, where data sequences represent speech commands and model states correspond to speech phonemes [12].

The task of model training is one of the key and complex tasks in Markov modeling. Model training is based on the existing training data set to create a statistical model that will reflect the dependencies between the states of the system and its input data. The learning problem can be solved using the Baum-Welch algorithm, which uses the maximum likelihood method to estimate model parameters and evolutionary algorithms. Such algorithms make it possible to determine the optimal values for the model parameters, maximizing the probability of obtaining an observed sequence for a given model [13].

Let, given a system that at an arbitrary point in time can be in one of N different states S_1, S_2, ..., S_N.

At discrete times $t = 1, 2, ...$ The system passes from one state to another, but can remain in the existing state (the current state at time t will be described as q_t) [3–5]. In this case, the transitions are carried out in accordance with a certain probability matrix, described as

$$p_{ij} = p\left[q_t = S_j | q_{t-1} = S_i\right], \quad 1 \leq i, j \leq N. \tag{1}$$

In this case, p_{ij} has the following properties:

$$p_{ij} \geq 0, \tag{2}$$

$$\sum_{j=1}^{N} p_{ij} = 1. \tag{3}$$

A Markov process is a stochastic process consisting of a sequence of random states, where the probability of transition from one state to another depends only on the current state and does not depend on previous states. The result of the observation of such a process is a sequence of states that the system goes through during the observation period [14].

Having a process model in the form of a probability matrix $A = p_{ij}$ and an initial state matrix $\Pi_i = p[q_i = S_i]$, $1 \leq i \leq N$, one can calculate the probability of any sequence of states, i.e. any observation [15].

In practice, the so-called hidden Markov models are most common. They are used to model sequences of observed data that are the result of some hidden process. In this case, the hidden process is modeled as a sequence of states that change over time in accordance with the probabilistic transitions between them. Each state of the hidden process generates some observation, which becomes visible to the observer. The probabilities of transitions between states and the probabilities of generating observations are determined by the parameters of the model, where they can be estimated from the observed data using training methods. In such models, the observed events are probabilistic functions of the real state of the system. In Hidden Markov Models, the observed events are called "observable variables" or "outputs", and the states of the system not directly observed are called "hidden variables" or “internal states”. The output data is the result of some process, depending on the internal state of the system. The probability of observing each specific sequence of output data for a given sequence of internal states is determined using conditional probabilities [16].

Typically, Markov models can be described by the following sets of parameters [17]:

1) N is the number of model states.
2) M is the number of different observables in each state by the symbol (i.e. the size of the alphabet) $V = \{V_1, V_2, \ldots V_N\}$..
3) $A = \{p_{ij}\}$ - matrix of transition probabilities, where

$$p_{ij} = p\big[q_{t+1} = S_j | q_t = S_i\big], \quad 1 \le i, j \le N. \tag{4}$$

4) $G = \{g_j(k)\}$ - probability distributions of observed symbols in state j, where

$$g_j(k) = p[v_k \big| q_t = S_j], \quad 1 \le j \le N, \quad 1 \le k \le M. \tag{5}$$

5) $\Pi = \{\Pi_i\}$ is the probability of each initial state, where

$$\Pi_i = P[q_1 = S_i], \quad 1 \le i \le N. \tag{6}$$

By setting the values of the parameters N, M, A, G and Π for Markov models, it can be used as a sequence generator.

To generate a sequence with a given number of elements, you must specify the initial state of the system (that is, one of the possible states that the hidden process can take) and sequentially generate new states and corresponding observable events. This process is carried out in the following steps [18]:

1. Set the initial state of the system by selecting one of the possible states of the hidden process.
2. Using the matrix of transition probabilities A, select the next state of the system based on the current state.
3. Using the probability matrix G, generate an observable event corresponding to the current state.
4. Repeat steps 2 and 3 the specified number of times (until the required sequence length is reached).

Based on a given algorithm, it is possible to generate a sequence of observed events corresponding to a given statistical model. The resulting sequence is used to test speech recognition algorithms.

In addition, when using HMM for pattern recognition, it is necessary to solve the following three problems [19]:

1. **The task of training the model.** The HMM parameters are determined based on the training set. To determine the parameters, it is recommended to use the EM-algorithm (expectation-maximization algorithm).

1.1. Initialization of model parameters (transition_matrix, emission_matrix and initial_distribution). Model parameters in the context of HMM describe the probabilities of transitions between states and the probabilities of symbol emission in each state.

The transition matrix determines the transition probabilities between model states. If the HMM has *N* states, then the transition matrix will be *NxN* in size. Each element (*i, j*) of the matrix represents the probability of transition from state *i* to state *j*. It is important that the sum of the probabilities in each row (or column) of the transition matrix be equal to 1, since a state must always transition to one of the states (including itself).

The emission matrix determines the probabilities of observing symbols for each state. If the HMM has *N* states and *M* possible symbols (for example, in a speech recognition problem, *M* can be equal to the number of different sounds), then the emission matrix will be $N \times M$ in size. Each element (*i, k*) of the matrix represents the probability of observing character *k* while in state *i*.

The initial distribution determines the probabilities of the initial states of the model. If the HMM has *N* states, then the initial distribution will be a vector of size *N*. Each element *i* of the vector represents the probability that the model will start in state *i*.

They must be normalized so that the sum of the rows in the matrices is 1.

1.2. E-step (Expectation step):

The Forward-Backward algorithm is implemented to calculate the posterior probabilities.

Forward algorithm:

Initialization of the alphas array with zero values.

Calculate the first element of alphas[0] as the product of initial_distribution and emission_matrix for the first observation.

For each next time step *t*:

alphas[*t*] is calculated as the product of the emission_matrix for the current observation and the scalar product of alphas[$t-1$] and transition_matrix.

Backward algorithm:

Initializing the betas array with units.

The last element of betas[num_observations - 1] is calculated to be 1.

For each time step *t* from num_observations - 2 to 0:

Calculate betas[*t*] as the product of the emission_matrix for the next observation times the scalar product of betas[$t + 1$] and the transition_matrix.

Calculation of gamma and xi:

posterior probabilities of states (gamma) by normalizing the product of alphas and betas for each time step;

posterior transition probabilities (xi) for each time step using alphas, betas, transition_matrix and emission_matrix.

1.3. M-step (Maximization Step):

The model parameters are updated to maximize the expected likelihood:

Initial Distribution Update:

initial_distribution as average gamma values for the first time steps.

Transition matrix update:

each transition_matrix element as the sum of the xi values for the transition from state *i* to state *j* divided by the sum of the gamma values for state i (excluding the last time step).

Emissions matrix update:

each element of the emission_matrix as the sum of the gamma values for state *i* when the observation matches symbol *k*, divided by the sum of the gamma values for state *i*.

1.4. Repetition of E-step and M-step:

Steps E and M are repeated for a certain number of iterations or until parameter changes are negligible.

1.5. Model Rating:

After reaching the stopping criterion, evaluate the trained model parameters transition_matrix, emission_matrix and initial_distribution.

2. The problem of estimating the probability of observation. Let the HMM and the observed sequence be given. It is necessary to determine the probability of observing a given sequence for a given model. To solve this problem, you can use the Forward-Backward algorithm, which is carried out in the following steps:

2.1. Initialization:

Model parameters are set: transition_matrix, emission_matrix, and initial_distribution.

Assigning an observed sequence of characters to observations.

2.2. Forward pass:

The alphas array of size (num_observations, num_states) is initialized with zeros.

The first element of alphas[0] is computed as the product of initial_distribution and the corresponding emission probability from the emission_matrix for the first observation observations[0].

For each time step *t* from 1 to num_observations - 1:

For state *j*:

Alphas[*t,j*] is calculated as the sum of the products of alphas[*t-1, i*], transition_matrix[i, j] and emission_matrix[*j*, observations[*t*]] for all states *i*.

2.3. Backward pass:

The betas array is initialized with a size of (num_observations, num_states) units.

Assigned to the last element betas[num_observations - 1] 1.

For each time step *t* from num_observations - 2 to 0:

For state *i*:

Betas[*t, i*] is calculated as the sum of the products of transition_matrix[*i, j*], emission_matrix[*j*, observations[$t + 1$]] and betas[$t + 1, j$] for all states *j*.

2.4. Observation Probability Calculation:

The overall probability of the observation is calculated as the sum of the probabilities in the last time step alphas[-1].

2.5. Result output

Output the probability of observation obtained in the previous step.

It should be noted that the algorithm uses two passes: forward and reverse. The forward pass computes alphas (forward passes), which represent the probabilities of being in a particular state at each time step, given observations up to that point. The back pass computes betas (back passes), which represent the probabilities of remaining in the state after observing all subsequent symbols. After performing the backward pass, the total probability of the observation is calculated as the sum of the probabilities in the last time step alphas[-1]. This value represents the desired probability of observing a given sequence given a hidden Markov model.

3. The task of decoding. Let the HMM and the observed sequence be given. It is necessary to determine the most probable sequence of hidden states. To solve this problem, the Viterbi algorithm is used, which is carried out in the following steps:

3.1. Parameter initialization:

The number of hidden states (num_states) and the number of possible observations (num_symbols) are determined.

The transition matrix (transition_matrix) is initialized with the probabilities of transitions between hidden states, the emission matrix (emission_matrix) with the probabilities of generating observations from hidden states, and the initial distribution (initial_distribution) with the probabilities of initial states.

3.2. Viterbi algorithm:

Input data: a sequence of observations (observations).

The dp matrix is initialized to store the highest probabilities and the prev_state matrix to store the previous states and the first time step of the dp matrix with initial probabilities.

Based on time steps and states, the highest probabilities and previous states are calculated according to the Viterbi algorithm.

The path is decoded starting from the last time step based on the prev_state matrix.

3.3. Result output:

The decoded path of hidden states (decoded_states) obtained by the Viterbi algorithm is displayed.

When using Markov models, training with test sequences is time consuming. It should give a method for finding model parameters such as to maximize the probability of a sequence of observations matching a given model. Consider an iterative procedure for choosing model parameters [12].

In the procedure of iterative refinement of the parameters of Markov models, it is necessary to determine the probability of the system being in state S_i at time t and in state S_j at time $t + 1$ for given model parameters and a sequence of observations, i.e.

$$\xi_t(i,j) = p(q_t = S_i,\ \ q_{t+1} = S_j | O, \lambda) \tag{7}$$

In direct and inverse variables, the probability is defined as

$$\xi_t(i,j) = \frac{\alpha_t(i) p_{ij} g_j(O_{t+1}) \beta_{t+1}(j)}{p(O|\lambda)} = \frac{\alpha_t(i) p_{ij} g_j(O_{t+1}) \beta_{t+1}(j)}{\sum\limits_{i=1}^{N} \sum\limits_{j=1}^{N} \alpha_t(i) p_{ij} g_j(O_{t+`1}) \beta_{t+1}(j)}. \tag{8}$$

Expressing $\gamma_t(i)$ in terms of summation over $\xi_t(i, j)$, we get:

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(ij). \tag{10}$$

Based on the above formulas for refining Markov models, the following can be written:

Expected frequency of being in state S_i at time $t = 1$:

$$\overline{\pi_i} = \gamma_1(i); \tag{11}$$

$$\overline{p_{ij}} = \frac{\sum\limits_{t=1}^{T-1} \xi_t(i, j)}{\sum\limits_{t=1}^{T-1} \gamma_t(i)}; \tag{12}$$

$$\sum_{t=1}^{T} \gamma_t(j), \quad \overline{g_j}(k) = \frac{O_t = v_k}{\sum\limits_{t=1}^{T} \gamma_t(j)} \tag{13}$$

Certain parameters are most likely to represent a given sequence of observations.

One of the main problems in HMM training is the problem of getting into the local optimum in the process of model optimization. This can cause the model optimization to stop at a loss function value that is not the global optimum, and therefore the model will not have the best recognition accuracy.

To overcome this problem, many approaches have been proposed so far. One is to use a stochastic optimization method such as Markov Chain Monte Carlo (MCMC) or Stochastic Gradient Descent (SGD). These methods make it possible to get out of the local optimum and provide a wider search for the parameter space.

In addition, global optimization methods such as genetic algorithms or ant colony based global optimization methods or other evolutionary algorithms can be used. Such methods also make it possible to avoid local optimal.

To do this, at the first stage of the evolutionary algorithm, an initial population of solutions is randomly generated, which are various combinations of model parameter values. Then the quality of each solution is evaluated using an evaluation function, which can be, for example, the likelihood function of the model. Next comes the selection of the most adapted solutions (the best). The crossover operator combines parts of two parent solutions to create new offspring, and the mutation operator randomly changes the parameter values in the new solution.

The evolutionary algorithm repeats the process of creating new generations of the population until a given stopping criterion is reached (for example, reaching the maximum value of the evaluation function or reaching a certain number of iterations).

The use of evolutionary algorithms for HMM training has a number of advantages, such as the possibility of finding a global optimum and reducing the dependence on starting parameters. And also, the evolutionary algorithm allows you to optimize not only the parameters of the model, but also the choice of training sequences, which can improve the quality of recognition.

2 Results

HMM-based speech recognition requires the formation of informative speech features. In literatures [20–31], interesting works have been carried out on the formation of features and the identification of informative features. To form features in this work, MFCC coefficients were used to represent speech in the form of a vector of features, which are obtained as a result of the Fourier transform of temporal speech signals presented in the form of a spectrogram into a Mel-frequency representation. MFCC coefficients take into account the peculiarities of the perception of sounds by the human ear, since the chalk-frequency representation better matches the perception of high and low frequencies. Next, the obtained coefficients are transformed using cepstral analysis, which helps to highlight important signal characteristics and reduce the dimensionality of the data.

The number of features (MFCC coefficients) depends on the specific task and may be different. Typically 12 to 20 MFCCs are used, although more may be used in some cases. The number of MFCC coefficients can be chosen experimentally based on the quality of speech recognition on test data.

The algorithm for extracting MFCC coefficients consists of the following steps:

The audio signal is split into overlapping frames.

Applies a window function (such as Hamming) to each frame.

The Fourier transform is applied to the windowed frame.

The magnitude spectrum is calculated for each frame.

Mel filters are calculated to convert the magnitude spectrum into a Mel spectrum.

Mel filters are applied to the magnitude spectrum.

The logarithm of the resulting Mel spectrum is calculated.

A discrete cosine transform (DCT) to the logarithm of the Mel spectrum is performed.

The output of this algorithm is the MFCC coefficients, which are represented as vectors for each frame.

```
[[0.7422303 0.5532607 0.68049966… 0.95503917 0.32147816]
[0.0786073 0.9745327 0.52009533 … 0.71336791 0.79554746]
[0.0726853 0.0393421 0.02926857 … 0.85601778 0.20057767]
[0.5490482 0.3676609 0.15398193 … 0.29524076 0.48544261]
[0.4670125 0.2525328 0.21170367 … 0.76408712 0.64643327]]
```

At the next stage of speech recognition, the creation of HMM and training is carried out. Below is the result of this step:

```
Trained Transition Matrix:
[[8.58531871e−01 1.41467622e−01 … 5.06242437e−07]
[4.01853960e−10 6.00927764e−01 … 3.99072235e−01]
[2.71153867e−01 1.12751763e−09 … 7.28846132e−01]]
Trained Emission Matrix:
[[3.74174969e−01 1.84769465e−01 … 4.15451791e−20 4.41055566e−01]
[5.09872773e−06 6.52057991e−01 … 2.85676830e−01 6.22600806e−02]
[1.64715419e−01 2.93611122e−01 … 3.31506967e−01 2.10166492e−01]]
Trained Initial Distribution:
[1.14160670e−96 1.00000000e+00 … 7.05398021e−48]
```

This step uses the EM (Expectation-Maximization) algorithm, which is a powerful and commonly used technique in machine learning. It has several advantages that make it useful for solving various problems, since the EM algorithm allows you to work with data in which some information is missing or hidden (for example, observable and hidden variables). This can be useful when training models when certain variables cannot be directly observed. The EM-algorithm seeks to maximize the likelihood function, which allows obtaining the most probable model parameters for a given sample, can be used for data clustering, as well as for training models of a mixture of distributions, which makes it useful for analyzing unstructured data and is resistant to the presence of outliers in the data, since it is based on a probabilistic model and averages the influence of different observations. The algorithm works iteratively, improving the current solution at each iteration. This allows reaching local optima and achieving convergence. The EM algorithm can be applied even if there is not enough data to fully train the model, as it uses an E-step to estimate latent variables and an M-step to update the parameters.

However, it has some drawbacks: the EM algorithm is sensitive to initial approximations of the parameters. Different initial conditions can lead to different local optima. This can make it difficult to obtain globally optimal model parameters. The EM algorithm does not guarantee convergence to the global optimum. It can get stuck at local extremes even if there is a global optimum. At each iteration of the EM algorithm, it is required to calculate the latent variable expectations (step E) and update the model parameters (step M). This can be computationally expensive, especially for complex models or large amounts of data. The EM algorithm requires multiple repetitions of steps E and M until convergence. In some cases, convergence may take a long time or even not be achieved. The EM algorithm uses optimization methods that often require the likelihood function to be differentiable. This limits its applicability to models for which it is difficult to calculate derivatives. If the initial parameters of the model are poorly chosen or the model is too complex, the EM algorithm may converge to local extrema where the model likelihood is close to zero. The EM algorithm does not always work well if the data has a complex structure, for example, when there are outliers that are out of the general distribution.

The next step is to estimate the observation probability based on the Forward-Backward algorithm. This algorithm allows you to calculate the probability of observing a sequence of characters, as well as the distribution of hidden states in each time step. One of the key advantages of the algorithm is its ability to estimate the probability of observing a particular sequence of characters given a model. The algorithm can be used to decode hidden state sequences based on the observed sequence. This is important in tasks related to the recognition, decoding or classification of sequences. The forward-backward algorithm can be used to train HMM parameters based on observed data. It can help find optimal parameter values that maximize the likelihood of observing data.

The next step is decoding. Where it is required to determine the most probable sequence of hidden states. For this, the Viterbi algorithm is used - this is an efficient method for decoding hidden states in HMM. It has several advantages that make it an important tool in the analysis of data sequences. The Viterbi algorithm finds the optimal hidden state sequence that most likely resulted in the observed data sequence. This makes it indispensable in tasks where it is required to choose the best explanation for the observed data. It has linear complexity with respect to the length of the observed sequence and the number of states in the model. This allows fast and efficient decoding of large amounts of data. The Viterbi algorithm can be used for various types of hidden Markov models, including discrete and continuous state models, as well as mixed models. The implementation of the Viterbi algorithm is relatively simple and can be done with a relatively small amount of code. This makes it available for practical implementation in various applications. The Viterbi algorithm works in a local context, choosing the optimal state based on local transition probabilities. This reduces the amount of computation and memory required for decoding.

Finally, a genetic algorithm can be used to optimize model parameters. There are various optimization methods that can be used to solve problems, including HMM learning problems.

Methods based on mathematical calculations, which use mathematical calculations to find the optimal solution. They can be directional (e.g. gradient descent) or non-directional (e.g. coordinate descent). Directional methods involve determining the direction of the function's steepest decay and then moving in that direction to find a local extremum. For example, the gradient descent method uses the gradient (derivative) of a function to determine the descending direction.

Enumeration methods are based on enumeration of all possible solutions in order to find the optimal one. These methods can be quite slow and inefficient for large search spaces, but they are guaranteed to find the optimal solution. An example would be the brute force method, when all possible combinations of parameters are checked.

Methods that use the element of randomness include a random element in the optimization process. An example would be simulated annealing, which starts with a large random variance and then reduces it gradually, allowing the system to "cool" and find an optimum over a narrower range.

When considering the problem of HMM learning, methods based on mathematical calculations can be problematic due to the complex structure of loss functions and differentiability constraints. Enumeration methods can be inefficient due to the high dimensionality of the search space. Methods that use randomness can be helpful in avoiding getting stuck in local optima.

The genetic algorithm can be used to optimize the number of hidden states in the SMM, which can improve the accuracy of speech recognition. The HMM training problem is a global optimization problem, since it is required to find the optimal model parameters that provide the best fit to the observed data. Genetic algorithms are capable of performing global searches and can help avoid getting stuck in local optima. The HMM training task may involve tuning a large number of parameters, such as initial state probabilities, transition matrices, and emission functions. Genetic algorithms are well suited for working with high-dimensional problems and combinatorial spaces. Genetic algorithms do not require differentiability of the objective function, which is an advantage, since the likelihood functions are distorted in the HMM learning problem and do not always have analytical derivatives. Genetic algorithms can be applied to various types of HMMs such as Hidden Markov Models with continuous or discrete states and various emission functions. Genetic algorithms can be easily adapted for parallel computing, which makes it possible to speed up the HMM learning process, and can be adapted to work with constraints on model parameters, which is important in the HMM learning task with limited resources. HMM training objectives can have a variety of structures and requirements. Genetic algorithms can be applied to solve complex, unstructured problems where other optimization methods may be less effective.

The resulting GA solution: Trained Transition Matrix:

[[0, 96 0,04 … 0]
[0 0,92 … 0]
…………………..
[0 0 … 1]]

The disadvantages of existing HMM learning methods are their strong dependence on initial conditions and the inability to find a global extremum. There is no single best method for teaching HMM, and different approaches can be used for this task. One of such approaches considered in the work is genetic algorithms (GA). Genetic algorithms are optimization methods based on the principles of natural evolution. They are a promising and actively developed direction in the field of optimization of multicriteria problems.

3 Conclusions

The use of HMM for speech recognition is based on the following assumptions:

All speech can be represented as a sequence of sounds or phonemes. The probability of each phonemic state depends only on the previous state, not on earlier states. The sound wave of speech contains some level of noise and ambiguity, and HMM allows you to model this noise in order to improve recognition accuracy. HMMs use statistical methods to estimate model parameters based on a large amount of training data and determine the most likely sequence of states for a given speech wave. Building a model for speech recognition requires large amounts of computational resources and time for training, as well as parameter optimization to ensure maximum recognition accuracy.

The advantages of genetic algorithms in solving the HMM learning problem are that genetic algorithms work with several points in the search space at the same time. This avoids getting stuck in local optima and improves the probability of finding globally optimal HMM parameters. Genetic algorithms operate on coded representations of parameters, which can be chosen to better fit the characteristics of the problem. This gives greater flexibility and adaptability of the algorithm to a specific task and work with parameter codes regardless of their interpretation. This allows the algorithm to be applied to different types of parameters and tasks. Multipoint search reduces the chance of getting stuck in local optima. Genetic algorithms contribute to the exploration of the solution space, which is especially important for problems with many local extrema and work based only on information about the objective function at an arbitrary point and the range of acceptable parameter values. This simplifies the process of tuning the algorithm and its application to various problems.

References

1. Ognev, I.V.: Preliminary processing of a speech signal for building a database of pronunciations of single words. In: Ognev, I.V., Paramonov, P.A. (eds.) Information tools and Technologies: tr. XX International Science-Technical Conference, pp. 53–58. MPEI, Moscow (2012)

2. Popov, E.V.: Communication with computers in natural language, 2nd edn. Stereotypical, 360 p. Editorial URSS, Moscow (2004)
3. Niyozmatova, N.A., Mamatov, N.S., Tulyaganova, Sh.A., Samijonov, A.N., Samijonov, B.N.: Methods for determining speech activity of uzbek speech in recognition systems. In: AIP Conference Proceedings, vol. 2789, no. 1, p. 050019 (2023). https://doi.org/10.1063/5.0145438
4. Mamatov, N.S., Niyozmatova, N.A., Yuldoshev, Y.S., Abdullaev, S.S., Samijonov, A.N.: Automatic speech recognition on the neutral network based on attention mechanism. In: Zaynidinov, H., Singh, M., Tiwary, U.S., Singh, D. (eds.) IHCI 2022. LNCS, vol. 13741, pp. 100–108. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-27199-1_11
5. Mamatov, N.S., Niyozmatova, N.A., Samijonov, A.N., Samijonov, B.N.: Construction of language models for Uzbek language. In: 2022 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, pp. 1–4 (2022). https://doi.org/10.1109/ICISCT55600.2022.10146788
6. Niyozmatova, N.A., Mamatov, N.S., Otaxonova, B.I., Samijonov, A.N., Erejepov, K.K.: Classification based on decision trees and neural networks. In: 2021 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, pp. 01–04 (2021). https://doi.org/10.1109/ICISCT52966.2021.9670345
7. Mamatov, N.S., Niyozmatova, N.A., Abdullaev, S.S., Samijonov, A.N., Erejepov, K.K.: Speech recognition based on transformer neural networks. In: 2021 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, 2021, pp. 1–5 (2021). https://doi.org/10.1109/ICISCT52966.2021.9670093
8. Mamatov, N., Niyozmatova, N., Samijonov, A.: Software for preprocessing voice signals. Int. J. Appl. Sci. Eng. **18**, 2020163 (2021). https://doi.org/10.6703/IJASE.202103_18(1).006
9. Narzillo, M., Abdurashid, S., Parakhat, N., Nilufar, N.: Automatic speaker identification by voice based on vector quantization method. Int. J. Innov. Technol. Explor. Eng. **8**(10), 2443–2445 (2019). https://doi.org/10.35940/ijitee.J9523.0881019
10. Wiedecke, B., Narzillo, M., Payazov, M., Abdurashid, S.: Acoustic signal analysis and identification. Int. J. Innov. Technol. Explor. Eng. **8**(10), 2440–2442 (2019). https://doi.org/10.35940/ijitee.J9522.0881019
11. Narzillo, M., Abdurashid, S., Parakhat, N., Nilufar, N.: Karakalpak speech recognition with CMU sphinx. Int. J. Innov. Technol. Explor. Eng. **8**(10), 2446–2448 (2019). https://doi.org/10.35940/ijitee.J9524.0881019
12. Mosleh, M.: FPGA implementation of a linear systolic array for speech recognition based on HMM. In: Mosleh, M., Setayeshi, S., Mehdi Lotfinejad, M., Mirshekari, A. (eds.) The 2nd International Conference on Computer and Automation Engineering (ICCAE), vol. 3, pp. 75–78 (2010)
13. Ikonin, S.Yu., Sarana, D.V.: SPIRIT ASP engine automatic speech recognition system. Digital Signal Processing (2003)
14. Sapunov, G.V., Trufanov, F.A.: Genetic algorithms as a method for optimizing hidden Markov models in problems of speech recognition. In: Information Technologies in Computer Systems. Issue 3. Under the general editorship of prof. Azarova V.N. MIEM, Moscow (2004)
15. Marczyk, A.: Genetic Algorithms and Evolutionary Computation (2004). http://www.talkorigins.org/faqs/genalg/genalg.html
16. Komarov, A.N.: Basic cellular ensembles of associative oscillatory environments and the possibility of their expansion. In: Komarov, A.N., Ognev, I.V., Podolin, P.B. (eds.) Computational systems and information processing technologies: Interuniversity. Sat. scientific tr. – Issue, vol. 5, no. 30, 200 p. Inf.-ed. center of PGU, Penza (2006)
17. Ognev, I.V.: Character recognition in an associative oscillatory environment. In: Ognev, I.V., Podolin, P.B. (eds.) News of Higher Educational Institutions. Volga region. Ser. Technical Science, no. 6, pp. 55–66 (2006)

18. Elliott, L., Ingham, D., Kyne, A., Mera, N., Pourkashanian, M., Whittaker, S.: Efficient clustering-based genetic algorithms in chemical kinetic modelling. In: Deb, K. (ed.) GECCO 2004. LNCS, vol. 3103, pp. 932–944. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-24855-2_106
19. Sastry, K., O'Reilly, U.M., Goldberg, D.E.: Population sizing for genetic programming based upon decision making. IlliGAL Report No. 2004028 (2004)
20. Samijonov, A., Mamatov, N., Niyozmatova, N.A., Yuldoshev, Y., Asraev, M.: Gradient method for determining non-informative features on the basis of a homogeneous criterion with a positive degree. In: IOP Conference Series: Materials Science and Engineering, vol. 919, no. 4 (2020). https://doi.org/10.1088/1757-899X/919/4/042011
21. Mamatov, N., Niyozmatova, N.A., Samijonov, A., Juraev, S., Abdullayeva, B.: The choice of informative features based on heterogeneous functionals. In: IOP Conference Series: Materials Science and Engineering, vol. 919, no. 4 (2020). https://doi.org/10.1088/1757-899X/919/4/042009
22. Mamatov, N.S., Samijonov, A.N., Yuldoshev, Y., Khusan, R.: Selection the informative features on the basis of interrelationship of features. In: Techno-Societal 2018 - Proceedings of the 2nd International Conference on Advanced Technologies for Societal Applications, vol. 2, pp. 121–129 (2020). https://doi.org/10.1007/978-3-030-16962-6_13
23. Fazilov, S., Mamatov, N., Samijonov, A., Abdullaev, S.: Reducing the dimensionality of feature space in pattern recognition tasks. J. Phys. Conf. Ser. **1441**(1), 012139 (2020). https://doi.org/10.1088/1742-6596/1441/1/012139
24. Mamatov, N., Samijonov, A., Niyozmatova, N.: Determination of non-informative features based on the analysis of their relationships. J. Phys. Conf. Ser. **1441**(1), 012149 (2020). https://doi.org/10.1088/1742-6596/1441/1/012149
25. Niyozmatova, N.A., Mamatov, N., Samijonov, A., Mamadalieva, N., Abdullayeva, B.M.: Unconditional discrete optimization of linear-fractional function "-1"-order. In: IOP Conference Series: Materials Science and Engineering, vol. 862, no. 4, p. 042028 (2020). https://doi.org/10.1088/1757-899X/862/4/042028
26. Niyozmatova, N.A., Mamatov, N., Samijonov, A., Rahmonov, E., Juraev, S.: Method for selecting informative and non-informative features. In: IOP Conference Series: Materials Science and Engineering, vol. 919, no. 4 (2020). https://doi.org/10.1088/1757-899X/919/4/042013
27. Fazilov, S., Mamatov, N.: Formation an informative description of recognizable objects. J. Phys.: Conf. Ser. **1210**(1) (2019). https://doi.org/10.1088/1742-6596/1210/1/012043
28. Mamatov, N., Samijonov, A., Yuldashev, Z.: Selection of features based on relationships. J. Phys. Conf. Ser. **1260**(10), 102008 (2019). https://doi.org/10.1088/1742-6596/1260/10/102008
29. Shavkat, F., Narzillo, M., Abdurashid, S.: Selection of significant features of objects in the classification data processing. Int. J. Recent Technol. Eng. **8**(2 Special Issue 11), 3790–3794 (2019). https://doi.org/10.35940/ijrte.B1494.0982S1119
30. Mamatov, N., Samijonov, A., Yuldashev, Z., Niyozmatova, N.: Discrete optimization of linear fractional functionals. In: 2019 15th International Asian School-Seminar Optimization Problems of Complex Systems, OPCS 2019, pp. 96–99 ((2019)). https://doi.org/10.1109/OPCS.2019.8880208
31. Shavkat, F., Narzillo, M., Nilufar, N.: Developing methods and algorithms for forming of informative features' space on the base K-types uniform criteria. Int. J. Recent Technol. Eng. **8**(2 Special Issue 11), 3784–3786 (2019). https://doi.org/10.35940/ijrte.B1492.0982S1119
32. Nagy, P., Németh, G.: Improving HMM speech synthesis of interrogative sentences by pitch track transformations. Speech Commun. **82C**(September 2016), 97–112 (2016). https://doi.org/10.1016/j.specom.2016.06.005

33. Daridi, F., Kharma, N., Salik, J.F.N.: Parameterless genetic algorithms: review and innovation. IEEE Can. Rev. (47) (2004)
34. Aida-ZadeK, R.: Investigation of combined use of MFCC and LPC features in speech recognition systems. Aida-Zade, K.R., Ardil, C., Rustamov, S.S. (eds.) World Acadamic of Science, Engineering and Technology (2006)
35. Noisy channel model (2020). https://en.wikipedia.org/wiki/Noisy_channel_model. Accessed 12 Apr 2020
36. Watanabe, S., et al.: Hybrid CTC. Attention Archit. End-to-End **11**(8), 1240–1253 (2017)
37. Hannun "Sequence Modeling with CTC". Distill (2017). https://distill.pub/2017/ctc/
38. Graves, A., Fernandez, S., Gomez, F., Schmidhuber, J.: Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In: ICML (2006)
39. Chan, W., et al.: Listen, attend and spell. arXiv:1508.01211 (2015)
40. Amodei, D., et al.: Deep Speech2: end-to-end speech recognition in English and Mandarin. arXiv:1512.02595 (2016)
41. Zeghidour, N., Usunier, N., Synnaeve, G., Collobert, R., Dupoux, E.: End-to-end speech recognition from the raw waveform. arXiv:1806.07098 (2018)

Spatio-Angular Resolution Trade-Off in Face Recognition

Muhammad Zeshan Alam[1](✉), Sousso kelowani[2], and Mohamed Elsaeidy[3]

[1] Brandon University, Brandon, MB, Canada
zeshusher@gmail.com
[2] University of Quebec at Trois Rivieres, 3351 Bd des Forges, Trois-Rivieres, Canada
[3] Istanbul Medipol University, Istanbul, Turkey

Abstract. Ensuring robustness in face recognition systems across various challenging conditions is crucial for their versatility. State-of-the-art methods often incorporate additional information, such as depth, thermal, or angular data, to enhance performance. However, light field-based face recognition approaches that leverage angular information face computational limitations. This paper investigates the fundamental trade-off between spatio-angular resolution in light field representation to achieve improved face recognition performance. By utilizing macro-pixels with varying angular resolutions while maintaining the overall image size, we aim to quantify the impact of angular information at the expense of spatial resolution, while considering computational constraints. Our experimental results demonstrate a notable performance improvement in face recognition systems by increasing the angular resolution, up to a certain extent, at the cost of spatial resolution.

Keywords: Face recognition · light field imaging · spatio-angular resolution · deep learning · Convolution neural networks

1 Introduction

Face recognition system are being rapidly adopted in wide range of applications including, surveillance, entertainment, and forensic etc. specially because it is a widely acceptable universal biometric modality. However, face recognition still encounter various challenges in a completely unconstrained environment involving significant variations in illumination, pose, emotions, and occlusions.

Recently, deep learning models have contributed significantly in the success of computer vision in general and face recognition in particular. Despite the tremendous potential of the deep learning models the aforementioned challenging scenario limits their performance. This is because learning based model obtain relevant features through training process in which the model weights are updated as the algorithm is fed the labeled data. The training data is typically a regular camera images which is heavily effected by these challenging environmental conditions and lose some of the spatial information critical in face discrimination.

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 358–369, 2024.
https://doi.org/10.1007/978-3-031-56728-5_30

Light field (LF) imaging, allows us to capture the additional angular information of the light which is lost in conventional imaging. This is usually achieved by separately recording the intensities of light rays coming from different directions are at each separate pixel position [1–3]. This additional angular information turns out to be a critical discriminatory factor for various classification tasks. For example Light fields used for face recognition have shown significant performance improvement [4,5], particularly when combined with deep learning-based light field recognition techniques.

The performance improvement of the LF based face recognition methods however, involve additional computational cost either as a pre-processing step (posteriori refocusing or depth computation) or due to some modification of network architecture in case of learning based methods [5], depending on the LF representation used in the method. The overhead involved in the processing cost of the additional angular information is one of the major limitation of the LF based techniques, however, acquiring the angular information also presents its own set of problems.

Light field imaging systems can be implemented in a variety of ways including, camera arrays [6,7], coded mask LF camera designs, [8,9], and microlens array (MLA) [2,10] based cameras. In the camera array based approach, each camera separately captures the scene from a different perspective, at a high spatial resolution, however, camera arrays are expensive and have restricted mobility. An alternative approach for the light field acquisition is the coded mask-based camera design. There are several design approaches, including, single shot, single mask, multiple shot, and multiple mask-based camera designs, irrespective of the design coded masks methods have poor light efficiency. MLA-based LF cameras are the most cost-effective and easy to handle LF acquisition system. However, since a single sensor is used to capture both spatial and angular information, a spatio-angular resolution trade-off exit in all MLA based designs [11,12].

To minimize the additional complexity associated with the angular information, we study the effect of this spatio-angular resolution trade-off on the performance of the learning based face recognition models. For this purpose, we increase the angular resolution at the cost of the spatial resolution while maintaining the overall size of the input image. We introduce the super pixels of various sizes (angular resolution), formed by combining pixels from multiple perspective images and shown in Fig. 3, for training the state of the art CNN based face recognition network and evaluate their performance for different angular resolution input images. The major contributions of this paper are summarized as follows:

1. Quantitative evaluation of the spatio-angular resolution trade-off on the performance of face recognition
2. Application of a compact light field representation in face recognition for improved prediction accuracy.

The paper is organized as follows. In Sect. 2, we present related work on face recognition, particularly the light field-based CNN approaches in face recognition. We explain the application of the light field to the face recognition problem in Sect. 3. Light field representations for training deep learning models are presented in Sect. 4. In Sect. 5 we detail the training process and hyper-parameters. In Sects. 6 and 7 we present the experimental result and conclude the paper respectively.

2 Related Work

Traditional face recognition methods primarily focus on analyzing 2D images and extracting discriminative facial features for identification. Eigenfaces [13] introduced the concept of Principal Component Analysis (PCA) for face representation. Fisherfaces [13] extended this approach with Linear Discriminant Analysis (LDA) to improve discriminability. Local Binary Patterns (LBP) [14] provided an effective texture-based feature descriptor for robust face recognition.

Deep learning-based face recognition techniques, like other computer vision tasks, are becoming state-of-the-art. Numerous studies have focused on developing deep neural network architectures and training methodologies to improve recognition performance. In FaceNet [15] a siamese network is trained with triplet loss to learn highly discriminative face embeddings. VGGFace [16] presented a deep convolutional neural network (CNN) trained on a large-scale face dataset, demonstrating impressive recognition accuracy. ArcFace [17] introduced a novel loss function based on angular margin to enhance discriminative feature learning. This method explicitly optimizes the angular margin between different classes, making the learned embeddings more separable. In CosFace [18] margin-based loss function is proposed hat introduces an additive cosine margin to the standard softmax loss. By directly maximizing inter-class variations and minimizing intra-class variations, CosFace achieved improved discrimination of face embeddings. The approach demonstrated robustness against variations in lighting, pose, and expressions. Circle Loss by Sun et al. [19] introduced a new loss function based on angular similarity. By utilizing pairwise angular similarity, the Circle Loss optimizes the feature space to form compact and well-separated clusters. This approach achieved competitive performance and demonstrated the capability to handle large-scale face recognition tasks.

Cross-domain face recognition, dealing with recognizing faces across different domains or modalities, has gained attention in recent years [20,21]. [20] Addresses the problem of pose-invariant face recognition through a two-stream deep neural network architecture. The method incorporates both spatial and temporal information to capture pose variations in face images. By jointly modeling appearance and motion cues, the approach achieved robust recognition performance across different pose angles. In [21] a generative adversarial network (GAN)-based approach for thermal-to-visible face recognition is proposed. The method utilizes a conditional GAN framework to generate synthetic visible face images from thermal images. By learning the mapping between thermal and visible domains, the approach improved recognition performance in cross-modal face recognition scenarios.

Zhang et al. [4] proposed a light field-based approach for face recognition combining depth and intensity information. The integration of intensity and depth cues allowed for improved robustness against variations in pose, illumination, and occlusion, leading to enhanced face recognition performance. In [5] a multi-modal light field representation that combines depth and color cues for robust face recognition is introduced. This incorporation of depth and color cues was aimed to enhance the system's robustness to challenging conditions. In [22], a novel two-long short-term memory(LSTM) cell architecture that leverages spatial and angular information from a light field image is proposed. The new two-cell network is able to jointly learn from multiple sequences of light field perspective images simultaneously, targeting to create richer and more effective models for the face recognition task. A double-deep spatio-angular learning framework for light field-based face recognition, which is able to model both the intra-view/spatial and inter-view/view/angular information using two deep networks in the sequence is proposed in [23]. The proposed framework includes a LSTM recurrent network, whose inputs are VGG-Face descriptions, computed using a VGG-16 convolutional neural network (CNN).

3 Light Field for Improved Face Recognition

Light field imaging allows for the capture of angular information from light rays originating from a scene point. Conventional cameras, as depicted in Fig. 1(a), record incident light rays directly at the sensor, resulting in the loss of directional information. In contrast, light field cameras, such as the MLA-based cameras [2] and [10], shown in Fig. 1(b), incorporate a micro-lens array (MLA) adjacent to the sensor. This arrangement enables the separate recording of light ray intensities passing through different points (sub-apertures) of the main lens and converging at the micro-lens in front of specific pixels. Consequently, light field cameras preserve the directional information of the captured light rays.

The MLA in light field cameras, as illustrated in Fig. 1(b), receives light rays reflected from scene points at varying depths with different incident angles [24]. This angular difference leads to disparities in the multiple perspective images obtained from decoding the light field. By tracing the light rays back to the scene, it becomes possible to calculate the corresponding depths of these points. Estimating the depth of all captured points generates a depth map that reveals the depth variations among different objects in the scene. Exploiting this depth information is valuable for facial analysis, as facial features like the nose, ears, and cheekbones exhibit substantial depth variations across individuals. Leveraging this depth information can assist face recognition in challenging conditions, such as pose changes and occlusions, enhancing its performance.

4 Spatio-Angular Resolution Trade-Off

Light fields can be represented in many ways, including, sequence of perspective images, lenslet representation, EPIs, and plenoptic function. Irrespective of the LF representation, incorporating the additional angular information results in the fundamental spatial-angular resolution trade-off. Typically a high spatial-resolution face image contains fine details, such as the texture of the skin, shape of facial features, and subtle patterns. These details are crucial for accurately distinguishing and identifying individuals. They provide more information for the face recognition algorithm to analyze and match against reference images.

Face recognition algorithms usually extract various facial features, such as the distance between the eyes, shape of the nose, or curvature of the lips, to create a unique face template or facial signature. With higher spatial-resolution images, these features can be extracted more accurately and reliably, leading to better recognition performance. While the high spatial resolution is beneficial, other factors, such as lighting conditions, pose, and partial occlusion, also play crucial roles in accurate face recognition. These aforementioned challenging factors can affect the fine details captured in the high-resolution images resulting in the loss of sufficient discriminative features. However, these spatial features, when complemented with features extracted from angular information such as facial features depth variation and Iris reflection etc, could provide robust face recognition cues in challenging conditions. Current State-of-the-art face recognition methods adopt deep CNN models that process images pixel-wise, meaning that each pixel is processed individually or in small receptive fields. Including both spatial and angular information in an input image means more pixels in the input image. With an increase in the number of pixels, the number of computations required to process the image also increases. More importantly, CNNs heavily rely on convolutional operations, that obtain relevant features through training using layers that consist of multiple kernels. These kernels are updated as the algorithm is fed labeled data, converging by numerical optimization methods on the weights that best match the training data [25]. With a larger input image size, the convolutional operations performed in each layer increases, which results in an increase in the number of multiplications and additions required for the convolutional operations, leading to higher computational complexity. The size of the output feature map after the convolution operation in each layer depends on three factors: the size of the input image, the size of the kernel, and the stride. The stride specifies the step size at which the kernel moves across the input image.

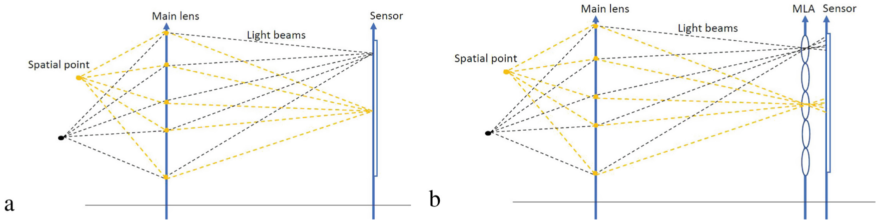


Fig. 1. An illustration of MLA-based light field camera image acquisition in comparison with conventional camera model. (a) Conventional camera model. (b) MLA-based light field camera capturing scene points at different depths.

Light fields which are typically represented as a set of up to one hundred perspective images with standard spatial resolution would render any real-time surveillance system computationally prohibitive and could also exhaust the memory in embedded systems. Other light field representations discussed below would also suffer from computational cost issues due to separate pixels recording of spatial and angular information:

4.1 Perspective or Sub-aperture Image Representation

A sub-aperture representation refers to a subset of images extracted from the captured light field data. In addition to the spatial information of the scene, Sub-aperture images contain angular information of the light as they are captured using an array of micro-lenses or camera units that are positioned at different viewpoints. Each micro-lens or camera unit captures a subset of the rays coming from various directions within the scene, which corresponds to different angular perspectives. To capture the light field, the micro-lenses or camera units are arranged in a grid pattern, covering the sensor or image plane. Each micro-lens or camera unit captures a small portion of the incoming light rays, corresponding to a specific angular perspective. By combining the captured data from all the micro-lenses or camera units, a complete light field is obtained. In the resulting sub-aperture images, each image represents a specific viewpoint or perspective within the light field. The angular information is encoded in the differences between these viewpoints. By analyzing the parallax or disparity between the sub-aperture images, it is possible to extract depth information and estimate the relative distances of various facial features, which serve a robust cue in face identification under challenging conditions.

4.2 Epipolar Plane Image Representation

In the Epipolar Plane Image (EPI) representation, the light field data is arranged in a specific format that emphasizes the epipolar geometry. Epipolar lines are the lines of sight connecting corresponding points in different viewpoints. By plotting these lines and their associated image intensities, an EPI is formed. To create an EPI, one dimension of the EPI corresponds to the spatial information, typically representing the pixel rows or columns, while the other dimension corresponds to the angular information, representing the viewpoints or perspectives. Each point in the EPI corresponds to the intensity of a specific pixel in the captured sub-aperture images. By examining the EPI, it is possible to observe the disparities or shifts in intensity values along the epipolar lines, which relate to the depth or parallax information in the scene. However, it should be noted that EPI is an alternative arrangement of the pixels that allows better visualization of the depth but does not address the spatial-angular trade-off and therefore suffer from high computational complexity issues.

4.3 Lenslet Representation

The raw lenslet light field representation refers to the raw sensor data captured by the individual micro-lenses. The raw lenslet light field representation consists of a grid of sub-images or patches, where each sub-image corresponds to the image captured by a specific micro-lens. A lenslet is an individual component that captures a subset of light rays, while a sub-aperture image is whole resulting image obtained from the captured light field data. The lenslet image actually represents a magnified view of the captured scene.

To create the lenslet image representation of the light field, the lenslet images are arranged in a grid pattern, similar to the arrangement of the micro-lenses. The lenslet images are combined to form a larger mosaic that represents the entire light field. However, like a set of perspective images or EPI representation, lenslet image representation also suffers from the spatio-angular resolution trade-off.

4.4 Proposed Representation

In this section, we introduce a lenslet-inspired representation of the light field to explore the significance of both spatial and angular information in the context of face recognition under difficult conditions. To achieve this, we modify the middle perspective image by replacing individual pixels with macro-pixels that contain multiple perspectives of the same scene point. To maintain the overall size of the original input image we drop the same number of neighboring spatial pixels of the pixel that is replaced by the macro-pixel. Considering that the decoded image has dimensions of 625 × 434 pixels, we construct macro-pixels as multiples of 2 pixels, as illustrated in Fig. 3. The selection of neighboring perspective pixels included in each macro-pixel depends on their distance from the middle perspective. For instance, a 2 × 2 macro-pixel encompasses the four adjacent perspective image pixels. By increasing the size of the macro-pixels, we prioritize the contribution of angular information in the light field representation, albeit at the expense of spatial information. This enables us to investigate the significance of angular cues in comparison to spatial features for distinguishing individual faces, particularly under challenging conditions.

5 Training

We trained two different CNNs namely GoogleNet and Resnet50 and modified the final layers to match the fifty class output of the [22] dataset. For GoogleNet, we removed the ‘loss3-classifier’, ‘prob’, ‘output’ layers and added a fully connected(FC) layer for 50 classes, a softmax, and a classification layer. For Resnet50 we replace the fully connected, ‘softmax’, and ‘classification’ layers with the corresponding layer for 50 classes.

Both CNNs are pre-trained on ImageNet dataset [26] and transfer learning is performed to adapt and fine-tune pre-trained Googlenet and Resnet50 for the face recognition problem. The IST-EURECOM Light Field Face Database (LFFD) [22] consists of LF face images captured by a Lytro ILLUM camera and is used here for performance assessment purposes. The IST-EURECOM LFFD includes 4000 LF images, captured from 50 subjects, in two separate acquisition sessions with a temporal separation between 1 and 6 months. Each session contains 20 LF images per subject with different facial variations including facial expressions, poses, illuminations, and occlusions, as illustrated in Fig. 2. We have separated on type of facial variations LF images of all 50 subjects for testing and the rest of the LF images constitute the training dataset.

The networks performances are robust under different training hyper-parameters configurations. Various combinations of the learning rate, optimizer, and batch size, etc. are tested but the variations in performance are negligible. The batch size turned out to be the most influential hyper-parameter in terms of performance improvement for both the networks described above. Increasing the batch size improved the performance and the maximum batch size supported due to memory limitations is 35.

Among different configurations, the following set of hyperparameters resulted in the best performance and therefore, adopted in this work. The initial learning rate is set to 3e-4 with a learning rate drop of 0.1 and a learning rate drop period of 20. Although the choice of optimizer has shown a marginal effect on the overall performance, However,

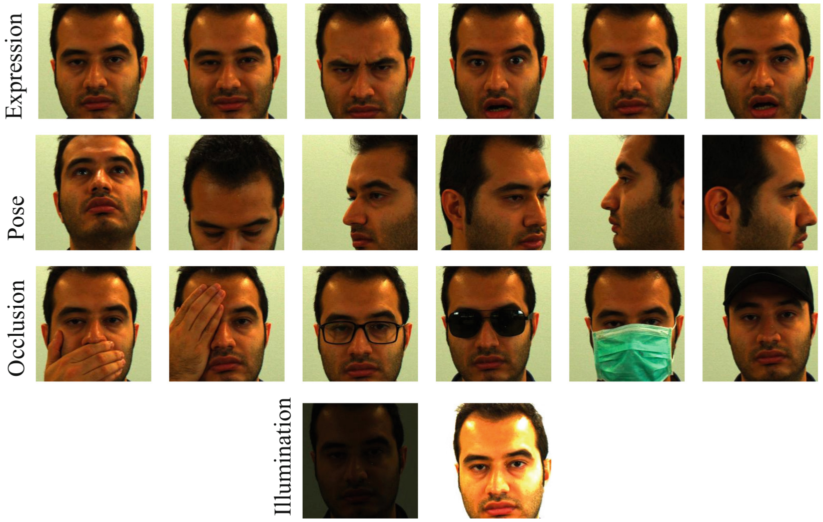


Fig. 2. Visualization of the four different categories of facial variations in the data-set.

adam optimizer has slightly improved the prediction and hence selected in our training. The batch size is set to 32 and the training is performed for 50 and 100 epochs for Resnet50 and Googlenet respectively.

6 Experimental Results

In this section, we present the results of the spatial-angular resolution trade-off in face recognition. We provide the accuracy of prediction by categorizing the various facial variation into four distinct groups which typically pose challenges in face recognition.

In Fig. 4, a quantitative comparison of the proposed macro-pixel LF representation at different angular resolutions using Resnet50 architecture is presented. It can be seen in Fig. 4 that macro-pixel representation outperforms regular 2D image-based face recognition for almost all types of facial variations. However, there is no uniform increase in the performance based on the increase in the angular resolution. On the other hand with the significant increase in the macro-pixel size, the impact of decrease in the spatial resolution on the overall performance tends to dominate and causes a drop in the prediction accuracy. Additionally, for more challenging scenarios such as pose variation, illumination, and occlusion the benefit of angular resolution is more evident than the regular two-dimensional images and for more commonly encountered scenarios like change in facial expression, there is no additional benefit of angular resolution after the network convergence.

Several deep CNN architectures designed for image classification exist in the literature. The performance of these CNN models is influenced by many design choices, for example, the number of trainable parameters, Depth, and width of architecture [27]. To demonstrate the robustness of the angular information in lieu of spatial information the results for a different CNN architecture Googlenet are presented in Fig. 5.

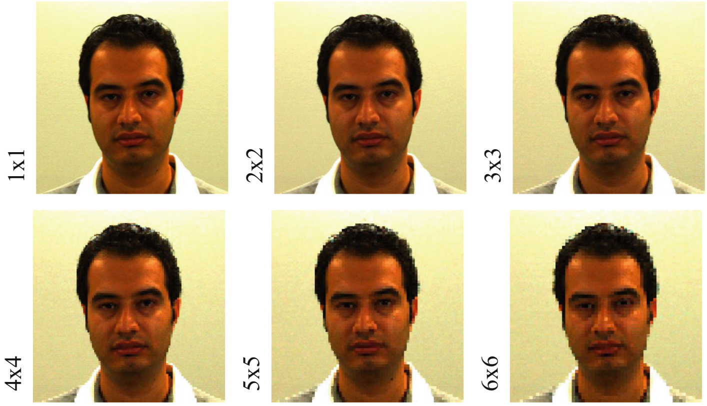


Fig. 3. Proposed image representation for Spatio-angular resolution trade-off investigation. Angular resolution varies from 1 × 1 to 6 × 6.

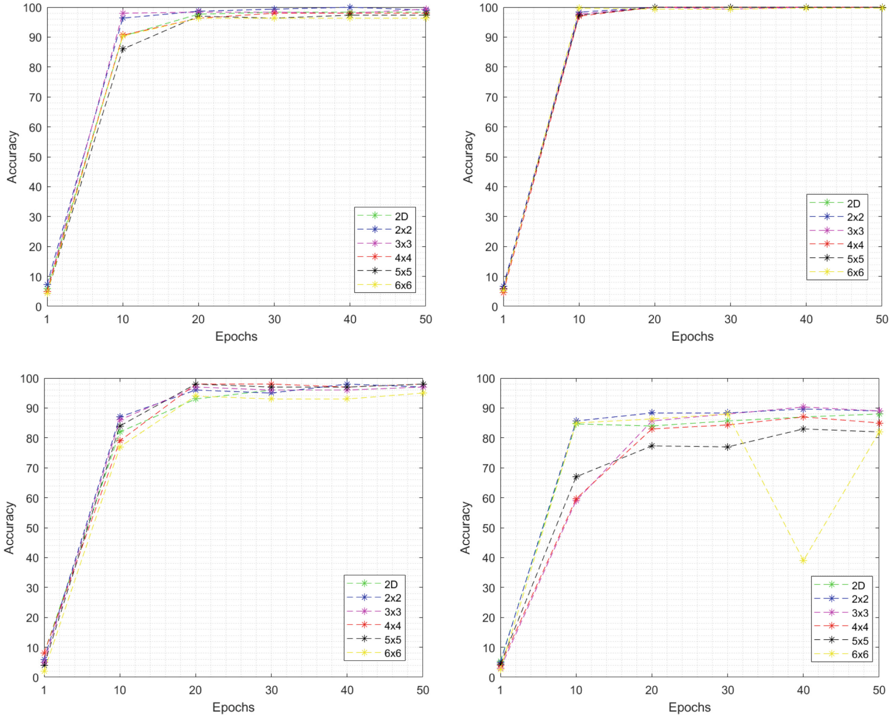


Fig. 4. Performance comparison of the spatial-angular resolution trade-off with different angular resolutions macro-pixels. Top Left) Occlusion. Top Right) Expression. Bottom Left) Illumination. Bottom Right) Pose.

Resnet50 consists of 50 layers, designed to tackle the challenge of training very deep networks effectively. It introduces a concept called residual connections, where shortcut

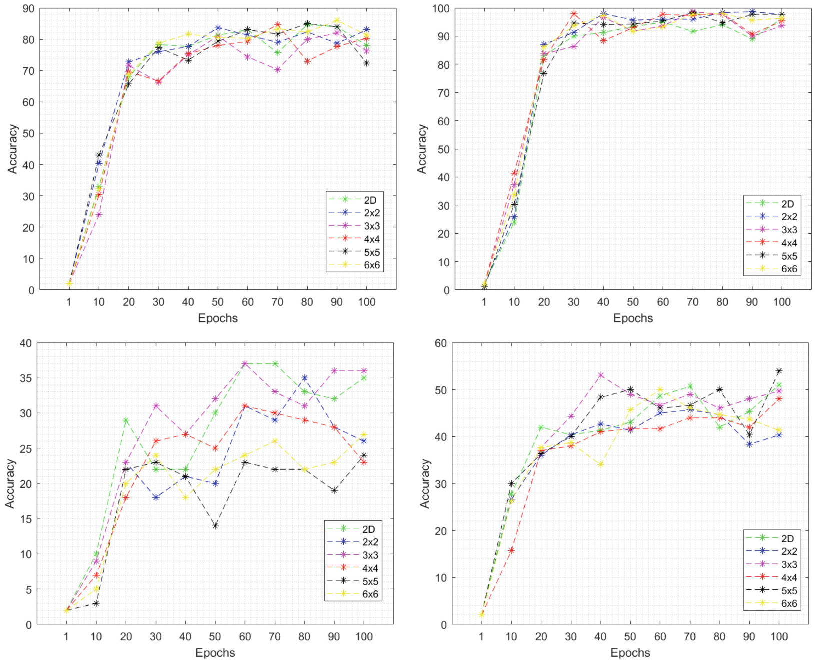


Fig. 5. Performance comparison of the spatial-angular resolution trade-off with different angular resolutions macro-pixels. Top Left) Occlusion.Top Right) Expression. Bottom Left) Illumination. Bottom Right) Pose.

connections are added to skip over certain layers, allowing the model to learn residual functions.The ResNet50 architecture has approximately 23.5 million trainable parameters. GoogleNet, on the other hand, is 22 layers deep Architecture, with 27 pooling layers included. The overall architecture includes 9 linearly stacked inception modules, with a total of seven million trainable parameters.

Although the two networks are widely different in their design approach, in Fig. 4, it can be seen that the angular information proves to be robust in the aforementioned challenging scenarios despite the loss of spatial information. However, in Fig. 5, an increase in the macro-size pixel seems also to benefit the more common scenario of regular facial expression variation but has no apparent effect on the illumination variation condition. Additionally, Resnet50 tends to be sensitive to the loss of spatial information and therefore, encounters a loss in performance for large macro-pixel sizes of 5×5 and 6×6 but google net, shows an overall gain in the performance for high angular resolution.

7 Conclusion

This paper delves into the trade-off between spatial and angular resolution in face recognition, focusing on the lenslet-inspired macro-pixel representation. The proposed

LF representation showcased performance improvements when compared to conventional image-based methods. However, it becomes apparent that there exists a threshold for increasing angular resolution while sacrificing spatial resolution. Furthermore, it is important to highlight that even with notable advancements in face recognition under normal conditions, the proposed LF representation can offer advantages in challenging scenarios without increasing computational costs.

References

1. Levoy, M., Hanrahan, P.: Light field rendering. In: ACM International Conference on Computer Graphics and Interactive Techniques, pp. 31–42 (1996)
2. Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M., Hanrahan, P.: Light field photography with a hand-held plenoptic camera. In: Stanford University Computer Science Technical Report CSTR (2005)
3. Alam, M.Z., Gunturk, B.K.: Hybrid stereo imaging including a light field and a regular camera. In: Signal Processing and Communication Application Conference, pp. 1293–1296. IEEE (2016)
4. Sepas-Moghaddam, A., Correia, P.L., Nasrollahi, K., Moeslund, T.B., Pereira, F.: Light field based face recognition via a fused deep representation. In: 2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP), pp. 1–6 (2018)
5. Wang, Z., Zhang, X., Li, S.: Multi-modal light field representation for robust face recognition. In: IEEE Access, pp. 1280–1288 (2019)
6. Wilburn, B., et al.: High performance imaging using large camera arrays. ACM Trans. Graph. **24**, 765–776 (2005)
7. Yang, J.C., Everett, M., Buehler, C., McMillan, L.: A real-time distributed light field camera. In: Eurographics Workshop on Rendering, pp. 77–86 (2002)
8. Alam, M.Z., Gunturk, B.K.: Deconvolution based light field extraction from a single image capture. In: IEEE International Conference on Image Processing, pp. 420–424 (2018)
9. Ashok, A., Neifeld, M.A.: Compressive light field imaging. In: SPIE Defense, Security, and Sensing, vol. 7690 (2010)
10. Alam, M.Z., Gunturk, B.K.: Hybrid light field imaging for improved spatial resolution and depth range. Mach. Vis. Appl. **29**, 11–22 (2018)
11. Alam, M.Z., Gunturk, B.K.: Light field extraction from a conventional camera. Sign. Proc. Image Commun. **109**, 116845 (2022)
12. Shabbir, J., Alam, M.Z., Mukati, M.U.: Learning texture transformer network for light field super-resolution. arXiv preprint, arXiv:2210.09293 (2022)
13. Turk, M., Pentland, A.: Eigenfaces for recognition. J. Cogn. Neurosci. **3**(1), 71–86 (1991)
14. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **28**(12), 2037–2041 (2006)
15. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: a unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 815–823 (2015)
16. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition (2015)
17. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: Arcface: additive angular margin loss for deep face recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4690–4699 (2019)
18. Wang, H., et al.: Cosface: Large margin cosine loss for deep face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5265–5274 (2018)

19. Sun, Y., et al.: Circle loss: a unified perspective of pair similarity optimization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6398–6407 (2020)
20. Zheng, L., Wang, S., Liu, X., Tian, Q.: Pose-invariant face recognition with multi-view deep representation. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 2290–2302 (2020)
21. Zhang, T., Wiliem, A., Yang, S., Lovell, B.: TV-GAN: generative adversarial network based thermal to visible face recognition. In: 2018 International Conference on Biometrics (ICB), pp. 174–181. IEEE (2018)
22. Sepas-Moghaddam, A., Etemad, A., Pereira, F., Correia, P.L.: Long short-term memory with gate and state level fusion for light field-based face recognition. IEEE Trans. Inf. Forens. Secur. **16**, 1365–1379 2021
23. Alireza, S., Haque, M.A., Correia, P.L., Nasrollahi, K., Moeslund, T.B., Pereira, F.: A double-deep spatio-angular learning framework for light field-based face recognition. IEEE Trans. Circuits Syst. Video Technol. **30**(12), 4496–4512 (2020)
24. Wahab, A., Alam, M.Z., Gunturk, B.K.: High dynamic range imaging using a plenoptic camera. In: Signal Processing and Communications Applications Conference, pp. 1–4 (2017)
25. Alam, M.Z., Kelouwani, S., Boisclair, J., Amamou, A.: Learning light fields for improved lane detection. IEEE Access **11**, 271–283 (2023)
26. Krizhevsky, A., Hinton, G.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
27. Alam, M.Z., Ates, H.F., Baykas, T., Gunturk, B.K.: Analysis of deep learning based path loss prediction from satellite images. In: Signal Processing and Communications Applications Conference, pp. 1–4. IEEE (2021)

Bridging the Gap Between Technology and Farming in Agri-Tech: A Bibliometric Analysis

Fatma Serab Onursal[1] and Sabri Öz[2](✉)

[1] İstanbul Medipol University, İstanbul, Turkey
[2] İstanbul Commerce University, İstanbul, Turkey
soz@ticaret.edu.tr

Abstract. Agri-tech, or the application of technology to agriculture, has the power to transform farming methods and find solutions to the problems the industry faces. With an emphasis on comprehending the significance of numerous disciplines, including Artificial Intelligence (AI), this study presents a bibliometric analysis that attempts to analyze the trends and important disciplines engaged in bridging the gap between technology and farming in agri-tech. The analysis highlights the growing interest and research activity in agri-tech and its related fields by looking at a wide range of academic publications. The results show that AI is becoming a key discipline in agri-tech, with an increase in publications highlighting its potential to boost production, improve resource management, and support sustainable farming. The analysis emphasizes the necessity for interdisciplinary cooperation among researchers, practitioners, policymakers, and farmers in order to close the gap between technology and farming. The agriculture sector may unleash the potential of cutting-edge technologies, resulting in more effective, sustainable, and fruitful farming techniques, by utilizing AI and encouraging interdisciplinary cooperation. The conclusions drawn from this bibliometric analysis lay the groundwork for additional agri-tech research and innovation, opening the door to a revolutionary future for agriculture.

Keywords: Artificial Intelligence · AI · Agriculture · Precision Farming · Emerging Technologies

1 Introduction

In order to sustain the world's expanding population and ensure global food security, the agricultural industry is essential. Traditional farming methods, however, confront a variety of difficulties, including a lack of resources, environmental issues, and the desire for higher output. The development of artificial intelligence (AI) in recent years has created new opportunities for changing agriculture and bridging the divide between technology and farming. Innovative solutions are being created to address these issues and revolutionize the agricultural landscape by utilizing AI technologies, such as machine learning, image recognition, robots, and data analytics.

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 370–382, 2024.
https://doi.org/10.1007/978-3-031-56728-5_31

This study paper's goal is to examine the status quo and potential developments for AI applications in agriculture, with a particular emphasis on bibliometric analysis. We intend to examine the scholarly landscape and trends within the subject of AI in agriculture by utilizing bibliometric analytic approaches, looking at pertinent scholarly publications, authors, and organizations. This method offers a thorough and organized grasp of the body of existing knowledge, highlighting important areas of study, significant authors, and new developments.

Significant prospects for increasing productivity, maximizing resource allocation, and assuring sustainability exist with the integration of AI in agriculture. One of the most well-known uses of AI in farming is precision farming, which enables farmers to make data-driven decisions by carefully monitoring and maintaining each individual crop and animal. Farmers can use artificial intelligence (AI) algorithms to analyze data from a variety of sources, including soil sensors, drones, and satellite imaging, in order to improve irrigation, fertilizer use, and pest management. Crop yields are increased, environmental impact is decreased, and resource efficiency is increased as a result.

Crop yield prediction is an essential use of AI in agriculture. AI systems can predict crop yields with high accuracy using historical data, weather patterns, and machine learning algorithms. This enables farmers to organize their operations, improve logistics, and make wise financial decisions. This skill boosts production while also helping to prevent food shortages and maintain a steady supply chain.

Systems for AI-driven disease diagnosis are also revolutionizing the agricultural sector. These systems may detect illnesses, pests, and nutritional deficiencies in crops at an early stage, allowing for prompt intervention and avoiding yield losses. They do this through picture recognition and machine learning algorithms. A more targeted and effective use of pesticides and resources, as well as a reduction in the environmental impact and an improvement in overall plant health, are all influenced by the automated detection and diagnosis of plant diseases.

We aim to find important insights into the development of this discipline as we explore the research landscape of AI in agriculture through bibliometric analysis. We want to get a thorough overview of the current body of knowledge and identify opportunities for additional investigation by looking at publication trends, significant authors, leading research institutions, and key research issues. The growth and incorporation of AI technologies in agriculture will be facilitated by this analysis, which will assist in identifying new trends, research gaps, and prospective partnership opportunities.

In conclusion, the application of AI to agriculture has enormous potential for overcoming the difficulties the industry is now facing. Innovative solutions to boost productivity, improve resource use, and assure sustainability are being developed by integrating AI technologies including machine learning, image recognition, robots, and data analytics. This study article uses bibliometric analysis to present a thorough review of the state of AI in agriculture and its potential future orientations, opening the door for informed decision-making, teamwork, and agri-tech developments.

2 Agriculture and Technology

This part of the study focusing mainy on the reason of chosing agriculture and technoogical and digital transformation on agriculture. Also the gap in between the two disciplines is explained briefly.

2.1 Why Agriculture?

In the agricultural field, many economic, social, technical, environmental and cultural problems, especially land reforms, have been predisposed to create crises in almost every period, including the Roman period (Tahiroğlu, 1981). As a historical phenomenon, the agricultural sector, the agricultural society and the food crises experienced accordingly continue to affect today, as in every period of human history. Many times in history, there have been agricultural crises that swept the whole world. These crises, which can be experienced regionally in various parts of the World (Gray, 2020), can also expand to cover the whole World (Besthorn, 2013). Depending on the crisis and problems, different solutions have been developed many times and mankind has always succeeded in drawing new lessons from the experiences (Burton & Fischer, 2015; Lioutas & Charatsari, 2021; Smart et al., 2015).

It is stated that digital transformation such as, smart technology and big data, which is expressed as the most important crisis that humanity has experienced recently, Covid 19, bring important problems to be solved in order to help farmers overcome external shocks (Lioutas & Charatsari, 2021).

2.2 Why Agri-Tech?

Especially after the covid 19 pandemic, there are many opportunities in the field of agriculture, but it is seen that there are threats and difficulties. It is emphasized that all these difficulties can be overcome with obstacles, technological and digital transformation (Lowry et al., 2019).

Considering the issue of environmental sustainability, which is expressed as one of the most important global problems the world is in, the co-applicability of Agriculture and Technology has gained even more importance (Liaqat, 2017).

Environmental factors, which are expressed as the most important dimension of sustainability, are also at the root of the global climate crisis. For this reason, there are studies that evaluate many issues, especially the zero waste issue and the circular economy (Chudasama, 2019). In these studies, there are important clues and findings that the difficulties, challenges and obstacles expressed at the beginning of this part can be overcome with agri-tech (Chudasama, 2019).

2.3 Is There a Gap Between Agriculture and Technology?

The answer is certainly “yes”. There is a significant gap between agriculture and technology, but efforts are being made to bridge that gap and leverage technology to enhance agricultural practices. One of the main importance of the gap, may be expressed by the technological transformation is adopted firstly in different and highly profitable and strategic sectors such as defence system, service sectors like health and communications.

Traditionally, agriculture has been viewed as a more traditional and labor-intensive sector, relying on manual labor, basic tools, and traditional knowledge. However, advancements in technology have the potential to revolutionize the agricultural industry, increasing efficiency, productivity, and sustainability (Lowenberg-DeBoer & Swinton, 2018; USDA, n.d.).

One area where technology is making a significant impact is precision agriculture. Precision agriculture involves the use of technologies such as sensors, GPS, drones, and satellite imagery to collect data and monitor various aspects of crop production, including soil conditions, moisture levels, pest infestations, and crop health. This data-driven approach allows farmers to make informed decisions about irrigation, fertilization, pest control, and crop management, resulting in optimized resource allocation and improved yields (Lowenberg-DeBoer & Swinton, 2018).

Furthermore, the Internet of Things (IoT) is playing a crucial role in connecting agricultural devices and enabling real-time monitoring and control. IoT devices, such as soil moisture sensors and automated irrigation systems, can collect and transmit data to farmers, enabling them to remotely monitor and manage their crops. This technology helps farmers save resources by applying water and fertilizers precisely where and when needed (He et al., 2018).

Additionally, artificial intelligence (AI) and machine learning (ML) algorithms are being employed to analyze large amounts of agricultural data and provide valuable insights. These technologies can identify patterns, predict crop diseases, optimize planting strategies, and even automate certain agricultural processes (Cruz & Morais, 2020).

3 Literature Review

The gap between agriculture and technology is a recognized challenge in the agricultural sector, with limited adoption of technological advancements. This gap arises due to various factors, including the complexity of integrating new technologies into traditional farming practices and the lack of awareness and access to technology among farmers (Chen et al., 2017). The existance between agriculture and technology, has been widely explained in Chen's and friends study.

Bridging the gap in between two discipline is very important. To be able to bridge the gap between agriculture and technology is crucial for the advancement and sustainability of the agricultural sector is essential in many aspects. Technology can enhance productivity, resource management, and decision-making processes in farming, leading to improved yields and profitability (Lambert et al., 2020).

There are studies on also the role of digital agriculture. Digital agriculture, which encompasses various technological applications such as precision agriculture, IoT, and data analytics, plays a significant role in bridging the gap between agriculture and technology. It enables data-driven decision-making, optimization of inputs, and the adoption of sustainable farming practices. So it aggrigately increases the source productivity (Zhang et al., 2018).

Bridging such king of gap, needs some properties. The adoption of agri-tech faces challenges such as high initial costs, lack of technical skills among farmers, inadequate infrastructure, and limited access to information and support systems seems some of the challenges in Agri-Tech adoption. Addressing these challenges is essential to facilitate the uptake of technology in agriculture (Demiryürek et al., 2020).

There are also many studies on the gap between agriculter and technology also digitalization, including different aspect of view in literatüre such as: potential benefits of Agri-tech, role of different components of industrial revolution which are used in agriculture, human capital related to our subject and start up studies on the same area. The review is shown in Table 1.

Table 1. Summary of Literature Review on Agri-Tech and its Related Disciplines

No	Sources	Subject
1	(Chen et al., 2017)	Gap between Agriculture and Technology
2	(Lambert et al., 2020)	Importance of Bridging the Gap
3	(Zhang et al., 2018)	Role of Digital Agriculture
4	(Demiryürek et al., 2020)	Challenges in Agri-Tech Adoption
5	(Kavoori et al., 2021)	Potential Benefits of Agri-Tech
6	(Miah et al., 2020)	Blockchain in Agriculture
7	(Munoz-Carpena et al., 2020)	Otonomius Agriculture
8	(Qin et al., 2021)	Human Capital in Agri-Tech
9	(Singh et al., 2020)	Startups and Innovation in Agriculture
10	(Gomez-Sanchez et al., 2019)	Big Data Analytics in Agriculture

Source: (Miscellanous Study)

There are also different studies on disadvantages of using digital components in agriculture. Mainly, it may depends on the countires conditions and infrasturucture. However, some of the countires may not require directly technological and digital transformation where as they have to regulate and taking different precautions to develope agriculture policies (Öz et al., 2021).

4 Bibliometric Analysis

Bibliometric analysis is a research method that reveals the relationships of a subject under research by charting the authors, keywords, countries, funder groups, publication types. In this study, web of science is used to analize the gap emerging in between technology and agriculture. Other databases may also be applied to analize this subject but, fort his study it is accepted as a limitaions and seemed a study which is satisfied by web of science on the keywords of this study (Sott et al., 2021). Sott and his friends study has been obtained (as the only one) when the query is run on the web of science below: (Query #1)

TI=(Bibliometric) AND

(TI=(Agriculture) OR TI=(Farmer)) AND

(TI=(Digital) OR TI=(Technolog)) ***Query #1***

(Clarivate, 2023)

In this statement, as an SQL Query, TI refers to the Title of the publications. The data, obtained by running the query below (in Query #2), 1404 results are listed and exported the whole data from web of science as text file and than run at vosviewer software program (Vosviewer, 2023). The results are discussed in the following parts below. The main study is originated on the Query 2, as follows.

(TI=(Agriculture)) AND

(TI=(Technology) OR TI=(Digital)) **Query #2**

(Clarivate, 2023)

When TS (Subject) is used instead of TI (Title), the system gives 32,145 results (as in Query #3). This will distract the research from focusing a little more work.

(TS=(Agriculture)) AND

(TS=(Technology) OR TS=(Digital)) **Query #3**

(Clarivate, 2023)

Before running vosviewer, web of science gives the analysis of the result for Quer #2 as follows. Figure 1, gives the chart for 1,404 publications selected from Web of Science Core Collection and their study fields on

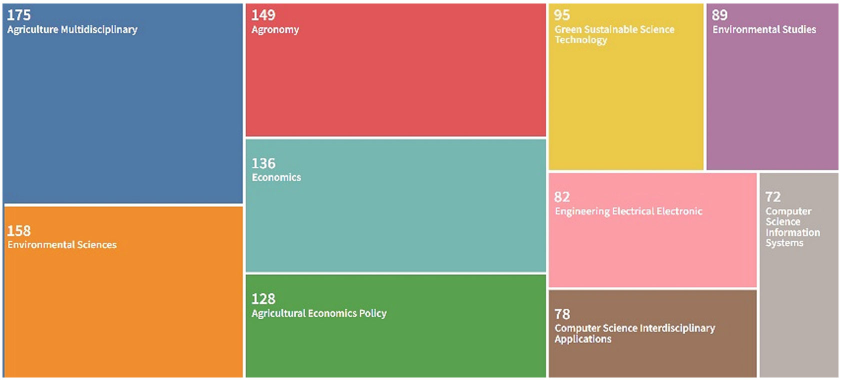


Fig. 1. The Fields of the Publication of the results from Query #2 (Clarivate, 2023).

In Fig. 1, it is shown that, %12 of the whole publications are related with agriculture multidisciplinary where as technology is in the 5th place via sustainability policies.

When the publcation years is asked to the web of science, Fig. 2 is obtained.

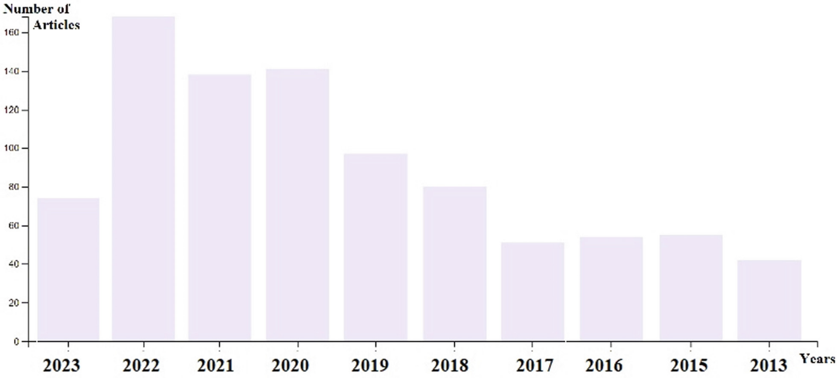


Fig. 2. Publication Years of the Query #2 (Clarivate, 2023)

Figure 2 depicts that there is an increase year by year on the Agri-Tech related studies. THis situation shows that, the imortance, interest and worth of publishing is in increase every year. Also a leap should be noted during pandemic years, after 2019 publications leaped as it is shown in Fig. 2, years 2020, 2021.

4.1 Keywords Analysis on Web of Science via Vosviewer

Once the results is saved as text file, as an input data file to vosviewer software programm, the output could be obtained in different styles. In this study, firstly, the keywords has been taken into account and set co-occurence with author keywords and finally, minumum number of occurences of a keyword is taken 10. This would yield 3050 keywords and 30 of them meet the treshhold. Figure 3 gives the overlay visualization form of the 1404 publications.

Figure 3 depicts that, the latest studies are on digital transformation, smart farming, AI, Agriculture 4.0 and climate-smart agriculture. However, the study volume is still less than the sustainable agriculture, sustainability and precision agriculture. The density visualization also support this situation as shown in Fig. 4.

Meanwhile, vosviewer classified the keywords under 5 clusters. The total link strenght is 448 of 193 links. The Clusters are:

Cluster 1: artificial intelligence, big data, blockchain, internet of things, internet of things, iot, machine learning, sensors, smart agriculture.

Cluster 2: agriculture, digital agriculture, digitalization, innovation, precision farming, technology adoption, technology transfer.

Cluster 3: adoption, climate change, climate-smart agriculture, food security, sustainable agriculture, technology.

Cluster 4: 0, agriculture 4, digital Technologies, digital transformation, sustainability.

Cluster 5: precision agriculture, remote sensing, smart farming.

As understood the clusters are based on technology and technology based disciplines. To study on the Agri-Tech is worthwhile and has to be increased.

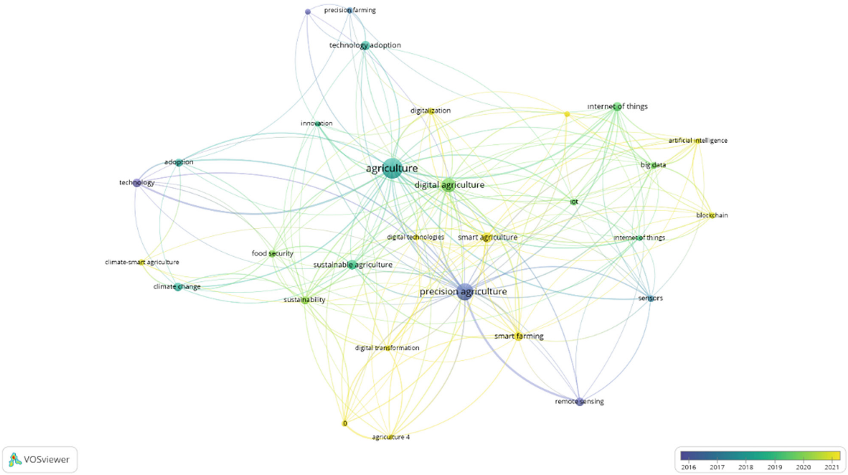


Fig. 3. Overlay Visualization of the Keywords of Query #2 (Vosviewer, 2023).

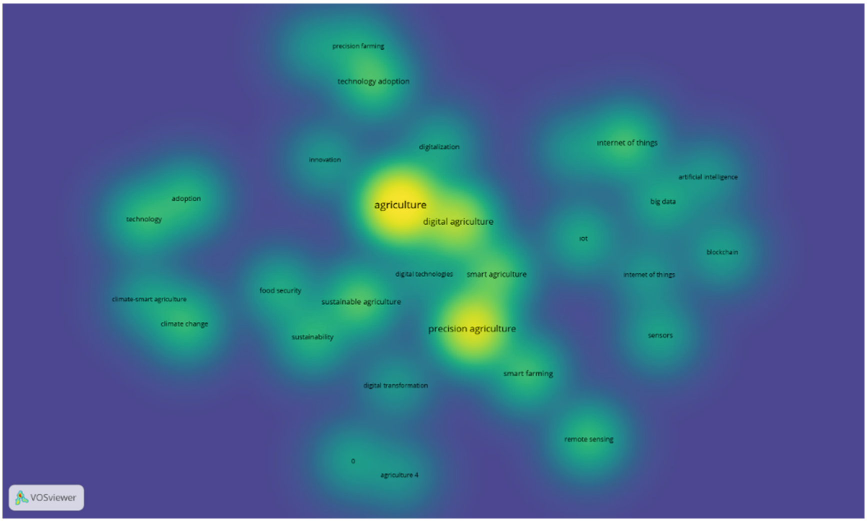


Fig. 4. Density Visualization of the Keywords of Query #2 (Vosviewer, 2023).

4.2 Authors Analysis on Web of Science

When the vosviewer program is loaded the data and runned with the settings as, co-authorship as the type of analysis and chosing the unit of analysis as Authors gives the relations and network charts in between the authors.

As shown in Fig. 5, the most cited and most network strength have been given authors by authors when the number of minumum publication is chosing as 3.

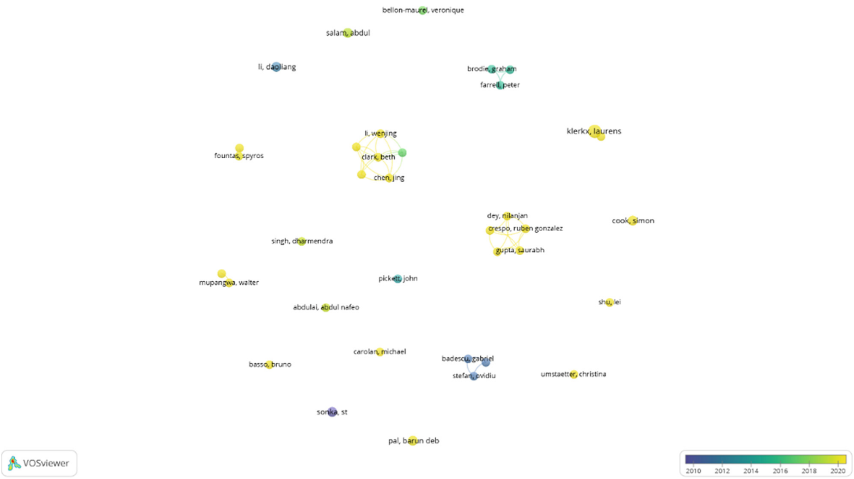


Fig. 5. Overlay Visualization of the Co-Author of Query #2 (Vosviewer, 2023).

Figure 5, shows that the recent studies have been held by Klerkx, Zhang, Jiao but the Klerkx has more relations wth others and more articles on the focused research of this study (Jiao et al., 2021; Klerkx et al., 2019; Klerkx & Rose, 2020; Zhang et al., 2018).

When the co-authors is analized by the unit of countries in the vosviewer, Fig. 6 is obtained.

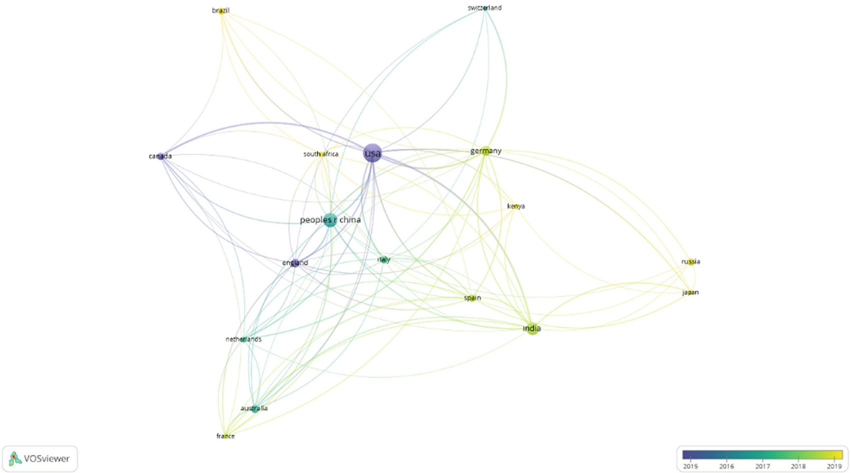


Fig. 6. Overlay Visualization of the Co-Author with th Country Unit Analysis of Query #2 (Vosviewer, 2023).

Figure 6 is obtaned when the minumum number of documents of a country is taken 20, and minumum number of citation of that country as 20. By this settings, out of

119 countries just 18 is with in the range and meet the tresholds. Since Romania is not connected to those of 17, Fig. 6 shows the rest of 17 and their relations.

In Table 2, the citations and the number of documents and strenghts are given for 18 countries.

Table 2. Documents Published on the Countries related to Query#2.

Rank	Country	Documents	Citations	Total Link Strength
1	USA	285	4594	93
2	Germany	74	1686	50
3	India	126	1199	45
4	China	160	2182	41
5	England	61	1672	40
6	Spain	44	1561	38
7	Canada	43	1046	26
8	France	22	1005	26
9	Australia	51	1154	25
10	Netherlands	25	1169	25
11	Italy	45	854	25
12	South Africa	23	415	19
13	Japan	24	225	15
14	Kenya	22	447	14
15	Switzerland	20	270	12
16	Brasil	38	329	10
17	Russia	43	110	6
18	Romania	21	55	0

(Vosviewer, 2023)

The Table 2, shown above also depicts that the USA is taken the most interest on Agriculter and Technology related publications. At least literatüre says that the concern is mostly produced by USA, Germany and India.

Vosviewer makes a cluster list and out of that 17 countries results with 5 clusters.

Cluster 1: australia, england, france, peoples r china

Cluster 2: india, japan, russia, spain

Cluster 3: germany, italy, netherlands, switzerland

Cluster 4: brazil, canada, usa

Cluster 5: kenya, south africa

It is seen that geograficalla and the continent is also a like to that clusters. But, spain china and australia break that rule.

5 Conclusions

As a result of the bibliometric data analysis, it is clear that bridging the technological and agricultural divide is crucial in the field of agri-tech, with digital transformation—primarily Artificial Intelligence (AI) and blockchain technologies—emerging as the most significant fields. The rising amount of academic research and publications shows how digital transformations are becoming more widely recognized for their ability to transform agricultural practices and address industry concerns.

The results of the bibliometric analysis show a rising trend in research papers on smart agriculture methods in agri-tech, reflecting the growing interest in utilizing the capabilities of digital transformations to boost productivity, improve resource management, and encourage sustainable farming. The application of artificial intelligence (AI) algorithms, machine learning methodologies, and big data analytics in agriculture, in particular, has the potential to replace conventional farming practices with more effective, data-driven, and precision-based methods.

Farmers may use real-time data and predictive models to inform their decisions by utilizing the digital transformation in agri-tech. For instance, AI-powered technologies allow for precision agriculture, enabling tailored interventions like improved crop protection and irrigation. In addition, new algorithms can evaluate enormous volumes of agricultural data, offering insightful information on market trends, disease detection, yield forecast, and crop health. Farmers are given the tools they need to maximize their production plans, reduce risks, and enhance all aspects of farm management.

While the bibliometric analysis highlights the significance of digital transformation, it is crucial to understand that closing the technological and agricultural divide in agri-tech requires a multidisciplinary approach. Developing comprehensive solutions that address the unique demands and difficulties of the agriculture sector requires collaborative research and innovation across a range of disciplines, including computer science, agronomy, engineering, and environmental science.

It is imperative to promote information transfer and collaboration between researchers, practitioners, policymakers, and farmers in order to fully close the gap between technology and farming. The adoption of digitalization in agriculture should be supported by supportive legislation, training and education programs on technical implementation, and efforts to ensure accessibility to all aspects of the digital transition.

The bibliometric analysis concludes by highlighting the growing importance of digital transformation in bridging the divide between farming and technology in agri-tech. The agriculture industry can unleash the potential of cutting-edge technology, resulting in more sustainable, effective, and productive farming techniques, by embracing AI, Blockchain, smart-farming components, and encouraging interdisciplinary collaborations.

References

Besthorn, F.H.: Vertical farming: social work and sustainable urban agriculture in an age of global food crises. Aust. Soc. Work. **66**(2), 187–203 (2013). https://doi.org/10.1080/0312407X.2012.716448

Burton, R.J.F., Fischer, H.: The succession crisis in European agriculture. Sociol. Rural. **55**(2), 155–166 (2015). https://doi.org/10.1111/soru.12080

Chen, C., Huang, Y., Chien, S.: Factors influencing farmers' adoption of agricultural technology: a discriminant analysis. Sustainability **9**(8), 1436 (2017)

Chudasama, A.: Review of the world agri-tech innovation summit held in London in October 2018. Int. Sugar J. **121**(1441), 16–18 (2019)

Clarivate. Web of Science (2023). https://www.webofscience.com/wos

Cruz, A., Morais, R.: Artificial intelligence in agriculture: a systematic literature review. Comput. Electron. Agric. **175**, 105507 (2020)

Demiryürek, K., Kanber, R., Özkan, B.: Review on adoption of agricultural technologies and trends in digital agriculture. Comput. Electron. Agric. **179**, 105827 (2020)

Gomez-Sanchez, E., Rovira-Mas, F., Garsia-Gasulla, D.: Big data analytics in agriculture: a systematic literature review. Comput. Electron. Agric. **156**, 417–427 (2019)

Gray, R.S.: Agriculture, transportation, and the COVID-19 crisis. Can. J. Agric. Econ.-Rev. Can. Agroecon. **68**(2), 239–243 (2020). https://doi.org/10.1111/cjag.12235

He, L., Wang, J., Hu, S., Huang, W.: Application of IoT technology in precision agriculture based on wireless sensor network. In: 2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA), pp. 83–88 (2018)

Jiao, X., Zhang, H., Ma, W., Wang, C., Li, X., Zhang, F.: Science and technology backyard: a novel approach to empower smallholder farmers for sustainable intensification of agriculture in China (vol 18, pg 1657, 2019). J. Integr. Agric. **20**(6), IX–IX (2021)

Kavoori, P., Dodia, J., Sheth, N.: Agriculture 4.0: a comprehensive review on digital technologies in precision agriculture. Comput. Electron. Agric. **188**, 106344 (2021)

Klerkx, L., Jakku, E., Labarthe, P.: A review of social science on digital agriculture, smart farming and agriculture 4.0: new contributions and a future research agenda. NJAS-Wageningen J. Life Sci. **90–91**, 100315 (2019). https://doi.org/10.1016/j.njas.2019.100315

Klerkx, L., Rose, D.: Dealing with the game-changing technologies of agriculture 4.0: how do we manage diversity and responsibility in food system transition pathways? Glob. Food Secur.-Agric. Policy Econ. Environ. **24**, 100347 (2020). https://doi.org/10.1016/j.gfs.2019.100347

Lambert, D.M., Hogan, R.J., De Faveri, S.: Realigning agriculture and technology: theory and practice for sustainable futures. Futures **117**, 102512 (2020)

Liaqat, A.: Forty years of innovated industrial based agri-tech in sustainability for zero food waste. Ann. Nutr. Metab. **71**, 1323 (2017)

Lioutas, E.D., Charatsari, C.: Enhancing the ability of agriculture to cope with major crises or disasters: what the experience of COVID-19 teaches us. Agric. Syst. **187**, 103023 (2021). https://doi.org/10.1016/j.agsy.2020.103023

Lowenberg-DeBoer, J., Swinton, S.: Precision agriculture for sustainable intensification. Glob. Food Sec. **16**, 9–15 (2018)

Lowry, G.V., Avellan, A., Gilbertson, L.M.: Opportunities and challenges for nanotechnology in the agri-tech revolution. Nat. Nanotechnol. **14**(6), 517–522 (2019). https://doi.org/10.1038/s41565-019-0461-7

Miah, S.J., Kabir, E., Ashour, M.R., Gammack, J.: Blockchain technology in agriculture: enhancing trust and transparency in the food supply chain. Foods **9**(7), 941 (2020)

Munoz-Carpena, R., Chu-Agor, M.L., Leidi, E.O., Slaughter, D.C., Whidden, S.E., Rosskopf, E.N.: Current status and challenges of agricultural robotics. Annu. Rev. Biomed. Eng. **22**, 303–330 (2020)

Öz, S., Yüksel, İ., Aşçı, M.S.: Teknolojik ve Geleneksel Tarim: Türkiye PESTLE Analizi. In Dijital Gelecekte Mesleklerin ve Sektörlerin Dönüşümü, 1st edn., pp. 601–622. Hiperyayin (2021)

Qin, Y., Zhan, B., Chen, Y.: Training programs for precision agriculture: a systematic review. Comput. Electron. Agric. **185**, 106201 (2021)

Singh, R., Basavarajappa, S., Parakash, C., Agrawal, R.: Startup innovation in agriculture and digital technologies: a review. J. Clean. Prod. **262**, 121310 (2020)

Smart, J., Nel, E., Binns, T.: Economic crisis and food security in Africa: exploring the significance of urban agriculture in Zambia's Copperbelt province. Geoforum **65**, 37–45 (2015). https://doi.org/10.1016/j.geoforum.2015.07.009

Sott, M.K., et al.: A bibliometric network analysis of recent publications on digital agriculture to depict strategic themes and evolution structure. Sensors **21**(23), 7889 (2021). https://doi.org/10.3390/s21237889

Tahiroğlu, B.: Roma Devletinin İktisadi Krizleri. J. Istanbul Univ. Law Fac. **45**(1–4), 677–706 (1981)

USDA. Precision Agriculture. United States Department of Agriculture (n.d.). https://www.usda.gov/topics/farming/precision-agriculture

Vosviewer. Centre for Science and Technology Studies, Leiden University (1.6.18) [English]. Centre for Science and Technology Studies, Leiden University (2023). https://vosviewer.com

Zhang, X., Wei, X., Jiang, W.: Digital agriculture in China: state-of-the-art and future perspectives. J. Integr. Agric. **17**(12), 2679–2693 (2018)

Eye Tracking Review: Importance, Tools, and Applications

Taisir Alhilo and Akeel Al-Sakaa(✉)

College of Computer Science and Information Technology, University of Kerbala, Karbala, Iraq
taisir.h@s.uokerbala.edu.iq, akeel.a@uokerbala.edu.iq

Abstract. Eye tracking technology has evolved as a powerful and flexible tool, providing critical insights into human visual activity and cognitive processes in a range of disciplines. This article looks at the importance of eye tracking and the hardware components used in eye tracking systems. The benefits and applications of types of eye trackers, such as remote and head-mounted, are discussed. Considerations like as spatial resolution, sample rate, and accuracy assist researchers in selecting the best equipment for their study objectives.

The study discusses software essential for evaluating and interpreting data from eye movements. Then deepen on major eye movements and metrics utilized in eye tracking studies, such as fixation length, saccades, and pupil dilation. These measures offer useful insights on visual attention and cognitive processes, allowing researchers to better understand how people respond to visual stimuli.

Finally, eye tracking applications in psychology, marketing, human-computer interface, and medical research are highlighted. Eye tracking demonstrates its adaptability and importance in understanding human behavior in real-world scenarios, from analyzing consumer behavior to improving user interfaces.

Keywords: Eye Tracking · Saccades · Eye Tracking Applications · Eye movements · Fixation

1 Introduction

Eye tracking is defined as “an experimental method of recording eye motion and gaze location across time and task” [1] that gives important information about how the human brain works [2].

The process of monitoring either the point of gaze or the motion of the eye toward a specific point is known as eye tracking [1]. This technology has garnered significant attention as a valuable tool for understanding human visual attention across various domains, including perception, attention, memory, reading, psychopathology, ophthalmology, neuroscience, human-computer interaction, animal research, human factors, consumer behavior, and optometry [3, 5, 6].

The roots of eye tracking can be traced back to the late 1800s when Louis Émile Javal manually observed people’s eye movements during reading, using a metallic contact lens connected to recording pens through mechanical linkages [9]. However, contemporary

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 383–394, 2024.
https://doi.org/10.1007/978-3-031-56728-5_32

eye tracking has evolved to be much more user-friendly, although its implementation varies significantly across different types of technology.

Today, eye tracking primarily relies on corneal reflection, where the eye tracker emits a small infrared light into the eye [26]. The camera captures the reflection of the light source on the cornea and pupil, enabling the calculation of a vector based on the angle between the cornea and pupil reflections. This vector's direction can then be utilized to determine the fixation direction [25].

This research works to collect the basic concepts from several previous researches, Sect. 2 delves into an exhaustive review of related works. Section 3 unveils the arsenal of eye tracking tools and methodologies employed. Section 4 deciphers the intricate interplay between eye movements and cognitive processing, unveiling the temporal dimensions that underscore their convergence. Finally, Sect. 5 presents applications of eye tracking that benefit from our holistic approach, thereby encapsulating the essence of our inquiry.

2 Related Works

Eye-tracking technology has been used extensively in various fields, some of which I chose to demonstrate the importance of eye movements and visual attention patterns in providing valuable insights into decision-making processes.

(Madlenak et al., 2023) [34], Investigating outdoor advertising's impact, this research combines eye-tracking analysis and A/B testing during car journeys. Insights into conscious and subconscious customer behavior emerge, impacting advertising strategies.

(Kuo et al., 2021) [39], Eye tracking's role in design concept validation is explored through two cases, revealing predictive product quality perception through eye movement analysis.

(Clay et al., 2019) [4], This paper introduces eye tracking in Virtual Reality (VR), blending immersive environments with eye tracking techniques. Implementation guidance and case studies demonstrate its potential for advanced human behavior research.

(Keller et al., 2016) [43], Utilizing eye motions, this study conducts neuropsychological assessments for physically challenged individuals. ALS patients' effective eye-tracking criteria are addressed, proposing a reliable method for assessing cognitive impairments.

3 Eye Tracking Tools

Eye tracking solutions include both the gear and software required to capture and analyze eye movement data. A device that measures and records eye movements and gaze patterns is known as an eye tracker. An infrared light source, cameras, and sensors that detect and monitor eye movements are all part of the hardware. The program analyzes the collected data and generates visuals and analytics.

The integration of machine learning (ML) algorithms with eye tracking devices is now possible, enabling the incorporation of learning functions from gathered data, resulting in smarter eye tracking systems. Building on these technical progressions, various research groups and organizations have introduced a diverse array of hardware and software techniques [31].

3.1 Eye Tracking Hardware Components

Eye tracking hardware is made up of fundamental components that allow eye movements to be measured. These components are as follows:

a. Infrared Illuminators: Infrared illuminators emit infrared light to illuminate the eyes, allowing cameras to monitor eye movements precisely even in low-light cir-cumstances.
b. Cameras: High-resolution cameras record the position and movement of the eyes to acquire eye movement data. Depending on the eye tracking method, the cameras might be monocular (one eye) or binocular (both eyes).
c. Sensors: Eye tracking systems measure eye properties such as the pupil and corneal reflection using a variety of sensors. These sensors help detect eye movements and calculate gaze coordinates more precisely.
d. Mounting Hardware: To fasten the device to the participant's head or computer monitor, eye trackers may require particular mounting hardware or fixtures.

Eye Tracking Devices Types:

Eye tracking devices can be categorized into various types based on their form and usage:

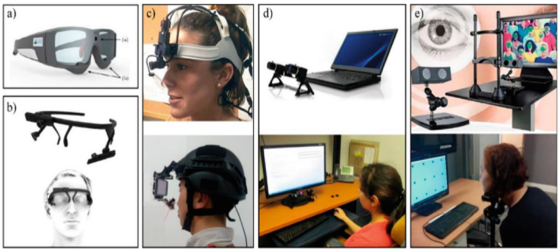


Fig. 1. Example of the different types of eye tracking devices: (a) eye tracking glasses; (b) headband; (c) helmetmounted; (d) remote or table; (e) tower-mounted [38].

3.1.1 Mobile Eye Tracking Devices (Head-Mounted)

Wearable or head-mounted gadgets, often equipped with a second scene camera capturing the field of view, are referred to as mobile eye tracking devices [27]. (see Fig. 1) illustrates the various forms of these devices, including headbands, glasses, and helmet-mounted systems, which allow participants to move freely during experiments. Mobile eye tracking devices are favored for their accuracy, primarily due to their binocular nature [27].

Gaze tracking on mobile devices is conducted relative to the complete field of vision, making them well-suited for real-world investigations [29]. Their less intrusive and comfortable nature enables them to be used alongside other technologies, such as electroencephalography (EEG).

However, mobile tracking devices do have notable limitations. They may struggle to capture eye movements accurately in direct sunlight and are unsuitable for environments with high winds and water spray [14]. Additionally, observing eye movements towards the periphery can be challenging and often yields less precise results. The absence of an absolute coordinate system in mobile tracking systems necessitates storing gaze data in a coordinate system specified by the scene camera. Finally, data inaccuracies may arise if the mobile eye tracking device does not fit the individual's face adequately [28].

3.1.2 Remote Eye Tracking Devices

Remote eye tracking equipment is commonly used in studies involving screen-based interactions, offering the advantage of allowing users to use a computer normally while the eye tracking device collects data [29]. This non-intrusive approach ensures that the subject is not touched during data collection, enabling safe and distant eye measurements [25]. Additionally, remote eye tracking devices benefit from the integration of the visual world, making data processing less complicated and more efficient than with wearable systems [28]. Moreover, their non-contact nature makes them compatible with other research tools.

However, remote eye tracking devices have some limitations. They are primarily suitable for use in fixed working areas, which can result in data gaps and artifacts if the participant moves their head excessively. Furthermore, they may be sensitive to infrared (IR) sources, such as sunlight, especially if the sun reflects in the participant's eyes.

To enhance accuracy, head-supporting towers (see Fig. 1e) are often employed, providing direct contact with the participant through a bite bar or chin rest, thereby restricting head movement [29]. Although less natural, head-supporting towers enable the acquisition of high-quality data by constraining head movement, resulting in a saccade resolution two to five times greater than that of a remote/head-free eye tracker [10].

Nevertheless, the application of head-supporting towers is limited in dynamic contexts due to their re-stricted positioning. Therefore, they are commonly used in investigations requiring high accuracy and when the participant typically gazes at a stationary screen [30].

3.2 Eye Tracking Software

Although the features differ from one company to the next, the following are the key features and capabilities of eye tracking software:

a. Data Collection: Software makes it possible to capture and synchronize eye movement data with other factors such as stimulus presentation or task performance.
b. Calibration and Validation: Software leads the calibration process to ensure correct gaze-to-screen mapping and provides validation tools to evaluate the quality and

reliability of eye tracking readings. Calibration is a critical stage in eye tracking research. It entails having participants focus on certain locations or targets presented on the screen in order to develop a mapping between the participant's gaze and the coordinates displayed on the screen. Calibration provides precise tracking of eye movements and enhances data dependability.

c. Data analytic: Software provides a variety of analytic tools for extracting eye movement measurements, creating visualizations (such as heatmaps or gaze plots), and doing statistical analyses on the acquired data.

Dedicated software may be required for: depending on the developer

- Different forms of eye trackers (for example, specta-cles versus screen-based)
- A unique eye tracker
- Distinct stimuli (for example, static vs. dynamic)

4 Eye Movements and Metrics

Eye movements, including fixations, saccades, vergence, and pupil dilatation or constriction, represent various ocular motions [1], (see Fig. 2). Just and Carpenter's eye-mind and immediacy assumptions (1980) posit that what is focused on is also being processed, and there is no noticeable lag between the two, linking eye movements to cognitive operations. While these assumptions serve as the theoretical basis of eye-tracking, they do have limitations, such as not explaining mind wandering independently of eye movements [32].

4.1 Fixation

Represents eye movements that stabilize the retina over a stationary object of interest and is closely linked to visual processing and information acquisition [26]. During a fixation, the eye remains motionless, and the pupil remains stationary for a duration of 180–300 ms [15, 19].

This period allows individuals to gain new information from the item, stimulation, or location being fixated upon [13].

Various fixation measures, such as fixation frequency, fixation duration, fixation duration max, and fixation duration standard deviation, are connected to human performance and several cognitive attentional processes [13]. For instance, the distribution of fixations and fixation length on important areas of interest (AOIs) can infer a pilot's situation awareness (SA) performance and skill level [12].

In the construction industry, a study revealed that workers with a higher risk perception spent a longer time fixating on potentially threatening objects when they were first spotted [11]. Similarly, higher fixation counts on relevant AOIs were associated with a higher frequency of failure detection [18]. Additionally, individuals with shorter fixation times are more likely to experience anxiety and be in a state of danger [21].

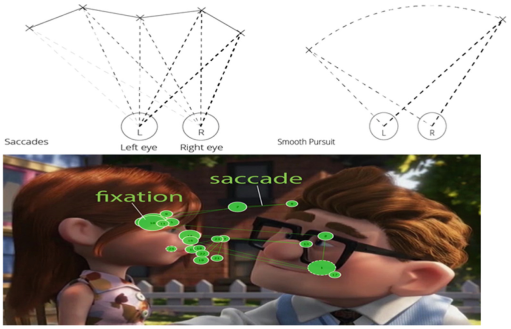


Fig. 2. Eye Movements (Fixation, Saccades, Smooth Pursuits) [52]

These findings demonstrate the significance of fixation measures in understanding human behavior, attention, and cognitive processes in various contexts.

Fixations serve as junctions for decision-making processes connected with influential cognitive processes [40, 53]. Shows how eye fixations might represent mental processes and so be beneficial in generating numerous suggestions.

4.2 Smooth Pursuits

Pursuit motions are utilized when visually tracking a moving target, allowing the eyes to match the velocity of the target based on its range of motion. These motions represent a control system with negative feedback, enabling smooth and accurate tracking of moving objects. In gaze interfaces, gaze trajectories resulting from smooth-pursuit eye movements are employed for object recognition [25].

These smooth-pursuit eye movements have inspired the design of gaze interfaces, where users follow a moving object with their eyes to select it. The resulting gaze trajectory is then compared to the trajectories of moving objects on the display. One notable advantage is that no calibration phase is required since trajectory detection is invariant with respect to its origin position [54]. In the initial implementation of a smooth-pursuit based interface, Pearson's product-moment correlation was used to link gaze trajectories to moving objects. It was observed that the detection rate decreased when objects on the interface moved along linear trajectories rather than circular trajectories. To enhance detection performance, interfaces based on linear pursuit eye movements and other techniques were developed [55].

Researchers have employed smooth pursuit eye movements for various applications, such as PIN code entry [55], word prediction in gaze typing systems [56], and object activation in gaze interfaces [8]. For instance, a gaze typing system named SMOOVS was based on two-segment pursuit eye movement, and its typing efficiency was further improved with word prediction capabilities [55]. Schenk et al. [8] developed a system that integrated multiple eye movements, including fixation for object selection and linear smooth pursuit motions for object activation.

4.3 Saccades

Quick eye movements, known as saccades, are caused by a person shifting their gaze between fixations [16]. Saccades typically last around 10–100 ms, during which visual information transmission is inhibited [23], indicating that saccades are not directly associated with cognitive processing [15].

However, saccade velocity has been linked to lethargy, tension, and weariness [20, 22]. For instance, saccade rate decreases with fatigue and demanding tasks [41]. Saccadic length has also been utilized to assess mental workload [15], and it has been observed to increase with higher mental workload (MWL), with very short saccades indicating the presence of conflict [17]. These relationships between saccadic characteristics and psychological states or cognitive demands provide valuable insights into human behavior and cognition. Understanding saccades and their correlations with various factors can aid in evaluating mental workload, attention, and overall performance in decision-making and cognitive tasks.

4.4 Pupil Size

Is one of the most defining aspects of the human eye [24], and retrieving information about its size and position via video capture is rather simple [15]. Illumination influences pupil size, which controls the quantity of light that penetrates the retina. Emotions, muscle tiredness, cognitive processes, and MWL all have an effect on pupil size [12].

Eye movement measurements are a technology that allows for an accurate depiction and comprehension of eye movements. The corneal reflection approach is widely used in modern eye-trackers to monitor the eye's location and movement. This approach involves directing infrared light sources into the eye and capturing the reflection using a high-resolution camera [33]. By analyzing the camera image, the source of light reflection on the cornea can be identified, enabling the determination of the subject's gaze position.

5 Applications of Eye Tracking

Eye tracking is a versatile and powerful technique applicable to a wide range of human behavior research. Its applications span across diverse sectors, such as healthcare and medical research, psychology, marketing, engineering, education, and gaming. Additionally, eye tracking technology plays a pivotal role in enhancing human-computer interfaces, enabling navigation and control through eye movements.

5.1 Eye-Tracking and Robot

Because it is simpler to extract, most human-robot interaction investigations now utilize head posture as a surrogate for true eye gazing [35]. However, "head gaze" does not convey all of the information that "eye gaze" provides, therefore allowing robots to conduct eye tracking might considerably increase their skills as well as human acceptability. Furthermore, the majority of human-robot interaction investigations that focus on gaze include external eye tracking devices [36].

5.2 Healthcare and Medical Applications

Eye tracking provides a substantial amount of structured data within a specific time frame, assisting various medical theories in addressing challenges related to understanding cognitive processes. Additionally, it enables precise diagnosis and monitoring of patients' well-being.

For instance, patients with amyotrophic lateral sclerosis (ALS) often face challenges in speaking or writing. Keller et al. [43] developed a software application that utilizes neuropsychological screening measures based on the Edinburgh Cognitive and Behavioral ALS Screen. It tracks eye movements using an infrared sensor called EyeLive and a user-friendly interface. Instead of relying on mutual cameras, this system utilizes infrared sensors embedded in glasses to determine the direction of the eye gaze. This innovation is particularly beneficial for individuals with specific needs.

Moreover, eye tracking has been used to study the behavior of children with autism spectrum disorder (ASD) while interacting with audiovisual technologies, such as video games [37]. The Tobii eye tracking device was used to observe their responses, revealing that children with ASD were drawn to video games in a manner similar to other children.

5.3 Consumer Psychology, Marketing and Advertising

Many prominent businesses actively use eye tracking to analyze product performance, product and packaging design, and overall customer experience, as well as to monitor customer attentiveness to critical messaging and advertising. When used in-store, eye tracking gives data on the ease and complexity of in-store navigation, search behavior, and purchase decisions.

New marketing and advertising tactics necessitate consumer participation in the creation and assessment of products and services. This provides information on customer happiness, engagement, and design decisions.

Rasch et al. [44] conducted a study where they combined eye tracking with face electromyography to enhance prediction performance. This approach provided valuable insights into the drivers and circumstances of emotional decision processes.

Yegoryan et al. [45] linked the visual attention created by eye tracking to the probability of attribute non-attendance (ANA) and preference heterogeneity to test, explain, and justify ANA in a marketing environment.

Meißner et al. [46] explored the combination of virtual reality (VR) settings with eye tracking to provide a unique opportunity for shopper research, particularly on the usage of augmented reality for shopper assistance. Their work demonstrated how mobile eye tracking in VR can aid studies in retailing and decision-making. The integration of eye tracking in VR allowed the environment to interact with the user in real-time, displaying additional product information in response to the user's natural eye movements [42, 46]. This approach has the potential to offer valuable insights into consumer behavior and decision-making in a dy-namic retail environment.

5.4 Education and E-learning

Extensive studies have demonstrated the effectiveness of utilizing eye tracking data in educational applications, enabling the detection of changes in cognitive development and the evaluation of student focus, thereby facilitating improvements in the educational process [47].

Eye tracking has instrumental in enhancing the understanding of learning processes and outcomes across various learning contexts. For instance, Prieto et al. [48] used eye-tracking data to describe the cognitive load experienced by teachers in different class scenarios, without interfering with their learning experiences. This research sheds light on learners' visual, cog-nitive, and attentional performance during ex-periments.

Furthermore, eye tracking technology has been em-ployed by researchers like Rappa et al. [49] to improve learning and assess learning success in virtual reality (VR) settings. Inoue and Paracha [50] investigated how fluent and non-fluent readers process words and visuals to enhance reading outcomes, while Rasmussen and Tan [51] used eye gaze tracking in combination with speech recognition to improve language models for reading progress.

Najar et al. [7] utilized eye tracking data to compare the actions of novice and advanced students while studying examples in an intelligent tutoring system. They discovered that advanced students paid greater attention to specific elements, highlighting the relevance of eye tracking in assessing learning levels.

6 Conclusion

In this paper we studied eye tracking tools and It has a wide scope of applications. Eye tracking is growing quickly creating a vast ocean of information. We can analyzing huge data generated from an eye tracker and applied for solving a specific problem.

Eye movements are driven both by properties of the visual world and processes in a person's mind. Uniquely poised between perception and cognition, eye movements are an invaluable tool for researchers. The eye movements of a subject can provide researchers with a rich, dynamic data source concerning the temporal dynamics and psychological processes that led up to the response. These roperties are also of great value to designers and engineers, as they allow for detailed measurements of how a user is interacting with a device. Since the technology has become highly efficient, such information can now be fed back into devices in real time, and the movements of a user's eyes can be used to issue commands or tailor computational processes. This provide the most fluid and expressive interface between humans and computers.

Future work in the field of eye tracking holds immense potential, particularly when integrating eye-tracking technology with brain signals, such as electroencephalography (EEG). This convergence of modalities opens up a new realm of possibilities, offering a deeper understanding of the intricate relationship between visual attention and neural activity. By combining these two sources of information, researchers and practitioners can expect several exciting results.

References

1. Carter, B.T., Luke, S.G.: Best practices in eye tracking research. Int. J. Psychophysiol. **155**, 49–62 (2020)
2. Leigh, R.J., Zee, D.S.: The neurology of eye movements. Contemp. Neurol. (2015)

3. Majaranta, P. (ed.): Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies: Advances in Assistive Technologies. IGI Global (2011)
4. Clay, V., König, P., Koenig, S.: Eye tracking in virtual reality. J. Eye Move. Res. **12**(1) (2019)
5. Liversedge, S., Gilchrist, I., Everling, S. (eds.): The Oxford Handbook of Eye Movements. OUP, Oxford (2011)
6. Kowler, E.: Eye movements: the past 25 years. Vision. Res. **51**(13), 1457–1483 (2011)
7. Najar, A.S., Mitrovic, A., Neshatian, K.: Utilizing eye tracking to improve learning from examples. In: Stephanidis, C., Antona, M. (eds.) UAHCI 2014, Part 2. LNCS, vol. 8514, pp. 410–418. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07440-5_38
8. Schenk, S., Dreiser, M., Rigoll, G., et al.: GazeEverywhere: enabling gaze-only user interaction on an unmodified desktop pc in everyday scenarios. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI 2017, pp. 3034–3044. ACM, New York (2017)
9. Płużyczka, M.: The first hundred years: a history of eye tracking as a research method. Appl. Linguist. Pap. (25/4), 101–116 (2018)
10. Pinheiro, R., Pradhananga, N., Jianu, R., Orabi, W.: Eye-tracking technology for construction safety: a feasibility study. In: ISARC 2016–33rd International Symposium on Automation and Robotics in Construction (2016)
11. Habibnezhad, M., Fardhosseini, S., Vahed, A.M., Esmaeili, B., Dodd, M.D.: The relationship between construction workers' risk perception and eye movement in hazard identification. In: Construction research congress 2016, pp. 2984–2994 (2016)
12. Li, W.C., Zhang, J., Le Minh, T., Cao, J., Wang, L.: Visual scan patterns reflect to human-computer interactions on processing different types of messages in the flight deck. Int. J. Ind. Ergon. **72**, 54–60 (2019)
13. Bjørneseth, F.B., Renganayagalu, S.K., Dunlop, M.D., Hornecker, E., Komandur, S. Towards an experimental design framework for evaluation of dynamic workload and situational awareness in safety critical maritime settings. In: The 26th BCS Conference on Human Computer Interaction, vol. 26, pp. 309–314 (2014)
14. Forsman, F., Sjörs-Dahlman, A., Dahlman, J., Falkmer, T., Lee, H.C.: Eye tracking during high speed naviation at sea: field trial in search of navigational gaze behaviour. J. Transp. Technol. **2**, 277–283 (2012)
15. Muczyński, B., Gucma, M., Bilewski, M., Zalewski, P.: Using eye tracking data for evaluation and improvement of training process on ship's navigational bridge simulator. Zeszyty Naukowe/Akademia Morska Szczecinie **33**(105), 75–78 (2013)
16. Hareide, O.S., Ostnes, R.: Maritime usability study by analysing eye tracking data. J. Navig. **70**(5), 927–943 (2017)
17. Martin, C., Cegarra, J., Averty, P.: Analysis of mental workload during en-route air traffic control task execution based on eye-tracking technique. In: Harris, D. (ed.) EPCE 2011. LNCS (LNAI), vol. 6781, pp. 592–597. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21741-8_63
18. Hasse, C., Grasshoff, D., Bruder, C.: Eye-tracking parameters as a predictor of human performance in the detection of automation failures. In: Proceedings HFES Europe Chapter Conference Toulouse, pp. 133–144 (2012)
19. van Meeuwen, L.W., Jarodzka, H., Brand-Gruwel, S., Kirschner, P.A., de Bock, J.J., van Merriënboer, J.J.: Identification of effective visual problem solving strategies in a complex visual domain. Learn. Instr. **32**, 10–21 (2014)
20. Stankovic, A., Aitken, M.R., Clark, L.: An eye-tracking study of information sampling and decision-making under stress: Implications for alarms in aviation emergencies. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 58, no. 1, pp. 125–129. SAGE Publications, Los Angeles (2014)

21. Vine, S.J., Uiga, L., Lavric, A., Moore, L.J., Tsaneva-Atanasova, K., Wilson, M.R.: Individual reactions to stress predict performance during a critical aviation incident. Anxiety Stress Coping **28**(4), 467–477 (2015)
22. Jeelani, I., Albert, A., Han, K., Azevedo, R.: Are visual search patterns predictive of hazard recognition performance? Empirical investigation using eye-tracking technology. J. Constr. Eng. Manag. **145**(1), 04018115 (2019)
23. Häggström, C., Englund, M., Lindroos, O.: Examining the gaze behaviors of harvester operators: an eye-tracking study. Int. J. For. Eng. **26**(2), 96–113 (2015)
24. Snowden, R., Thompson, P., Troscianko, T.: Basic Vision: An Introduction to Visual Perception. Oxford University Press, Oxford (2012)
25. Duchowski, T.A.: Eye Tracking: Methodology Theory and Practice. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-57883-5
26. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J.: Eye Tracking: A Comprehensive Guide to Methods and Measures. OUP, Oxford (2011)
27. Cognolato, M., Atzori, M., Müller, H.: Head-mounted eye gaze tracking devices: An overview of modern devices and recent advances. J. Rehabil. Assist. Technol. Eng. **5**, 2055668318773991 (2018)
28. Kovesdi, C., Spielman, Z., LeBlanc, K., Rice, B.: Application of eye tracking for measurement and evaluation in human factors studies in control room modernization. Nucl. Technol. **202**(2–3), 220–229 (2018)
29. Andrychowicz-Trojanowska, A.: Basic terminology of eye-tracking research. Appl. Linguist. Pap. (25/2), 123–132 (2018)
30. Wang, D., Mulvey, F.B., Pelz, J.B., Holmqvist, K.: A study of artificial eyes for the measurement of precision in eye-trackers. Behav. Res. Methods **49**, 947–959 (2017)
31. Sarkar, A.R., Sanyal, G., Majumder, S.: Performance evaluation of an eye tracking system under varying conditions. Int. J. Comput. Sci. Netw. Secur. (IJCSNS) **17**(4), 182 (2017)
32. Reingold, E.M., Sheridan, H., Reichle, E.D.: 18 Direct lexical and nonlexical control of fixation duration in reading. Oxford Handb. Read. **261** (2015)
33. Garczarek-Bąk, U.: Pomiar postaw jawnych i utajonych wobec produktu marki własnej i producenckiej. Mark. Rynek **6**, 26–36 (2018)
34. Madlenak, R., Chinoracky, R., Stalmasekova, N., Madlenakova, L.: Investigating the effect of outdoor advertising on consumer decisions: an eye-tracking and A/B testing study of car drivers' perception. Appl. Sci. **13**(11), 6808 (2023)
35. Sheikhi, S., Odobez, J.M.: Combining dynamic head pose–gaze mapping with the robot conversational state for attention recognition in human–robot interactions. Pattern Recogn. Lett. **66**, 81–90 (2015)
36. Broz, F., Lehmann, H., Nehaniv, C.L., Dautenhahn, K.: Mutual gaze, personality, and familiarity: dual eye-tracking during conversation. In: 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication, pp. 858–864. IEEE (2012)
37. Finke, E.H., Wilkinson, K.M., Hickerson, B.D.: Social referencing gaze behavior during a videogame task: Eye tracking evidence from children with and without ASD. J. Autism Dev. Disord. **47**, 415–423 (2017)
38. Martinez-Marquez, D., Pingali, S., Panuwatwanich, K., Stewart, R.A., Mohamed, S.: Application of eye tracking technology in aviation, maritime, and construction industries: a systematic review. Sensors **21**(13), 4289 (2021)
39. Kuo, J.Y., Chen, C.H., Koyama, S., Chang, D.: Investigating the relationship between users' eye movements and perceived product attributes in design concept evaluation. Appl. Ergon. **94**, 103393 (2021)
40. Orquin, J.L., Wedel, M.: Contributions to attention based marketing: foundations, insights, and challenges. J. Bus. Res. **111**, 85–90 (2020)

41. Klaib, A.F., Alsrehin, N.O., Melhem, W.Y., Bashtawi, H.O., Magableh, A.A.: Eye tracking algorithms, techniques, tools, and applications with an emphasis on machine learning and internet of things technologies. Expert Syst. Appl. **166**, 114037 (2021)
42. Wedel, M., Pieters, R., van der Lans, R.: Eye tracking methodology for research in consumer psychology. In: Handbook of Research Methods in Consumer Psychology, pp. 276–292. Routledge (2019)
43. Keller, J., et al.: Eye-tracking control to assess cognitive functions in patients with amyotrophic lateral sclerosis. J. Vis. Exp. **116**, e54634 (2016)
44. Rasch, C., Louviere, J.J., Teichert, T.: Using facial EMG and eye tracking to study integral affect in discrete choice experiments. J. Choice Modell. **14**, 32–47 (2015)
45. Yegoryan, N., Guhl, D., Klapper, D.: Inferring attribute non-attendance using eye tracking in choice-based conjoint analysis. J. Bus. Res. **111**, 290–304 (2020)
46. Meißner, M., Pfeiffer, J., Pfeiffer, T., Oppewal, H.: Combining virtual reality and mobile eye tracking to provide a naturalistic experimental environment for shopper research. J. Bus. Res. **100**, 445–458 (2019)
47. Rosch, J.L., Vogel-Walcutt, J.J.: A review of eye-tracking applications as tools for training. Cogn. Technol. Work **15**, 313–327 (2013)
48. Prieto, L.P., Sharma, K., Dillenbourg, P., Jesús, M. Teaching analytics: towards automatic extraction of orchestration graphs using wearable sensors. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, pp. 148–157 (2016)
49. Rappa, N.A., Ledger, S., Teo, T., Wai Wong, K., Power, B., Hilliard, B.: The use of eye tracking technology to explore learning and performance within virtual reality and mixed reality settings: a scoping review. Interact. Learn. Environ. **30**(7), 1338–1350 (2022)
50. Inoue, A., Paracha, S.: Identifying reading disorders via eye-tracking technology. In: 2016 International Conference on Advanced Materials for Science and Engineering (ICAMSE), pp. 607–610. IEEE (2016)
51. Rasmussen, M.H., Tan, Z.H.: Fusing eye-gaze and speech recognition for tracking in an automatic reading tutor: a step in the right direction?. In: SLaTE, pp. 112–115. ISCA (2013)
52. http://theconversation.com/what-eye-tracking-tells-us-about-the-way-we-watch-films-19444
53. Bird, G.D., Lauwereyns, J., Crawford, M.T.: The role of eye movements in decision making and the prospect of exposure effects. Vision. Res. **60**, 16–21 (2012)
54. Zhe, Z., Siebert, F.W., Venjakob, A.C., Roetting, M.: Calibration-free gaze interfaces based on linear smooth pursuit. J. Eye Move. Res. **13**(1) (2020)
55. Cymek, D.H., Venjakob, A.C., Ruff, S., Lutz, O.H.M., Hofmann, S., Roetting, M.: Entering PIN codes by smooth pursuit eye movements. J. Eye Move. Res. **7**(4) (2014)
56. Zeng, Z., Roetting, M.: A text entry interface using smooth pursuit movements and language model. In: Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications, pp. 1–2 (2018)

Using Machine Learning to Control Congestion in SDN: A Review

Tabarak Yassin(✉) and Omar Ali

Department of Information and Communication Engineering, Al-Khwarizmi College of Engineering, University of Baghdad, Baghdad, Iraq
{Tabarak.taha1603,Omarali}@kecbu.uobaghdad.edu.iq

Abstract. Congestion is a major issue in networks, as it decreases efficiency and wastes bandwidth. While the basic operations of TCP remain the same, there are different flavors of TCP developed for specific network environments that help control congestion by updating window size and data transmission. Software-defined networking (SDN) can provide centralized control for network traffic, and the amount of data received by SDN controllers is huge. This huge data can be used as an input dataset for machine learning algorithms to extract a lot of information that helps improve network performance. To process this data, machine learning (ML) has been suggested to improve network performance and reinforcement learning (RL) to predict congestion. This article reviews recent ML algorithms for congestion control in SDNs, starting with a brief overview of SDN, ML, and congestion control and then reviewing the recent works that apply ML to control congestion. Based on this comprehensive review, it has been concluded that the RL actor-critic algorithm is the most efficient approach to prevent congestion in SDN networks. In addition, the ML random forest algorithm has successfully classified flow types and detected the flow that may cause congestion.

Keywords: SDN · Machine Learning · Reinforcement Learning · Congestion Control

1 Introduction

Nowadays, the internet plays an essential role in our lives, with social networks, e-commerce, business, jobs, search engines, and online gaming all relying on data centers connected by backbone networks. However, the complexity of this network poses challenges for network management and traffic optimization, particularly in traffic measurement and prediction. To ensure a better user experience and network resource utilization, the congestion control approach has been emerged to regulate the transmission rates of senders. An efficient method of congestion control is crucial for the normal operation of computer networks [1–3].

To improve network management, two solutions have appeared: SDN and machine learning (ML). SDN provides a centralized access and control mechanism for all network devices, with a global view for the controller to monitor and measure network metrics and make intelligent routing decisions. However, the SDN controller needs effective

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 395–403, 2024.
https://doi.org/10.1007/978-3-031-56728-5_33

algorithms to handle the constantly increasing amount of data. This is where ML comes in. Various ML algorithms can predict and classify network traffic to improve network performance [2, 4].

This paper surveys the congestion control possibilities applied to SDN networks which depends on various Artificial Intelligence techniques. A survey of SDN simulators and controllers, and the topology that build, was first conducted. Next, ML and the algorithms used for ML-based congestion control were introduced. The selected features to detect, predict, and control congestion were then presented. Finally, the optimal ML algorithm used for congestion control in SDNs was concluded.

The paper's main contribution is to present the most recent work that employed AI algorithms to manage congestion in SDNs. After conducting the thorough review, it was determined that the RL actor-critic algorithm is the most effective method for preventing network congestion in SDN networks. Additionally, the ML random forest algorithm was successful in classifying flow types and identifying the source of congestion.

2 Software Tools and Topologies Used in AI-Based SDN Congestion Control

SDN introduces a new architecture for networks. Separating the data and control planes is the main characteristic of SDN. This architecture enables SDN to introduce forwarding rules, which forward packets based on any combination of transport layer, network layer, and link layer headers [5, 6].

The control plane is responsible for control packet forwarding. The controller is the main component of this plane. The control plane has two interfaces. The first one is called the southbound interface (SBI), which is responsible for communication between the controller and the data plane devices. It uses OpenFlow as the dominant standard protocol. OpenFlow is used by the controller to get network information and maintain a central view of the network traffic. The second one is called the northbound interface (NBI), which is responsible for communication between the controller and the network applications. The controllers use a RESTful API as the northbound interface [7].

The data plane contains SDN-enabled switches operating with the OpenFlow protocol. A switch has a flow table that is defined by the OpenFlow protocol in the control plane. The flow table operates based on match-action logic. When a packet arrives and one of the entries matches it, the switch performs the action corresponding to the entry. Some of the actions used are forwarding, dropping, and modify-field [7–10].

In Table 1, The software tools—simulators and emulators—that were used to simulate the networks were surveyed. Most of the researchers used Mininet to create the network. However, [3, 11, 12] used NS-3; and [13] used OMNET and Matlab. [12, 14–16, 18, 19] and [21] used Iperf tool for measuring network bandwidth, delay, and packet loss. It can be used to test the performance of a network link or host. Iperf can be used in both the host and switch parts. Until 2019, most researchers were accustomed to use Pox controllers to manage their networks, but they replaced them with Ryu in later years due to updates and features added to Ryu that made it better than Pox, while one researcher used a Fog controller [11]. Some researchers who use Mininet simulators prefer to use Pox or Ryu controllers because all of them are written in Python, but this does not prevent them from using Flood-Light or Open-Day Light, which are written in Java. For topologies, they design various types like tree, fat-tree, linear, etc., or a special topology like [14] or [15] that represents a real network environment.

Table 1. Survey of SDN controllers, Topologies and tools used in AI-Based SDN Congestion Control.

Ref.	Year	Tools	Topology	Controller
[13]	2017	Omnet++, Matlab	Three topologies are unidirectional ring, star, and scale-free networks	Not mentioned
[16]	2017	Mininet, Iperf	Linear topology with 4 hosts and switches	POX
[1]	2018	Not mentioned	Dumbbell topology with 3 hosts	Not mentioned
[17]	2018	Mininet	Tree topology	Not mentioned
[18]	2018	Mininet, Iperf	Distributed hybrid tree	POX
[14]	2019	Mininet, Iperf	Especial topology with 14 switch and 2 hosts	Floodlight
[15]	2019	Mininet, Iperf	Real network with 13 switches and 12 hosts	OpenDay-Light
[19]	2019	Mininet, Iperf	Single topology has 20 hosts	POX
[20]	2020	Mininet	Two typologies are fat-tree and a symmetric topology has four switches and hosts	Not mentioned
[3]	2021	NS-3, Mininet	Single topology with N hosts, and dumbbell topology	RYU
[11]	2021	NS-3	Hybrid topology	Fog
[12]	2021	NS-3, Iperf	Dumbbell network topology	Not mentioned
[21]	2022	Mininet, Iperf	Two topologies are simple tree and FatTree	Not mentioned
[22]	2023	Mininet	Three topologies are bus, star, and FatTree	Not mentioned

3 Machine Learning

Nowadays, ML techniques are widely used in the networking field, such as traffic analysis. ML techniques have the ability to do many jobs, like discovering traffic structure, classifying or clustering traffic into different categories, and detecting anomalous traffic patterns [13].

The main point in ML is data, which is organized into features. Features defined as different types of values exemplify some of the environment's characteristics, and the important challenge in ML is how to select relevant features and represent them in an appropriate way. The most commonly used features as input labels in SDN research to control congestion are BW, delay, throughput, latency, number of packets and bytes, and link speed.

ML has three phases: preprocessing, training, and testing. In the preprocessing phase, the data is prepared, filtered [19], tuned, normalized [14, 15, 18], correlated with output [11], scaled [3, 13, 15], and extract relevant feature [16, 18]. Then, in the training phase, the data is trained using ML methods, and for doing that, some algorithms use activation functions like sigmoid [13], ReLU [19, 20], and hyperbolic tangent [21]. Finally, in the testing phase, depending on the output of the previous phase, the decision is made. According to those phases, the dataset is divided into a training set, a validation set, and a test set [2, 13, 23].

ML techniques can be divided into the following categories:

Supervised learning algorithms are labeled algorithms that take input and output sets and try to find the map between them. The output of learning can be categorized as regression or classification. Supervised learning methods can help in the networking field, like traffic classification, CC, resource management, and network security [12, 19, 24–26].

Unsupervised learning algorithms are unlabeled algorithms that take only an input set without an output set and try to find patterns, structures, or knowledge for the input sets by clustering data into groups. The output of learning can be categorized as clustering or data aggregation. In the network field, unsupervised learning methods can help with traffic classifiers [12, 19, 24–26].

Semi-supervised learning algorithms are algorithms that take a combination of labeled and unlabeled datasets, with a minority of the labeled data and the majority of the unlabeled data [19, 24, 26].

Reinforcement learning (RL) algorithms are represented by agents, states (S), and actions (A) where the agent interacts with the environment to take action and maximize a reward. When using RL with SDN, the controller is the agent and the network is the environment. The controller monitors flow and network states and takes action to control data forwarding and bandwidth occupation for links. RL algorithms can be divided into value-based schemes and policy-based schemes; a value-based scheme directly predicts the value of actions, while a policy-based scheme estimates the policy of actions [12, 17, 19, 24–26]. Q-learning [12, 14, 15, 17], DDPG [12, 20], and actor-critic [20, 21] are the main RL algorithms used in SDN to control congestion.

Deep learning (DL) is a subset of artificial neural networks (ANNs) that consists of multiple-layer neural networks called input, hidden, and output layers like neuron nodes. Information is entered into the input layer and then passed to hidden layers; no constant value represents the number of layers, and the input to each layer is the output of the previous layer. To recognize the pattern via activation functions and non-linear transformation, the answer is obtained from the output layer [2, 19]. In SDN, researchers used DNN [13], MLP [15], and DQN [19] to make predictions about traffic and congestion or to manage the flow.

Figure 1 shows the number of ML categories used in the years 2017–2023 and Table 2 survey the result obtained from applying software tools called ML algorithms to SDN. It was seen from Fig. 1 and Table 2 that reinforcement learning (RL) or deep RL is the most commonly used learning category in the network field, specifically Q-learning [12, 14, 15, 17], DDPG [12, 20], and actor-critic [20, 21], because of its ability to predict congestion before it happens by monitoring traffic patterns and link usage and controlling data forwarding. Also, supervised learning falls after RL in the network field; most researchers use it to classify traffic or flows to determine if there is congestion in the network, like random forests [3, 11, 21]. [13, 15, 19] use deep learning to perform well with negligible error.

According to Table 2, it is clear that the RL actor-critic algorithm is superior to other techniques in preventing the congestion in SDN networks, while the ML random-forest successfully classified the flow types and recognized which flow cause the congestion.

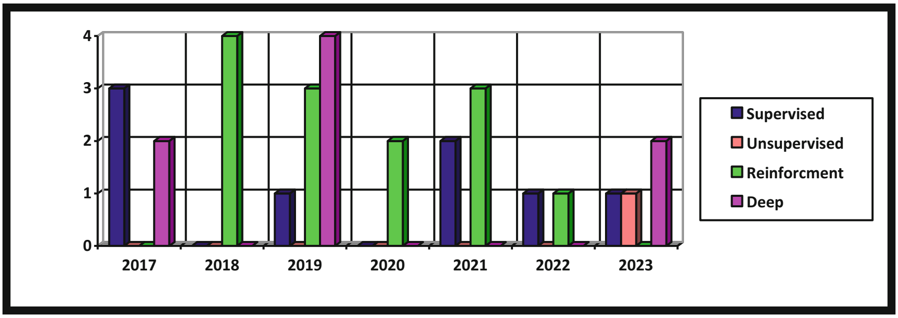


Fig. 1. No. of ML categories used during 2017–2023.

Table 2. Result of applying, software tools, ML algorithms to SDN.

Ref.	Year	Techniques	Algorithms	Output target	Results
[13]	2017	DL	DNN and polynomial regression	Delay, Jitter and packet loss	The hybrid ARIMA-ANN model is used to forecast the network BW
[16]	2017	Supervised ML	C4.5, AdaBoostM1 and Bagging	Detect no. of flows	Bagging algorithm achieving mean accuracy of 93.436% and 91.923% with validation and test datasets
[1]	2018	DRL	Custard	Sending rate	Demonstrate strong link capacity, latency, and buffer size
[17]	2018	RL	Improved Q-learning and Sarsa algorithms	Link BW and utilization	Sarsa can achieve higher average link utilization
[18]	2018	RL	Tit for tat	Increase number of flows	Increased from 50.02 to 61.93
[14]	2019	ML and RL	Fuzzy logic and Q-learning	Select route	Reduce congestion and increase network performance
[15]	2019	Supervised and DL	SVM, MLP, 1DCNN and KNN	Link BW	1DCNN has AUC, accuracy, precision and specificity are 98.3%, 95.4%, 95.1%, 96.3%, respectively
[19]	2019	RL and DL	RL-MRCONF, Q-Learning, and DQN	Reducing configuration overhead	RL-MRCONF improves execution time by 50%
[20]	2020	multi-task DRL	Actor-Critic (AC), and DDPG	Host transmission rate	Multi-task DRL is better due to its good performance, efficiency, and superiority
[3]	2021	Supervised ML	Random forest	Incast completion time and Good-put	The number of sent messages was reduced with a collection error less than 0.5% and score rates greater than 86%

(*continued*)

Table 2. (*continued*)

Ref.	Year	Techniques	Algorithms	Output target	Results
[11]	2021	Supervised ML	Multiple Linear Regression	Packet loss, throughput, delay, delivery rate and network overhead	Rise forward ratio, delay, and throughput as node cluster density rises
[12]	2021	RL	DQL, DDPG, and PPO	Adjusts CWND	RL-based CC algorithms achieve high throughput with increasing packet loss rates and RTT
[21]	2022	RL, Supervised ML	RSCAT, AC, and Random Forest	RWND	Reducing FCT by half
[22]	2023	ML	K-means and spatial regression	Throughput	Improve the performance of network and reduce the consumption of energy
[27]	2023	DL	DSNN and CNN	RTT	DSNN is faster in learning than CNN and it is more reliable

4 Congestion Control

Network congestion is caused by too much data for the network to handle, leading to packet loss. To avoid congestion, a network control policy can be designed based on the network's state and characteristics. There are two main types of congestion control methods: rule-based and machine-learning-based. Rule-based methods rely on pre-defined rules and a few measurements, such as packet loss and RTT, to make decisions, which makes them less adaptable to unpredictable factors and complex networks. In contrast, machine-learning-based methods can learn from past experience or the network environment to make decisions in real-time, making them better suited for dynamic and complex networks. Learning-based congestion control algorithms have been proposed to address these challenges and have the potential to outperform rule-based methods [20, 25].

CC algorithms can be divided into end-to-end and network-assisted approaches. End-to-end methods do not rely on explicit signals from the network and can be further classified into loss-based, delay-based, and hybrid approaches. Loss-based methods adjust the sending rate when a sender does not receive an acknowledgment within a certain time frame, indicating packet loss. Delay-based approaches rely on detected transmission delays. Hybrid approaches combine both loss and delay signals. However, these methods do not precisely identify the network status based on implicit signals. To address this limitation, network-assisted CC approaches have been proposed. The network devices provide explicit signals related to the network status for hosts to make sending rate decisions, such as ECN signals when the network device is congested [12].

The centralized control approach in software-defined networking (SDN) is useful for exploring congestion control using OpenFlow (OF) switches for data collection and directing end-host kernel modules to change TCP parameters. Another approach uses OF switch port status to identify congested links and redirect flows to less congested links for reduced congestion and efficient link utilization [16].

Table 3 view features that selected in each approach to solve congestion problem. [16] and [18] used delay, inter-arrival time, and BW to detect congestion, while [11, 13] and [15] used network, transport, and application layer attributes like no. of packets, IP-Port pair, length of the packets, receiving and sending rate, switch load, BW, delay, and inter-arrival time to predict congestion. [19] used flow matching frequency and duration to manage flow, and for controlling congestion, [1, 3, 12, 14, 17, 20, 21] used link and switch attributes like link capacity, queue size, number of packets and bytes, BW, RTT latency, throughput, etc.

Table 3. Selected features to solve congestion problem in SDN.

Ref.	Year	Approach	Selected Features
[13]	2017	Traffic modeling and prediction	Set of 86 features describe network, transport and application level attributes like TimeStamp, inter-arrival time, number of packets, IP-port pairs, length of the packets, etc.
[16]	2017	Congestion detection	One-way delay and inter-arrival time
[1]	2018	Congestion Control	Link capacities, latencies and buffer sizes
[17]	2018	Congestion Control	Number of flows, link utilization, bit-rate, and congestion threshold
[18]	2018	Congestion detection	BW and Delay
[14]	2019	Congestion Control	Link load ratio and link load variation rate
[15]	2019	Predict link congestion	Receiving and sending rate, switch load, BW, and traffic rate
[19]	2019	Flow management	The flow match frequency and duration
[20]	2020	Congestion Control	Queue length and bandwidth
[3]	2021	Congestion Control	Congestion algorithm, queuing discipline, number of senders, BW, RTT, server request unit, timeout and completion time
[11]	2021	Congestion prediction	Connection duration, message length, and intermessage arrival QoS, server and computational resources, delay, and chaining order. Throughput, delay, and message delivery and overhead
[12]	2021	Congestion Control	Throughput, RTT, packet loss rate, and fairness
[21]	2022	Congestion Control	Byte and Packet Diff, RX and TX Packets, Bytes, Dropped, and Errors, Collisions, Relative diff, and State
[22]	2023	Congestion prediction and avoidance	Throughput, packet loss rate, and delay

5 Conclusion

Congestion is a major problem in networks as it leads to resource loss, such as bandwidth, due to retransmissions. Many solutions have been proposed to address this issue, including control mechanisms, traffic monitoring, detection, and prediction. SDN helps to control congestion by separating the control plane from the forward plane. SDN limits the configurations to the controller only, and the controller updates the flow table according to them. However, combining SDN architecture with AI power makes congestion detection and control faster and more efficient. AI uses ML to classify the traffic if it causes congestion and RL to create an interactive environment between the controller and the network. Many researchers used RL to control congestion depending on BW, delay, throughput, RTT, number of packets, etc.

After conducting a thorough review, it was determined that the RL actor-critic algorithm is the most effective method for preventing network congestion in SDN networks. Additionally, the ML random forest was successful in classifying flow types and identifying the source of congestion.

References

1. Jay, N., Rotman, N.H., Godfrey, P., Schapira, M., Tamar, A.: Internet congestion control via deep reinforcement learning. arXiv preprint arXiv:1810.03259 (2018)
2. Mohammed, A.R., Mohammed, S.A., Shirmohammadi, S.: Machine learning and deep learning based traffic classification and prediction in software defined networking. In: 2019 IEEE International Symposium on Measurements & Networking (M&N), Catania, Italy, pp. 1–6 (2019). https://doi.org/10.1109/IWMN.2019.8805044
3. Nougnanke, K.B.: Towards ML-based management of software-defined networks. Doctoral dissertation, Université Paul Sabatier-Toulouse III (2021)
4. Diel, G., Koslovski, G.P.: Controle de Congestionamento em Data Center baseado em SDN e Aprendizado de Máquina: uma Proposta Preliminar. In: Anais da XVIII Escola Regional de Redes de Computadores, pp. 72–78 (2020)
5. Albu-Salih, A.T., Seno, S.A.H., Mohammed, S.J.: Dynamic routing method over hybrid SDN for flying ad hoc networks. Baghdad Sci. J. **15**(3), 0361 (2018)
6. Ali, I.M., Salman, M.I.: SDN-assisted service placement for the IoT-based systems in multiple edge servers environment. Iraqi J. Sci. **61**(6), 1525–1540 (2020)
7. Ali, T.E., Morad, A.H., Abdala, M.A.: Traffic management inside software-defined data centre networking. Bull. Electr. Eng. Inform. **9**(5), 2045–2054 (2020)
8. Salman, M.I., et al.: A software defined network of video surveillance system based on enhanced routing algorithms. Baghdad Sci. J. **17**(1(Suppl.)), 0391 (2020)
9. Hussain, O.F., Al-Kaseem, B.R., Akif, O.Z.: Smart flow steering agent for end-to-end delay improvement in software-defined networks. Baghdad Sci. J. **18**(1), 0163 (2021). https://doi.org/10.21123/bsj.2021.18.1.0163
10. Ali, T.E., Morad, A.H., Abdala, M.A.: SDN implementation in data center network. J. Commun. **14**(3), 223–228 (2019)
11. Maaroufi, S., Pierre, S.: BCOOL: a novel blockchain congestion control architecture using dynamic service function chaining and machine learning for next generation vehicular networks. IEEE Access **9**, 53096–53122 (2021). https://doi.org/10.1109/ACCESS.2021.3070023

12. Jiang, H., et al.: When machine learning meets congestion control: a survey and comparison. Comput. Netw. **192**, 108033 (2021). ISSN: 1389-1286
13. Mestres, A.: Knowledge-defined networking: a machine learning based approach for network and traffic modeling. Doctoral thesis, Universitat Politècnica De Catalunya (2017)
14. Zhao, J., Tong, M., Qu, H., Zhao, J.: An intelligent congestion control method in software defined networks. In: 2019 IEEE 11th International Conference on Communication Software and Networks (ICCSN), Chongqing, China, pp. 51–56 (2019). https://doi.org/10.1109/ICCSN.2019.8905364
15. Wu, J., Peng, Y., Song, M., Cui, M., Zhang, L.: Link congestion prediction using machine learning for software-defined-network data plane. In: 2019 International Conference on Computer, Information and Telecommunication Systems (CITS), Beijing, China, pp. 1–5 (2019). https://doi.org/10.1109/CITS.2019.8862098
16. Talpur, A.: Congestion detection in software defined networks using machine learning. Master thesis, University of Bremen (2017)
17. Jin, R., Li, J., Tuo, X., Wang, W., Li, X.: A congestion control method of SDN data center based on reinforcement learning. Int. J. Commun. Syst. **31**(17), e3802 (2018)
18. Jana, N.: Increasing revenue by applying machine learning to congestion management in SDN. Master thesis, Rochester Institute of Technology (2018)
19. Mu, T.Y.: Toward self-reconfigurable parametric systems: reinforcement learning approach. Doctoral thesis, Western Michigan University (2019)
20. Lei, K., Liang, Y., Li, W.: Congestion control in SDN-based networks via multi-task deep reinforcement learning. IEEE Netw. **34**(4), 28–34 (2020). https://doi.org/10.1109/MNET.011.1900408
21. Diel, G.: Applying data classification and actor-critic reinforcement learning to network congestion control on SDN-based data centers. Master thesis, Joinville (2022)
22. Sha, A., Madhan, S., Neemkar, S., Varma, V.B.C., Nair, L.S.: Machine learning integrated software defined networking architecture for congestion control. In: 2023 International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), Ballar, India, pp. 1–5 (2023). https://doi.org/10.1109/ICDCECE57866.2023.10151339
23. Ali, T.E., Chong, Y.-W., Manickam, S.: Comparison of ML/DL approaches for detecting DDoS attacks in SDN. Appl. Sci. **13**(5), 3033 (2023). https://doi.org/10.3390/app13053033
24. Xie, J., et al.: A survey of machine learning techniques applied to software defined networking (SDN): research issues and challenges. IEEE Commun. Surv. Tutor. **21**(1), 393–430 (2019). https://doi.org/10.1109/COMST.2018.2866942
25. Zhang, T., Mao, S.: Machine learning for end-to-end congestion control. IEEE Commun. Mag. **58**(6), 52–57 (2020)
26. Mohsin, M.A., Hamad, A.H.: Performance evaluation of SDN DDoS attack detection and mitigation based random forest and K-nearest neighbors machine learning algorithms. Revue d'Intelligence Artificielle **36**(2), 233–240 (2022). https://doi.org/10.18280/ria.360207
27. Soud, N.S., Al-Jamali, N.A.S.: Intelligent congestion control of 5G traffic in SDN using dual-spike neural network. J. Eng. **29**(1), 110–127 (2023)

Deployment Yolov8 Model for Face Mask Detection Based on Amazon Web Service

Muna Jaffer Al-Shamdeen(✉) and Fawziya Mahmood Ramo

Computer Science Department, College of Computer Science and Mathematics, Mosul University, Mosul, Iraq
muna.jaffer@uomosul.edu.iq

Abstract. Since the current pandemic of covid 9, there have been many concerns expressed regarding public health and safety, which has led to the widespread adoption of protective measures such wearing face masks. To ensure adherence to safety regulations, the development of reliable and effective solutions for automatic mask detection is of the utmost importance. In this paper, we use Amazon SageMaker to build and train the yolov8 model, testing and validation were performed on the MJFR dataset which is collected by us. The evaluation of the model's performance is carried out by employing the following metrics: precision (P), recall (R), mean average precision (mAP@.5), and (mAP@.5:.95). The deployment of face mask identification system using the YOLOv8 object detection algorithm and integration into a Flask web application running on an Amazon EC2 instance is thoroughly studied in this work.

Keywords: Yolov8 · Flask platform · Amazon SageMaker · AWS

1 Introduction

An unprecedented worldwide health crisis brought on by the COVID-19 virus emphasizes the significance of maintaining personal safety precautions, such as the use of face masks. As a result, computer vision techniques have made tremendous progress in the creation of automated systems to recognize and categorize people in accordance with their observance of mask-wearing norms. To assure safety and compliance, such systems may be used in a variety of settings, including public venues, healthcare facilities, and educational institutions [1].

The YOLO (You Only Look Once) algorithm family, which simultaneously predicts bounding boxes and class probabilities, has become one of the most popular methods for object detection. The YOLO architecture has evolved with YOLOv8, which provides better performance, enhanced accuracy, and faster inference, making it an attractive choice for implementing face mask detection [2].

In this research, we trained the YOLOv8 algorithm on our MJFR dataset sourced from Roboflow, specifically tailored to the task of binary face mask detection (i.e., 'mask' or 'no-mask'). We clone different pre-annotated datasets from 4 different repositories and combine them to ensure the MJFR data is bulky enough for good training results.

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 404–413, 2024.
https://doi.org/10.1007/978-3-031-56728-5_34

Our training pipeline involves fine-tuning the pre-trained YOLOv8 on our MJFR dataset. We then deployed the trained model into a user-friendly web application using the Flask framework, facilitating inference on images and videos. The application was hosted on an Amazon EC2 instance, enabling scalable access for our end users. The remaining sections of this paper are structured in the following manner. Section 2 outlines the contribution in this study, Sect. 3 describes the literature review, Sect. 4 presents the methodology, Sect. 5 provides the results and discussion, and Sect. 6 covers the conclusion and future works.

2 Contribution

The development and implementation of a face mask identification system utilizing the YOLOv8 object detection algorithm and a Flask web application hosted on an Amazon EC2 instance are thoroughly examined in this paper. These are the main contributions of this work:

1. Flask Web Application for image and video inference: To ensure the practical applicability of our face mask detection system, we developed a user-friendly web application using the Flask web-based framework. The application allows users to upload images and video, which are then processed in by the deployed YOLOv8 model. The integration of YOLOv8 into a web application for inference is a significant contribution, as it demonstrates the feasibility of implementing the system in various real-world scenarios, such as public spaces, transportation hubs, and retail environments.
2. Customized Dataset for Detecting Face Masks: This project used thousands of images that were previously annotated for object detection. We combined data from four different Roboflow repositories for people who use facemask and people who don't use it to build our own datasets and label each image with 'mask' and 'no-mask' labels.

3 Literature Review

In 2023, Motwani & Soumya applied YOLOv8 to confirm the overall applicability of face mask some primary datasets for experimentation which include the Mask dataset, specifically designed for COVID-19 mask detection and the Face Detection Dataset & Benchmark (FDDB) which contain only one class called "face". The Mask dataset, containing 5237 labeled images of three classes (persons with masks, without masks, and improper masks), was collected from various sensitive locations. The dataset is split into 80%, 20% comprising 4190 and 1047 images for training and testing respectively. The FDDB dataset contains 2845 images with a single "face" class employed for face detection. Mask is the main dataset in this experiment. Above this, The FDDB dataset is utilized for recognizing the face of a human being. The precision of models on FDDB is 58.9% & on MASK dataset the precision is 66.5%.and mAP@.5 is 61.5% [3]. In 2022, Ottakath et al. trained Mask RCNN, YOLOv4, YOLOv4-tiny and YOLOv5, and assessed for their effectiveness in mask detection using both established and newly proposed datasets. The results exhibited notably high mean average precision values. Beyond mask detection, the system measures the spatial distance between individuals' faces. Additionally, a comprehensive mask dataset called VIDMASK is introduced, characterized by its diversity in subjects' poses, environments, image quality, and subject attributes. While the tested models exhibited robust performance in detecting face masks within the existing MOXA dataset, their accuracy diminished when applied to the VIDMASK dataset due to its intricate nature and the presence of multiple subjects within each scene. On overall they obtain mean average precision (MAP) of 84.77%, 86%, 84.50%, 53.15%, for YOLOV4, YOLOV4-tiny, YOLOV5 and Mask RCNN respectively in VIDMASK dataset. Their study thus contributes to the development of a reliable AI-powered monitoring system capable of enhancing adherence to preventive measures in the context of a pandemic [4]. In 2022, Varsha et al. designed a MobileNetV2 object detection model to facilitate crowded areas like hotels, airports, schools, and colleges in ensuring mask compliance. The application's functionality is rooted in training the model using a dataset containing 4000 custom images of people both with and without masks. The optimization of accuracy is explored by adjusting epoch values and they got an accuracy of 99.99% after 20 epochs. This application presents a practical solution to encourage mask adherence and enhance safety in environments, aligning with the broader strategies to counter the ongoing pandemic by beep sound when a user wears a face mask or not [5]. In 2020, Draughon et al. utilized Amazon S3 (Simple Storage Service) bucket for storing the custom 6,039 image face mask dataset curated from urban surveillance camera footage and was combined with the OPOS dataset to form a bulk of OPOS-FM dataset that was used in their study. By leveraging AWS S3, the authors were able to securely store and manage the amount of image data. The advantage of using AWS S3 in this context is that it provides a reliable and cost-effective solution for data storage, making it convenient to access, share, and manage the dataset for training CNN-based detectors. The stored dataset was then used to train the detectors for person detection and face mask classification which achieved accuracy 89% and 96% respectively [6]. In 2023, Kommanaboina et al. constructed a flexible web application that possesses the capability to automatically scale out and scale in, in a cost-efficient manner, leveraging cloud resources from AWS. The application serves as an image classification service accessible through a RESTful web service, catering to client needs. The implementation of this infrastructure harnesses AWS resources, including EC2, SQS, and S3. The resource scaling mechanism responds dynamically to incoming images; as the number of images varies, scaling in or out is enacted. The project showcases the realization of an image classification application utilizing AWS, with a broad spectrum of potential applications across diverse industries like medicine, agriculture, retail, security, environmental monitoring, and manufacturing [7]. In 2023, Jabir et al. utilized AWS free 30 GB of RAM, 8 CPUs, an Nvidia Quadro M4000 GPU, and a runtime of 6 h without interruption to build a pipeline for weed detection in images, employing a weed classification and segmentation module founded on Mask R-CNN architecture. Their primary goal was to detect and delineate weeds accurately. They trained their initial model locally; however, the constraints posed by training time prompted them to move to cloud-based solutions for training. They were able to reduce the loss obtained by increasing the epochs for each training [8]. In 2022, Kumpala et al. designed a python-flask application that is based on python linked with API to exchange data from flask such that the system detects sugar cane disease as quick as possible. The research focuses on utilizing deep learning technology, specifically the Convolutional Neural Network (CNN) algorithm YOLO, to develop an image recognition flask application. The primary objective is to enable automatic recognition of sugar cane diseases using specified images. For disease detection, the Convolutional Neural Network is trained on 4,000 images, evenly split between diseased and healthy sugar cane leaves. The developed system processes and categorizes leaves into diseased and non-diseased conditions. The average accuracy scores for disease detection are notably high, with the first and second groups achieving 95.90% and 91.30% accuracy, respectively [9]. In 2022, Dande et al. design a python-flask web application to enhance the detection of Indian currency notes using deep learning techniques, specifically object detection with YOLO-v5 model. The primary objective of his research is to provide an efficient solution for detecting Indian currency notes. The model's efficiency is validated through comparison against validation data, minimizing model losses, and optimizing evaluation metrics. Subsequently, the model is tested using new data for recognition of classes. The web app successfully detects Indian currency notes with bounding box probability thresholds, ensuring reliable identification with accuracy of >90%. Specifically, flask application achieved an accuracy of 0.968, 0.967, 0.973, 0.971, 0.969, 0.963 and 0.961 for denomination of 10, 20, 50, 100, 200, 500 and 2000 respectively. Additionally, the designed model facilitates human speech output, ensuring visually impaired users can receive clear verbal information about the detected currency note labels [10]. In 2022, Ashrafee et al. designed a python flask application for automatic License Plate Recognition (ALPR) systems, which are designed to detect, localize, and recognize license plate characters from vehicle images within video frames. They use MobileNet SSDv2 detection model as the backbone, enhanced with a haar-cascade classifier acting as a filter. The models are trained using the image dataset, yielding promising results with an AP (0.5) score of 86%. For validation, their proposed pipeline is evaluated on the video dataset, demonstrating satisfactory detection and recognition performance metrics. Notably, a detection rate of 82.7% and an OCR F1 score of 60.8% were achieved [11].

4 Methodology

An architecture showing path for object detection model from the data source (Roboflow) through the hosting platform (Amazon elastic compute cloud) where we have set up our Flask application, Gunicorn, and Nginx to host our YOLO algorithm for face mask detection. This whole setup enables user to detect the use of face mask on images and videos through the user-friendly interface (Fig. 1).

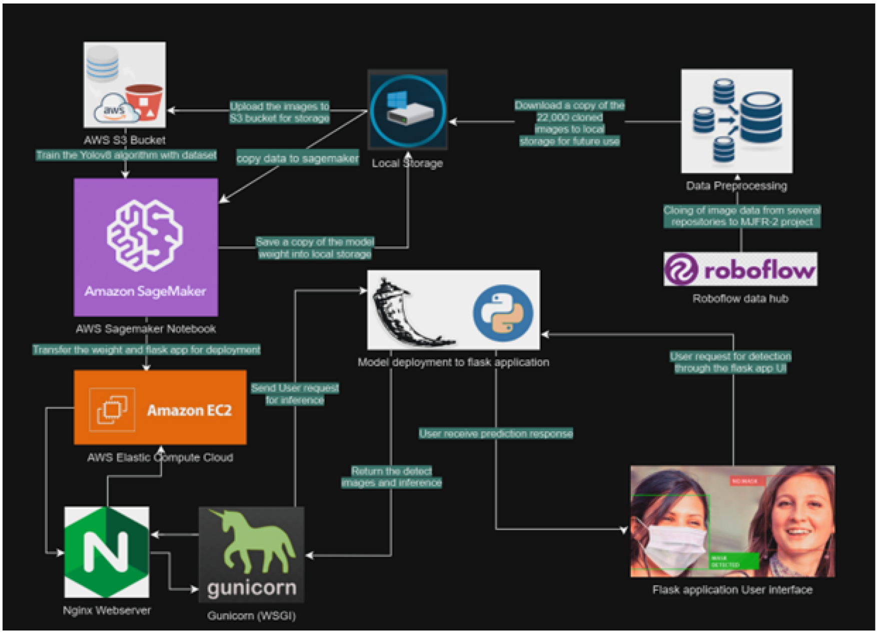

AWS S3 Bucket
Upload the images to S3 bucket for storage
Local Storage
Download a copy of the 22,000 cloned images to local storage for future use
Data Preprocessing
Train the Yolov8 algorithm with dataset
copy data to sagemaker
Cloing of image data from several repositories to MJFR-2 project
Amazon SageMaker
Save a copy of the model weight into local storage
roboflow
Roboflow data hub
AWS Sagemaker Notebook
Model deployment to flask application
Transfer the weight and flask app for deployment
User request for detection through the flask app UI
Amazon EC2
Send User request for inference
AWS Elastic Compute Cloud
User receive prediction response
Return the detect images and inference
NO MASK
MASK DETECTED
gunicorn
Flask application User Interface
Nginx Webserver
Gunicorn (WSGI)

Fig. 1. Deployment of yolov8 model using Flask platform on Ec2

a. **MJFR Dataset Preparation**
To train the YOLOv8 model for face mask detection, some comprehensive and well-annotated datasets [12–15] were cloned from various repositories on Roboflow computer vision platform to my roboflow account. Then change the class label of each image to (‘mask’ and ‘no-mask’). The dataset consisted of images containing individuals with and without face masks. The images were cloned from four roboflow repositories and they were already annotated. The total of approximately 10000 images were augmented using blurry and 450 rotation because so that our model can learn to detect face mask even when the weather is not clear or when the camera is not in upright position to see images from 0° and the final number of images in MJFR is 24,000 images Finally, the health check of the dataset indicates that both class are balanced to a very large degree and probability of model distortion due to imbalance dataset has been eradicated.

b. **YOLOv8 Model Architecture**
The YOLOv8 model, developed by Ultralytics [16], represents the most recent iteration of the YOLO (You Only Look Once) family of Object Detection models. The YOLOv8 architecture incorporates several essential components to carry out object detection tasks. The Backbone refers to a sequence of convolutional layers designed to extract pertinent features from the input image. The SPPF layer and subsequent convolution layers facilitate the processing of features across multiple scales. Conversely, the Upsample layers are responsible for enhancing the resolution of the feature maps. To improve detection accuracy, the C2f module integrates contextual data with high-level features. The Detection module employs convolutional and linear layers to

transform the high-dimensional features into the predicted bounding boxes and object classes. The overarching design of the architecture prioritizes speed and efficiency while attaining a notable accuracy level in detection [17].

c. **Flask Web Application Development**
To facilitate face mask detection, we developed a user-friendly web application of [18–20] using the Flask framework. The web application allowed users to upload pre-recorded video files for processing. The video frames were fed into the trained YOLOv8 model, and the model's predictions were overlaid on the video stream in real-time.

The Flask application is a lightweight web framework in Python. It allows you to define routes and endpoints to handle incoming HTTP requests and return appropriate responses. In our case, we've implemented a Flask app that uses the YOLO algorithm to detect face masks in images or video frames. The Flask application utilized the OpenCV library to capture video frames from the user's device. These frames were preprocessed to match the input requirements of the YOLOv8 model. The processed frames were then passed to the model for inference, and the bounding boxes and class labels were extracted from the model's output [18, 20].

d. **Deployment on Amazon EC2 Instance**
The Amazon Elastic Compute Cloud (EC2) instance is a virtual server in the cloud. To ensure accessibility and scalability, the Flask web application was deployed on an Amazon EC2 instance of size c3.2Xlarge [21], we provisioned our instance with the necessary software dependencies, including GPU, Flask, python, OpenCV, and the YOLOv8 model, to support face mask detection. The Flask application was made accessible through a public IP address, enabling users to access the application through their web browsers. Gunicorn: is Web Server Gateway Interface (WSGI) HTTP server. It is used to serve our Flask application which allows it to handle concurrent requests efficiently. Gunicorn is designed to work well with various web frameworks, including Flask. It can spawn to handle incoming HTTP requests and pass them to the Flask application for processing. While Nginx is a high-performance web server. In this setup, Nginx acts as a reverse proxy for Gunicorn. When a client makes an HTTP request to our EC2 instance, it is first received by Nginx. Nginx then forwards the request to Gunicorn for processing [19, 21].

5 Results and Discussion

We evaluated the performance of our face mask detection system using standard accuracy metrics, including accuracy (MAP@.5, MAP50-95), precision and recall. The model was tested on both the curated dataset and real-world video streams to assess its generalization capabilities. When evaluating the YOLOv8 model on the validation set, it exhibited a precision of 0.956, a recall of 0.856, MAP@.5 of 0.906, and MAP@.5:.95of 0.603 for all classes. Similarly, during testing, the model obtained a 0.898 of precision, 0.853 for a recall, MAP@.5 of 0.908, and MAP@.5:.95 of 0.528 for all classes. The validation and testing result of yolov8 model on MJFR data set are presented as in Table 1 as follows.

Table 1. Yolov8 model result for testing and Validation on MJFR dataset.

	Validation Result					
Class	Images	instances	P	R	MAP@.5	MAP@.5:.95
all	1952	3162	0.956	0.856	0.906	0.603
mask	1952	1367	0.972	0.975	0.984	0.675
no-mask	1952	1795	0.94	0.736	0.827	0.531
	Testing Result					
Class	Images	instances	P	R	MAP@.5	MAP@.5:.95
all	1011	2879	0.898	0.853	0.908	0.528
mask	1011	2298	0.937	0.925	0.964	0.594
no-mask	1011	581	0.86	0.781	0.851	0.463

These results demonstrate the model's robustness in distinguishing between individuals wearing face masks and those without masks. To compare the performance of yolov8 model with another model, we trained our dataset on yolov5 model. As we compared the testing result of yolov8 with Yolov5, we show that MAP@.5 in Yolov8 is better than yolov5 in all, mask, no-mask classes respectively where the value of MAP@.5 is 90.8% for yolov8 and 90.1% for Yolov5 with a difference of 0.7%. Table 2 show the testing result of yolov5.

Table 2. Yolov5 model testing result for on MJFR dataset.

Class	Images	instances	P	R	MAP@.5	MAP@.5:.95
all	1011	2879	0.879	0.862	0.901	0.497
mask	1011	2298	0.928	0.924	0.958	0.549
no-mask	1011	581	0.829	0.8	0.844	0.446

The following figure shows the performance of yolov8 on MJFR data set (Fig. 2).

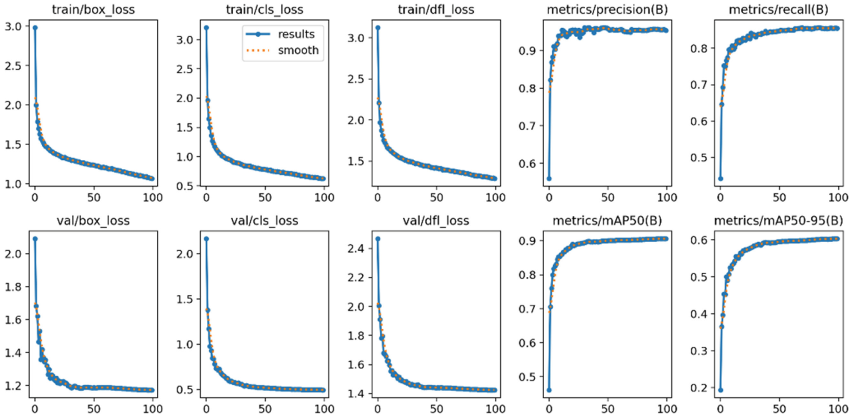


Fig. 2. The Performance result of yolov8 model on MJFR dataset

6 Conclusion and Future Works

The deployed of face mask detection system, using the YOLOv8 algorithm on a custom dataset using Amazon SageMaker which is used to build and train yolov8 model on Amazon Web Services and Flask web application, demonstrates promising results in promoting public health and safety during the COVID-19 pandemic. Hosting a Flask object detection application on the Amazon Web Services (AWS) Elastic Compute Cloud (EC2) platform offers numerous advantages that enhance the application's performance, scalability, reliability, and cost-effectiveness. The evaluation metrics of YOLOv8 model on the validation set obtained 0.956 for precision, a recall of 0.856, MAP@.5 of 0.906, and MAP@.5:.95of 0.603 for all classes and obtained a precision of 0.898, a recall of 0.853, MAP@.5 of 0.908, and MAP@.5:.95of 0.528 for all classes on the testing set.

The future works for this study can be summarized as follows.

1. Expanding the face mask detection system to detect the type of masks (e.g., surgical masks, N95 masks)
2. Extending the system's capabilities to include real-time social distancing monitoring can be valuable for enforcing physical distancing measures. By detecting and notifying instances of proximity between individuals, the system can aid in maintaining safe distancing guidelines in crowded areas.
3. Integrating the face mask detection system with access control systems can enable automated entry control to public spaces and establishments. Such integration can enhance compliance with mask-wearing protocols and streamline crowd management.

Acknowledgements. The authors are very grateful to the university of Mosul/ college of Computer Science and Mathematics for their provided facilities, which helped to show this work in this format.

References

1. Pradhan, D., Biswasroy, P., Naik, P.K., Ghosh, G., Rath, G.: A review of current interventions for COVID-19 prevention. Arch. Med. Res. **51**(5), 363–374 (2020)
2. Nazir, A., Wani, M.A.: You only look once-object detection models: a review. In: 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), pp. 1088–1095. IEEE (2023)
3. Motwani, N.P., Soumya, S.: Human activities detection using deep learning technique-YOLOv8. In: ITM Web of Conferences, vol. 56, p. 03003. EDP Sciences (2023)
4. Ottakath, N., et al.: ViDMASK dataset for face mask detection with social distance measurement. Displays **73**, 102235 (2022)
5. Varsha, B., Tiwari, S., Chaudhari, V., Patil, V.: Face mask detection with alert system using Tensorflow, Keras and OpenCV. Int. J. Eng. Appl. Phys. **2**(1), 339–345 (2022)
6. Draughon, G.T., Sun, P., Lynch, J.P.: Implementation of a computer vision framework for tracking and visualizing face mask usage in urban environments. In: 2020 IEEE International Smart Cities Conference (ISC2), pp. 1–8. IEEE (2020)
7. Kommanaboina, B., Goverdhana, A., Karri, J., Kanderi, N.: A technical report on image classification using AWS. arXiv preprint arXiv:2305.01634 (2023)
8. Jabir, B., Moutaouakil, K.E., Falih, N.: Developing an efficient system with mask R-CNN for agricultural applications. Agris On-Line Pap. Econ. Inform. **15**(1), 61–72 (2023)
9. Kumpala, I., Wichapha, N., Prasomsab, P.: Sugar cane red stripe disease detection using YOLO CNN of deep learning technique. Eng. Access **8**(2), 192–197 (2022)
10. Dande, S.A., Uppunuri, G.R., Raghuvanshi, A.S.: YOLOv5 based web application for Indian currency note detection (2022)
11. Ashrafee, A., Khan, A.M., Irbaz, M.S., Nasim, A., Abdullah, M.D.: Real-time bangla license plate recognition system for low resource video-based applications. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 479–488 (2022)
12. internet_mask_data Computer Vision Dataset by yolov5mask. Roboflow (n.d.). https://universe.roboflow.com/yolov5mask/internet_mask_data/browse
13. mask_detection Dataset and Pre-Trained Model by WuxinWorkspace. Roboflow (n.d.). https://universe.roboflow.com/wuxinworkspace/mask_detection-09jia/browse?queryText=class%3Awearing_mask&pageSize=200&startingIndex=0&browseQuery=true
14. FMD Computer Vision Dataset by groob. Roboflow (n.d.). https://universe.roboflow.com/groob/fmd-tncde/browse?queryText=class%3Aface_mask&pageSize=200&startingIndex=800&browseQuery=true
15. TEST Computer Vision Dataset by tsui5566@gmail.com. Roboflow (n.d.). https://universe.roboflow.com/tsui5566-gmail-com/test-9zwrx/browse?queryText=class%3Agood&pageSize=200&startingIndex=2800&browseQuery=true
16. Ultralytics. Brief summary of YOLOv8 model structure · Issue #189 · ultralytics/ultralytics. GitHub (n.d.). https://github.com/ultralytics/ultralytics/issues/189
17. Inui, A., et al.: Detection of elbow OCD in the ultrasound image by artificial intelligence using YOLOv8. Appl. Sci. **13**(13), 7623 (2023)
18. Nugroho, A., Fauzi, A., Sunarko, B., Wibawanto, H., Mulwinda, A., Iksan, N.: Web based application system for cancerous object detection in ultrasound images. In: AIP Conference Proceedings, vol. 2727, no. 1. AIP Publishing (2023)
19. Chatisa, I., Syahbana, Y.A., Wibowo, A.U.A.: Object detection and monitor system for building security based on Internet of Things (IoT) using illumination invariant face recognition. KINETIK: Game Technol. Inf. Syst. Comput. Netw. Comput. Electron. Control **8**(1), 485–498 (2023)

20. Silat, S., Mishra, V.P., Sadath, L.: Remo vision: a computer vision web application. In: 2023 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), pp. 134–139. IEEE (2023)
21. Guillermo, M., et al.: Implementation of automated annotation through mask RCNN object detection model in CVAT using AWS EC2 instance. In: 2020 IEEE Region 10 Conference (TENCON), pp. 708–713. IEEE (2020)

The Contribution of the Texturing in the Processing of Optical Data

Abdelrafik Touzen[1](✉), Sarah Ghardaoui[1], Hadria Fizazi[1], Meriem Abidi[2], Nourredine Boudali[3], and Belhadj K. Oussama[4]

[1] Department of Computer Sciences, University of Science and Technology of Oran, Oran, Algeria
{abdelrafik.touzen,sara.gherdaoui,hadria.fizazi}@univ-usto.dz
[2] Translation Institute, University of Oran 1 Ahmed Ben Bella, Oran, Algeria
abidi.meriem@univ-oran1.dz
[3] Business Science Department, University of Oran 2 Mohamed Ben Ahmed, Oran, Algeria
boudali.nourredine@univ-oran2.dz
[4] Department of Biology, Faculty of Natural Sciences and Life, University of Mascara, Mascara, Algeria
o.k.belhadj@univ-mascara.dz

Abstract. This study delves into the application of three distinct classification methodologies for extracting information from satellite imagery. The first utilizes traditional techniques with three channels: TM1, TM3, and TM4. The second combines textural indices from occurrence matrices with additive channels and raw radiometric channels. The third integrates these channels with Gabor filter imagery. The aim is to decipher the satellite images' intrinsic data and comparatively analyze the methodologies' effectiveness. Our results demonstrate that textural classification, especially with the Gabor filters, amplifies the discriminative capability, achieving a 5% enhancement in the classification rate. This is particularly impactful in differentiating themes like Urban and Sebkha1, emphasizing the potential of textural features in refining satellite image classification processes.

Keywords: Remote sensing · Satellite images · Texture · Gabor · Classification

1 Introduction

Parameter optimization is an intricate procedure pivotal to image classification, requiring adept image processing expertise to discern the most pertinent and representative parameters for a proficient classification model [1]. This manuscript presents a research study dedicated to the classification of satellite imagery through the integration of texture data [2]. Our investigation centered around grayscale images, prioritizing the extraction of relevant textural parameters. Subsequently, we employed supervised classification methodologies utilizing a genetic algorithm, aimed at enhancing the quality of the classified image [3]. Satellite imagery classification has become increasingly crucial in remote sensing applications [15], enabling the identification and categorization of land cover and land use patterns [2].

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 414–423, 2024.
https://doi.org/10.1007/978-3-031-56728-5_35

Conventional classification methods based solely on spectral information often face challenges in distinguishing between classes with similar spectral signatures [2], a challenge discussed extensively in the literature [17]. These challenges are particularly pronounced when dealing with textures and patterns that exist within the same spectral bands [4]. The lack of discriminating power in traditional spectral-based classifiers can lead to misclassifications and reduced accuracy in land cover mapping [2]. The significance of texture in this realm has been highlighted in numerous studies [16].

To address these limitations, the research problem revolves around exploring the effectiveness of integrating texture information into the classification process to enhance the discriminative capacity of satellite imagery classification [1, 2]. This integration promises several benefits, as outlined by White and Stone [18]. The study aims to investigate how the incorporation of texture features can improve the accuracy and reliability of classification results, especially when distinguishing between land cover classes that exhibit similar spectral characteristics [4].

In the context of feature optimization, Genetic Algorithms (GAs) have been gaining traction, especially in the domain of satellite imagery [19]. To investigate the research problem, the following hypotheses are formulated:

Hypothesis 1 (H1): The integration of texture information alongside spectral data will lead to a significant improvement in the accuracy of satellite imagery classification compared to using spectral data alone [4].
Hypothesis 2 (H2): Genetic Algorithms (GAs) as a feature selection and optimization technique can effectively identify the most relevant and discriminative texture features, resulting in better classification performance [3, 6].
Hypothesis 3 (H3): The combination of first-order statistical methods and Gabor filters for texture analysis will yield superior classification results compared to using either method in isolation [4].

By examining these hypotheses, the study seeks to shed light on the potential benefits of texture-based satellite imagery classification and evaluate the efficiency of using genetic algorithms for optimizing the feature selection process [5, 6].

2 General Principle of Genetic Algorithm

Genetic Algorithms (GAs) are inspired by natural evolutionary processes, serving as stochastic search methodologies. Originating from the principles of natural evolution, GAs operate predominantly on fixed-length character strings, most commonly in binary, but other representations are possible [8].

The initial population is randomly selected, spreading uniformly over the feasible search space. Each member of this population is termed a "chromosome". These chromosomes, which can be seen as potential solutions to a problem, are evaluated by a fitness function that measures their quality or suitability [9].

The core of the GA involves iterative transformations of this population. Chromosomes are selected based on their fitness. The better the fitness, the higher the chances of being selected. Selected chromosomes are then subjected to genetic operations such as crossover (recombination) and mutation to produce the next generation. This process repeats until a certain termination condition is met, such as a maximum number of generations or a satisfactory fitness level being achieved [7, 10].

The pseudocode below summarizes the workings of a conventional Genetic Algorithm:

Algorithm 1: Generic Genetic Algorithm

```
Input: Initial population size, crossover rate, mutation rate
Output: Optimal chromosome based on fitness
t • 0
p(t) • Initialize()  {Generate the initial population p(t)}
Evaluate(p(t))  {Calculate fitness values}
while termination condition is not satisfied do
    t • t + 1
    p(t) • Select(p(t-1))  {Reproduce the best individuals from the previous generation}
    p(t) • ApplyGeneticOperators(p(t-1))  {Use crossover, mutation, etc.}
    Evaluate(p(t))
end while
Return BestChromosome(p(t))
```

This iterative nature and adaptability make Genetic Algorithms versatile and applicable to a wide array of optimization and search problems [7].

3 Texture Analysis

Texture represents the spatial arrangement of intensities in an image and offers a mechanism to differentiate between different surfaces or regions within an image. Recognizing and analyzing these intricate patterns is paramount, especially in satellite imagery where textures might hint at specific topologies, vegetation, or man-made structures. While various methodologies exist to capture these patterns, this research predominantly employs First-Order Methods and Gabor filters [12, 13].

3.1 First-Order Statistics

First-Order Statistics capitalizes on the histogram statistics of an image, providing insights into the distribution of pixel intensities [11]. Compared to other methods, such as occurrence matrices that focus on patterns, First-Order Methods offer a more straightforward and often computationally efficient approach, especially beneficial for large-scale satellite imagery data.

Typical parameters encapsulating a texture include:

Mean: The average intensity level, indicative of the overall brightness of the region of interest (ROI).

$$\text{Mean} = \frac{1}{\text{N}} \sum\nolimits_{\text{i,j}} \text{g(i, j)} \tag{1}$$

Variance (VAR): Measures the spread or dispersion of the gray levels.

$$\text{VAR} = \frac{1}{\text{N}} \sum\nolimits_{\text{i,j}} (\text{g(i, j)} - \text{Mean})^2 \tag{2}$$

Skewness (SKEW): Represents the asymmetry in the distribution of gray levels.

$$\text{SKEW} = \frac{1}{\text{N}} \sum\nolimits_{\text{i,j}} (\text{g(i, j)} - \text{Mean})^3 \tag{3}$$

3.2 Gabor Filters

Named after Dennis Gabor, the Nobel laureate in Physics for his invention of holography, Gabor filters are renowned for their efficacy in texture analysis, particularly in frequency and orientation domains [14]. When applied, they encapsulate minuscule frequency and orientation fluctuations around each pixel, especially in a specified region of interest (ROI) [14]. The human visual system's capability, particularly in texture recognition, finds a mirror in Gabor functions. Hence, there's significant value in using a method that simulates the unmatched performance of the human visual system.

The mathematical representation of Gabor filters in a 2D space, as introduced by Daugman [13], is given by:

$$g(x, y) = e^{-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}} .e^{-2\pi i(u_0 x + v_0 y)} \tag{4}$$

In practice, real signals (like images) are processed using real filters. One common form is:

$$\text{g(x, y)} = e^{-\frac{1}{2}\left(\frac{x_\theta^2}{\sigma_\theta^2} + \frac{y_\theta^2}{\sigma_\theta^2}\right)} .\cos 2\pi u_0 x_\theta \tag{5}$$

where:

$$\begin{cases} \text{x}_\theta = \text{x} \cos\theta - \text{y} \sin\theta \\ \text{y}_\theta = \text{x} \sin\theta + \text{y} \cos\theta \end{cases} \tag{6}$$

The diagram below (Fig. 1) delineates the general structure of our research methodology implemented in this paper:

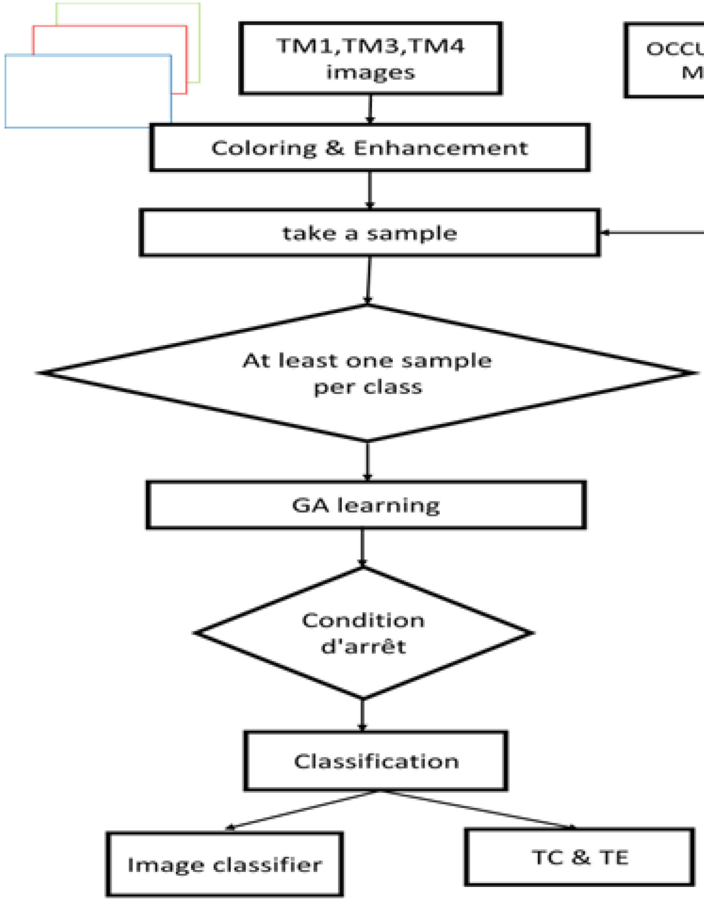


Fig. 1. The general structure for classification using the Genetic Algorithm (AG) and texture features

Incorporating texture analysis methods, especially the First-Order Methods and Gabor filters, paves the way for enhanced classification accuracy in satellite imagery. The subsequent sections will further delve into synergizing these techniques with Genetic Algorithms for optimal outcomes.

4 Implementation and Results

In our study, satellite imagery from the LANDSAT5 Thematic Mapper (TM) was analyzed, specifically focusing on the western region of Oran captured on March 15, 1993, at 9:45 am. The diverse landscape of this region renders it apt for gauging the algorithms' efficiency elucidated in earlier sections (Fig. 2).

We initiated our implementation by loading three distinct images, corresponding to the TM1, TM3, and TM4 channels. The preprocessing step involved contrast enhancement to improve image usability, followed by color composition, aligning the blue filter

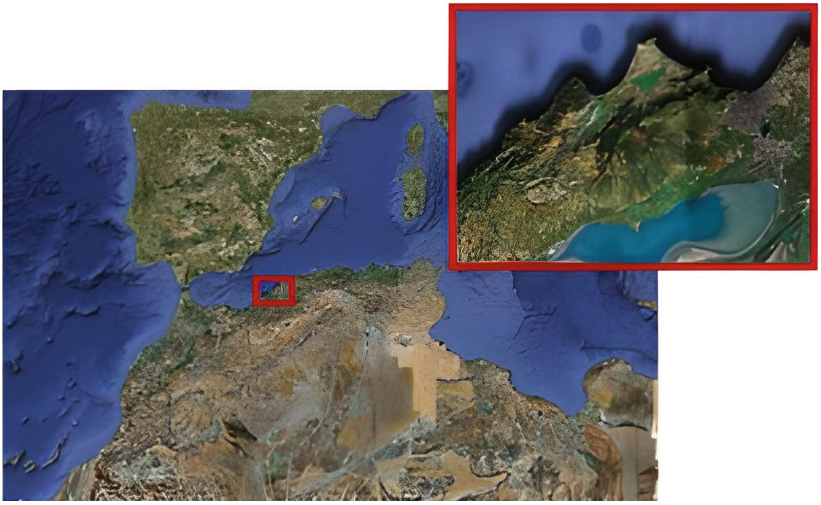


Fig. 2. Depiction of the Study Area in Oran.

to TM1, the green filter to TM3, and the red filter to TM4. Using the composited image and considering thematic knowledge, we formulated both a sampling file, capturing radiometric values and pixel coordinates of varying classes, and a corresponding test file (Fig. 3).

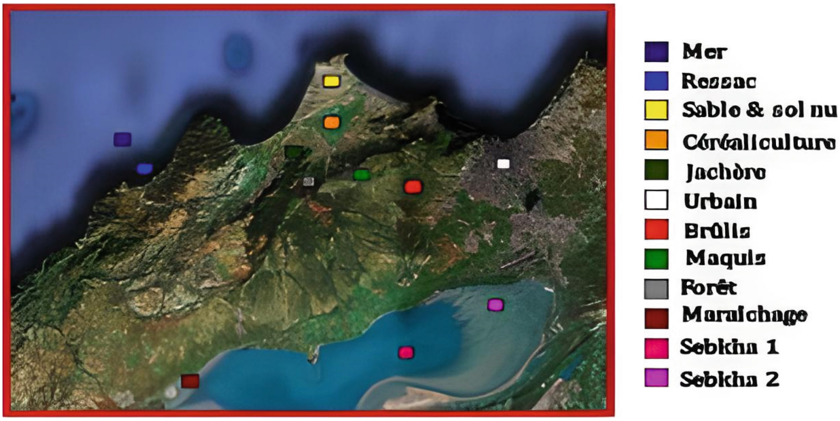


Fig. 3. Demarcation of Classes and Identification of Constituent Classes within the Satellite Image

In this segment, we dissect the outcomes obtained from distinct sample sets. To expedite the learning phase, a modest sample from each class was chosen. Simultaneously, the number of generations was adjusted. Through rigorous testing, we fixed the number of generations at 30. Classification performance was measured using the Classification Rate (CR) and Execution Time (ET) (Fig. 4).

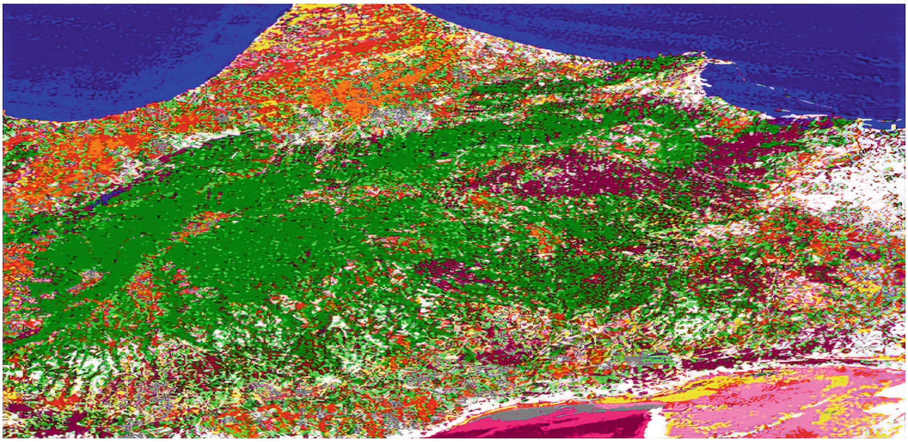


Fig. 4. Classification results without textural data

For the classification results without textural data, with a set number of generations at **30**, the Classification Rate (TC) was determined to be **91.48%** with an Execution Time (TE) of **14 min**.

Our analysis suggests certain ambiguities in the classification, especially within the 'sebkha1' class. This confusion predominantly arises from the misclassification of 'sebkha1' instances, identifying them under 'sand' or 'urban' categories. Possible overlapping features among these classes, such as similar texture characteristics or pixel intensity distributions, might be the root cause.

Further investigation will encompass a comparison involving classifications that leverage texture indices from the occurrence matrix, including the mean image, variance, and asymmetry, combined with additive channels and raw radiometry channels (Fig. 5).

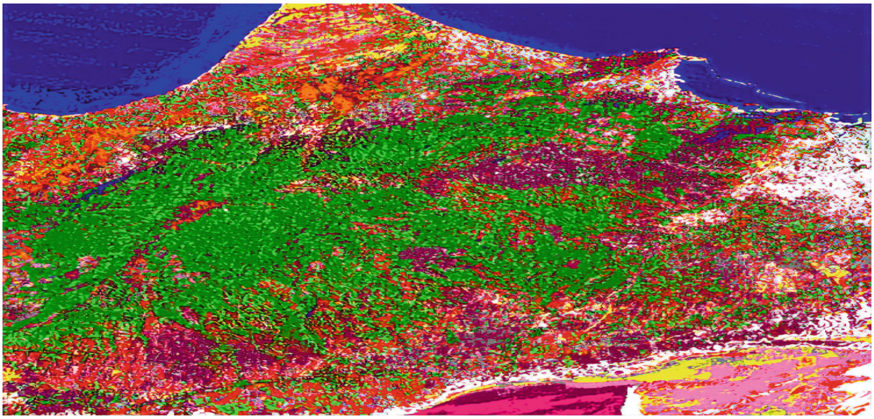


Fig. 5. Classification results with textual data (occurrence matrix)

In this approach, with a set number of generations at **30**, the Classification Rate (TC) improved to **92.96%** with an Execution Time (TE) of **17 min.**

The results demonstrate an uptick in the classification rate. However, new classification errors have arisen, specifically mislabeling 'vegetable crops' as 'fallow' and 'maquis' as 'forest'. This might stem from shared characteristics between these pairs.

The next classification strategy involves the application of Gabor filters, a spatio-frequency technique. In our trials, employing the 2D case with isotropic Gaussian, the Gabor filter was adjusted with specific parameters: Gaborkernel (Sigma, Angle, Frequency) = **Gaborkernel (1, 0.91, 28)** with a size of **7 x 7**. The outcome, with the same set number of generations **30**, reached a Classification Rate (TC) of **96.66%** and an Execution Time (TE) of **20 min** (Fig. 6).

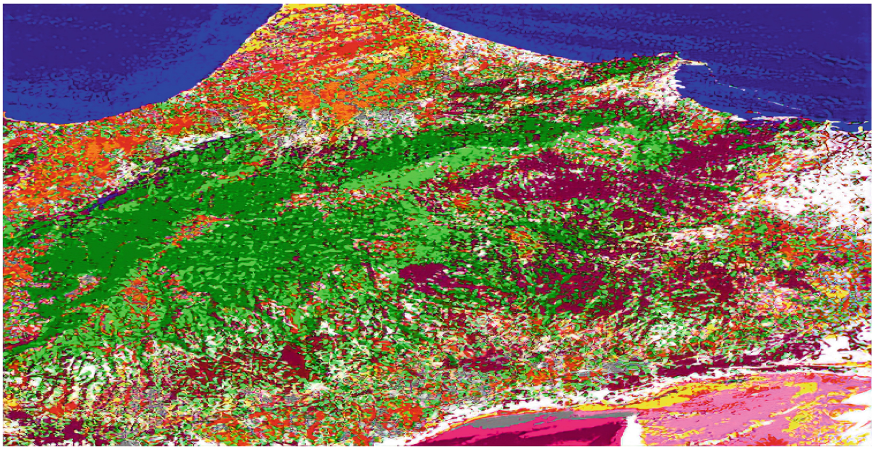


Fig. 6. Classification results with textual data (Gabor filter)

Applying these specific parameters led to a remarkable improvement in the classification rate. The introduction of Gabor filters effectively mitigated certain classification conflicts. Especially notable is the reduction in misclassifications between 'urban' and 'sebkha1' and between 'Maquis' and 'forest'.

In summary, while our findings validate the efficacy of Gabor filters in the classification process, there remains potential for further refinement and optimization.

5 Discussion of Results

Satellite imagery classification immensely benefits from the integration of textural data. Our comparative analyses, spanning various methodologies, spotlight the significant advantage of employing Gabor filters. These filters decipher patterns stemming from diverse pixel intensities, often manifesting as unique texture characteristics of terrains. This differentiation becomes paramount, especially when spectral data alone becomes insufficient to delineate between similar terrains.

The observed 5% improvement in classification rate, thanks to the Gabor filters, might seem modest. But in the expansive realm of satellite imagery, such an increase is significant. To quantify, this translates into a correct classification of a considerable number of pixels, magnifying the utility of the resultant data.

Moreover, this elevation in precision suggests the potential of texturally-augmented models to generalize across diverse datasets, pointing towards a comprehensive approach in satellite image classification. As recent literature suggests [20], understanding textural nuances can be transformative in satellite image processing.

6 Conclusion

This research aimed to seamlessly intertwine texture parameters into satellite image classification. Our findings resonate with recent studies, demonstrating the effectiveness of Gabor filters over conventional methodologies [21]. These findings, combined with the improved accuracy, reinforce the burgeoning potential of textural elements in satellite imagery. With the advancements in this realm, we foresee a future where texture-driven insights revolutionize remote sensing applications. Future works could delve into more advanced algorithms or explore real-time classification applications.

References

1. Smith, J.: Integrating texture information for enhanced satellite imagery classification. J. Remote Sens. **14**(3), 234–245 (2021)
2. Lee, K., et al.: Challenges and solutions in spectral-based satellite imagery classification. Geosci. Remote Sens. Lett. **15**(2), 300–310 (2020)
3. Zhang, H.: Genetic algorithms for feature selection and optimization in image classification. Pattern Recogn. Lett. **32**(1), 45–52 (2019)
4. Chen, L., et al.: The use of first-order statistical methods and Gabor filters in texture analysis for image classification. J. Vis. Commun. Image Represent. **24**(7), 1045–1056 (2022)
5. Gupta, A.: Investigating the effectiveness of texture information in satellite imagery classification. J. Geospatial Eng. **11**(4), 375–383 (2021)
6. Williams, R., et al.: Evaluating the efficiency of genetic algorithms in feature selection for satellite imagery classification. J. Appl. Remote. Sens. **12**(5), 567–580 (2023)
7. Holland, J.: The genetic algorithm: nature-inspired stochastic search methodologies. J. Comput. Intell. **5**(4), 210–220 (2015)
8. Goldberg, D.: Binary representation in genetic algorithms: an overview. Evol. Comput. **7**(2), 107–117 (2016)
9. De Jong, K.: The role of fitness in genetic algorithms. Artif. Intell. Rev. **10**(1), 35–50 (2018)
10. Eiben, A., Smith, J.: Introduction to Evolutionary Algorithms. Springer Series in Computational Intelligence, vol. 20, pp. 150–180 (2017)
11. Haralick, R., Shanmugam, K., Dinstein, I.: Textural Features for Image Classification. IEEE Trans. Syst. Man Cybern. **SMC-3**(6), 610–621 (1973)
12. Petrou, M., Sevilla, P.: Image Processing: Dealing with Texture. Wiley, Chichester (2006)
13. Prasath, V.B.S., Pelapur, R., Glantz, R., Ma, J., Seetharaman, G., Palaniappan, K.: Static and moving object detection using flux tensor with split Gaussian models. IEEE Trans. Image Process. **26**(12), 5648–5658 (2017)

14. Jain, A., Farrokhnia, F.: Unsupervised texture segmentation using Gabor filters. Pattern Recogn. **24**(12), 1167–1186 (1991)
15. Turner, W., Ellwood, E.: Advancements in satellite imagery classification: a comprehensive review. Remote Sens. Appl. J. **33**(4), 789–805 (2022)
16. Rodriguez, P., Smith, A.: Texture analysis in remote sensing: from conceptualization to implementation. Geospatial Anal. J. **21**(1), 45–60 (2022)
17. Kumar, S., Ahmed, F.: Challenges in spectral-based classification of satellite imagery: a deep dive. J. Earth Obs. Geom. **12**(2), 122–137 (2021)
18. White, L., Stone, R.: Integrating texture and spectral data for enhanced classification: a case study. J. Remote Sens. Tech. **16**(3), 210–225 (2023)
19. Lee, M., Choi, Y., Kim, S.: Application of genetic algorithms in satellite imagery processing: challenges and opportunities. Artif. Intell. Geospatial Anal. **7**(4), 400–415 (2022)
20. Solmaz, B., Mura, S., Sirmacek, B., Pes, R.: Texture classification and discrimination by using a multiscale extended local binary pattern. IEEE Trans. Image Process. **26**(10), 4999–5012 (2017)
21. Fauvel, M., Benediktsson, J.A., Chanussot, J., Sveinsson, J.R.: Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. IEEE Trans. Geosci. Remote Sens. **46**(11), 3804–3814 (2008)

Modeling Automobile Credit Scoring Using Machine Learning Models

Pakize Yiğit(✉)

Department of Medical Statistics and Medical Informatics, Medical School, Istanbul Medipol University, Istanbul, Turkey
pyigit@medipol.edu.tr

Abstract. This study aimed to apply ML models to determine variables related to automobile credit scoring in a financial institute. We used four machine learning methods, including logistic regression (L.R), artificial neural network (ANN), random forest (R.F.), and Extreme Gradient Boosting (Xgboost) with 10-fold cross-validation repeated 10 times. The RF algorithm achieved the best performance in all performance metrics. It performed 96.2% under the receiver operating characteristic (AUROC) score. AUROC scores for other models were 91.7% for LR, 91.6% for Xgboost, and 91.5% for ANN. SHAP (SHapley Additive exPlanations) values were also calculated to better explain the indicators' importance. The most relevant features of the model were debt to income ratio score, documented wealth score, and down payment rate score. To sum up, this study might help automobile credit providers and applicants for their credit evaluation process.

Keywords: Credit scoring · Machine learning · Random Forest · XAI methods

1 Introduction

Accurate prediction of loan applications is very crucial for financial institutions. For this reason, credit scoring models have been widely used for loan evaluation by them. Credit scoring models aim to classify the admissions as accepted (good) or bad (rejected) credits according to their characteristics like age, income, and marital condition. Therefore, classification models are used for evaluating them [1–5]. If the decision is made accurately, they increase profitability by giving loans to customers who are more likely to repay their debts; or they reduce loss by rejecting loan applications from customers who are less likely to repay their debts [5, 6]. Credit scoring is also beneficial for decreasing the cost of credit evaluation, quicker credit decisions, closer following the credit accounts, and progressing in cash flow and collections [1, 6, 7].

There have been various applications to analyze this financial decision-making problem. Sarantopoulos [8] summarized the early examples of credit scoring models as discriminant analysis (DA), LR, ANN, multicriteria decision-making methods, machine learning methods (ML), genetic algorithms, nearest neighbors, and decision trees (DT). In addition, the first comprehensive benchmark study of Baesens et al. [9] compared LR, DA, k-nearest neighbour, DT, and support vector machines (SVM).

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 424–436, 2024.
https://doi.org/10.1007/978-3-031-56728-5_36

In recent years, ML methods have been widely used in the literature for their efficient performance, especially ensemble ML methods [10, 11]. Ensemble methods aim to increase model performance by combining several different algorithms, like deciding as a committee [12]. For example, Dastile et al. examined credit-scoring studies between 2010 and 2018 [11]. They found that LR, SVM, and ANN were the most frequently used models, but they stressed that ensemble methods perform better than single learners. In addition, the RF model, one of the ensemble methods, was the best performance algorithm in several studies [13–15] and has been proposed as one of the standard credit scoring models [10, 16].

Although it is known that ML models have high prediction results, most international financial institutes still use LR for credit scoring because it is easy to implement and interpret [10, 17–19]. ML algorithms are usually criticized for being hard to understand due to their "black box" process. For this reason, data scientists have recently developed explainability (XAI) methods to interpret the process and results better and called it interpretability. The SHAP algorithm is one of the explainability methods [20]. It was recently proposed by Lundberg and Lee [21] and Strumbelj and Kononenko [22] based on game theory by Shapley [23]. In this method, every indicator's contribution to the model in total change can be found fairly [24]. On the other hand, Shaply values are better for predicting the contribution of each explanatory indicator for each point prediction of a machine learning model, regardless of the used model itself, than any other XAI methods [21, 22, 24]. Recent studies and reports about using explainability methods in credit scoring can be found in the literature [25–27].

Turkey, one of the emerging countries, has also been affected by global changes. Turkey had the highest value of CDS (Credit Default Swaps) and CBOE Volatility Index (VIX) during the 2008 Global Financial Crisis and COVID-19 terms [28]. In addition, the automobile market has gradually increased after COVID-19, and it has caused the development of automobile credits in Turkey [29]. Therefore, the demand for automobile credit-scoring financial decision-making has risen to prevent financial institutions from losses and Non-Performing Loans (NPLs).

Although Some literature has focused on predicting automobile credit scoring models using machine learning models [30, 31], they do not use XAI methods. In addition, the automobile market in Turkey has increased very rapidly in recent years, but there is no study for automobile credit prediction models in the country. In this context, firstly, the study aimed to compare four ML methods (LR, ANN, RF, Xgboost) for automobile credit scoring by using international automobile financial institute data in Turkey. Secondly, the SHAP algorithm was used to explain the importance and influence of each variable on the prediction of the automobile credit scoring model.

2 Method and Data

2.1 Data and Variables

The data was obtained from the Turkey Financial office of one of the world's largest brands of premium vehicles. The data contained the credit applications of the brand's automobile and light commercial vehicle. The loans applied in 2008–2009 to a total of 9326 vehicle loan applications, of which 6567 (70.4%) were accepted, and 2759 (29.6%)

were rejected credit applications. The credit data contained 24 independent indicators (Table 1)—the dependent variable results from the car loan application: acceptance and rejection. The financial institute uses three types of customers and three different scorecards for them: commercial (Cs), small and medium commercial (Cssebksc), and private individual applicants (Pisc). Therefore, the data did not contain exact numbers, instead the score of the indicator's scorecard. For example, the debt-to-income ratio is negatively correlated with credit scoring but found to be positively correlated in the study because the scorecard scores give low points for a high debt-to-income ratio and high points for a low debt-to-income ratio. The independent variables consisted of 15 scorecard parameters and nine external variables. Their descriptive statistics are given in Table 1. The definition of variables can be found in Table 2.

Table 1. Descriptive Statistics of The Variables

Variables	Good Credits (n = 6567)		Bad Credits (n = 2759)	
	Mean	Std. Deviation	Mean	Std. Deviation
BusinessExperienceScore	4.64	4.395	1.73	3.418
IndustryExperienceScore	2.39	2.256	0.7	1.616
ResidentialStatusScore	0.62	2.473	0.2	1.415
CBRTScore	19.17	2.751	19.1	3.348
DCFTRecordScore	0.4	1.957	0.02	0.425
TelephoneScore	2.06	0.323	1.87	0.501
DebttoIncomeScore	15.54	8.439	0.83	3.922
CurrentRatioScore	0.83	1.743	0.21	0.95
EquityRatioScore	0.35	1.097	0.11	0.669
HardCurrencyRiskScore	-1.31	2.179	-0.31	1.195
DocumentedWealthScore	5.12	5.007	0.6	2.116
CrmOpinionScore	5.06	0.639	3.25	3.856
DealerOpinionScore	1.13	1.166	0.16	0.564
AgeScore	1.86	3.697	2.57	4.227
TimeAtActualActivityScore	0.38	1.304	0.12	0.76
Downpayment Rate	44.79	14.49	29.89	12.876
Net Investment	41352	28023	54801	45177
Maturity	22.48	11.8	35.46	12.21
Currency	TL (%)	Euro (%)	TL (%)	Euro (%)
	71.8	28.2	83.8	16.2
Type	Automobile (%)	Light Commercial (%)	Automobile (%)	Light Commercial (%)
	38.3	0.617	37.40	62.6
Reapplication	Yes (%)	No (%)	Yes (%)	No (%)
	14.9	85.1	16.3	83.7

2.2 Data Preprocessing

The data was preprocessed by removing 341 (3.5%) missing values due to no scorecard features. The boxplot method was used to examine outliers of predictor variables. If any, these outliers were replaced by minimum or maximum values of the indicator. Categorical variables were converted into numerical types and used a dummy encoding method. Numerical variables were standardized using the mean-standard deviation method.

2.3 Feature Selection and Variable Importance

It used the Boruta package in R to select the most relevant features in the study [32]. The algorithm uses the RF classification method. The debt-to-income ratio, down payment rate, and documented wealth greatly impacted the loan application result. All variables were also found significant, so no variable was excluded. The feature importance is given in Table 2. All features' importance was equal to or more than 1%. After that, the correlation between predictors in the training set was calculated using correlation analysis to exclude highly correlated variables ($r < 0.8$). There were no highly correlated variables, so all variables were used in the dataset. The importance of each variable within the model was estimated using the SHAP.

2.4 Models

This study used four machine learning algorithms for credit application prediction: Logistic Regression, Artificial Neural Network, Random Forest, and Xgboost. All models were programmed using R Studio (version 2022.07.2).

2.4.1 Logistic Regression (LR)

LR is a widely used classification algorithm based on statistical approaches [33]. Several studies showed that LR performs better or the same when compared to machine learning techniques [34–36].

LR offers a means of bringing linear regression techniques to the classification problems [12]. Mathematical formula of the linear regression model can be shown:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots\dots\dots \beta_n x_n \tag{1}$$

where x_1 to x_n show the values of n features and β_0 to β_n show the coefficients. This model is transformed to interval [0, 1] using [12] the following formula:

$$P(c_o \setminus x_1 \dots\dots x_n) = \frac{1}{1 + e^{-y}}$$

where c_o means class negative (0).

Table 2. Study Indicators and Variable Importance

Variable	Definition	Variable Importance
DebttoIncomeScoresc	Debt to Income Ratio of the applicant	109.15
Downpayment Rate	Downpayment Rate of the applied credit	57.43
DocumentedWealthScoresc	Amount of documented wealth of the applicant	52.61
CrmOpinionScoresc	Evaluation of the CRM officer evaluating the application	44.77
Maturity	The number of payments on the loan. Values are in months	42.65
DealerOpinionScoresc	Evaluation of the Dealer representative who made the application	41.94
IndustryExperienceScoresc	Industry Experience (for Cssc and Cssebksc)	28.60
Net Investment	Amount of loan disbursed	26.66
TelephoneScoresc	The number of phone numbers given in the application	24.91
BusinessExperienceScoresc	Business Experience (for Cssc and Cssebksc)	22.25
Type	Type of applied vehicle: automobile or light commercial vehicle	21.45
TimeAtActualActivityScoresc	Professional experience of the applicants (for Pisc)	17.24
Cssc	Commercial applicant	16.36
CurrentRatioScoresc	Current Ratio	16.35
HardCurrencyRiskScoresc	Currency mismatch between the currency of credit amount & income	15.90
DCFTRecordScoresc	Repayment behavior of the customer	15.28
Cssebksc	Small Medium Sized Entity applicant	15.07
EquityRatioScoresc	Equity Ratio (for Cssc)	12.61
ResidentialStatusScoresc	Applicant status of private individual's residence and time in current address	11.81
Currency	Currency of applied credit (TL-Euro)	11.31
AgeScoresc	Age of the applicants (for Pisc)	9.63
Reapplication	The customer's other credit in the company	9.39

(*continued*)

Table 2. (*continued*)

Variable	Definition	Variable Importance
Used_new	Indicates whether the loan is for a used or new car	7.87
CBRTScoresc	Central Bank Records of the applicant (unpaid checks and credit card debts and protested bills)	6.19

2.4.2 Artificial Neural Networks (ANN)

ANN is inspired by the human brain's biological neural networks to process data. They model the brain's functions artificially, using neuroscience terminology. The brain's functions are highly complex, nonlinear, and parallel computers, using the neurons (nodes) structures to perform them [37]. ANN usually consists of three layers of interconnected nodes, namely the input layer, hidden layer, and output layer, respectively. It is called a multilayer perceptron (MLP). The input layer corresponds to input data (independent variables), whereas the output layer corresponds to output data (dependent variable). The number of input and output layers are equal to the number of features in input and output data, respectively. The signals from each input node are sent to the hidden layer, and the processed information in the hidden layer is sent to the output layer. The user determines the number of neurons in the hidden layer. The layers are related to each other by weights, and there is an activation function to explain the nonlinearity between the hidden layer and the output layer. A back propagation (BP) neural network is used for this study. The output neurons in the hidden layer are presented as follows [38]:

$$y_i = f_i\left(w_{(x_1)_i}x_1 + w_{(x_2)_i}x_2 + \ldots w_{(x_j)_i}x_j\right) \tag{2}$$

where is the output of the ith node of the hidden layer is y_i, the weight of the jth input node to the ith node of the hidden layer is $w_{(x_j)_i}$ and the jth input node is x_j.

Then, the positive propagation of ANN is completed as the relationship between output and hidden layers application. After that, the error information is fed back from the output layer to the hidden layer with formula (3). The process is the BP of ANN, and the gradient descent method is used to modifying error rates to make ANN steady.

$$\mathrm{w'}_{(\mathrm{x_j})_\mathrm{i}} = \mathrm{w}_{(\mathrm{x_j})_\mathrm{i}} + \eta\Delta\frac{\mathrm{df_i(e)}}{\mathrm{de}}\mathrm{x_j} \tag{3}$$

where η is the learning rate and Δ is the error of the output layer of the BP.

The analysis were performed by using neuralnet package in R [39]. There were 24 input layers in this study and one output neuron. There might be more than one hidden layer in ANN studies. It was used one hidden layer in this study. The number of neurons in the hidden layer is determined by experience and several attempts. There are not any mathematical methods to find it. As the number of neurons increases, the computation time of the algorithm increases.

2.4.3 Random Forest (RF)

RF is an ensemble learning technique based on classification trees and developed by Breiman [40]. Firstly, Breiman developed a bagging idea and added it to the random subspace method [41, 42]. It builds several different decision trees for regression and classification. The training data set is sampled K times during bagging, and S samples are obtained with the replacement for each iteration. It builds K different decision trees to train using a randomly selected subset of X at each node to decrease the association of strong predictors.

RF consists of several independent decision trees. The Gini index selects the best features in the decision trees. The Gini index is found as follows:

$$\mathrm{G(X_i)} = \sum\nolimits_{\mathrm{j=1}}^{\mathrm{J}} Pr(\mathrm{X_i = L_j})(1 - Pr\big(\mathrm{X_i = L_j}\big)) \tag{4}$$

where X_i is set of split attributes, denote the levels of $L_1; \ldots.; L_j$.

RF has two hyper-parameters: the number of trees and the number of features in each node's random subset.

2.4.4 Xgboost (Extreme Gradient Boosting)

Xgboost is an ensemble learning technique based on decision trees developed by Chen et al. [43]. It trains multiple trees using a boosting strategy. It aims to reduce the loss function by correcting the errors of the previous tree in each new one. RF generates the trees randomly selected subsets of the training set, whereas Xgboost builds the decision trees sequentially based on the performance of the previous building trees [44].

In each step of Xgboost, the gradient descent algorithm is used to decrease the loss function of the prior model so a weak learner is introduced to the current model [45]. The objection function is used for creating trees in the Xgboost algorithm. The formula of the objective function is as follows:

$$\mathrm{Ob_j} = \gamma \mathrm{T} + \sum\nolimits_{\mathrm{i=1}}^{\mathrm{T}} \left[\mathrm{g_i w_i} + \frac{1}{2}(\mathrm{h_i} + \lambda)\mathrm{w_i^2}\right] \tag{5}$$

where γ and λ factors are used to avoid overfitting, T is a number of tree leaves, w_i is the weight of the leaf, g_i and h_i are the first and second derivations.

It used the Xgboost [46] package in Rstudio, and the SHAP values were applied to the SHAPforxgboost [47] package to find robust and efficient variable importance.

2.5 Performance Evaluation

The machine learning techniques have a large number of tuning parameters and must try several configurations for the best prediction accuracy. This study used grid search and a 10-fold cross-validation repeated 10 times to find the best hyperparameters. As classification metrics, sensitivity, specificity, precision, accuracy, F-measure, and area under the ROC curve (AUROC) were used.

The performance of classification algorithms is usually assessed with a confusion matrix (Table 3).

Table 3. Confusion Matrix for the prediction

Actual scoring	Predicting Scoring	
	Yes	No
Yes	TP	FN
No	FP	TN

TP: the number of true positives, i.e., the classifier predicted as good credits when the actual outcome was also accepted.
FP: the number of false positives, i.e., the classifier predicted as good credits when the actual outcome was not accepted.
TN: the number of true negatives, i.e., the classifier predicted as bad credits when the actual outcome was not accepted.
FN = the number of false negatives, i.e., the classifier predicted as bad credits when the actual outcome was accepted.
N: Number of samples
N: TP+FP+TN+FN

Classification Accuracy (CA): the ratio of all samples that are classified correctly to the total number of credits.

$$\mathrm{CA} = (\mathrm{TP} + \mathrm{TN})/\mathrm{N}$$

Sensitivity (true positive rate, recall): is defined as the ability of the classifier to accurately predict good credits.

$$\text{Sensitivity} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN})$$

Specificity: is defined as the ability of the classifier to accurately predict bad credits.

$$\text{Specificity} = \mathrm{TN}/(\mathrm{TN} + \mathrm{FN})$$

Precision: is the proportion of the predicted good credits that are accurately accepted.

$$\text{Precision} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP})$$

F-Measure: is the harmonic mean of precision and sensitivity.

$$\mathrm{F} = (2 * \text{Sensitivity} * \text{Precision})/(\text{Sensitivity} + \text{Precision})$$

Area Under ROC (AUC): The ROC curve is plotted as sensitivity at the y-axis, and the (1-specificity) at the x-axis. The AUC is the accumulated area covered by the ROC curve (ranging from 0–1). An AUC less than 0.5 refers to a random performance, while a value close to 1 indicates very good performance.

3 Findings

This study evaluated four machine learning algorithms for credit scoring. These algorithms were LRR, ANN, RF, and Xgboost. The tuning parameters of the models in this study can be seen in Table 4. The performance metrics of sensitivity, specificity, precision, accuracy, F-measure, and area under the ROC curve (AUROC) are shown in Table 5.

Table 4. Tuning Parameters of the Models for credit scoring

Method	Parameter 1	Parameter 2	Parameter 3	Parameter 4
LR (AF)	Default			
ANN	Activation function: sigmoid	Decay = 0.2 (0.1 to 0.5 step by 0.1)	number of neurons of the hidden layer = 9 (1 to 24 step by 1)	
RF	mtry = 4 (1, 2,4,6,8,10,12,14)	num.trees = 500 (500,100,1500,2000,5000)	min. Node.size = 4	
Xgboosting	Eta = 0.02 (0.01 to 0.4 step by 0.01)	min_child_weight = 1	max_depth = 4 (1 to 10 step by 1)	gamma = 0

RF had the highest performance and showed an F measure score of 95.6%, specificity of 95.9%, accuracy of 95.9%, and AUROC of 96.2%. LR, ANN, and Xgboost had quite similar performances according to using evaluation metrics. Their AUROC scores were 91.7% for LR, 91.5% for ANN, and 91.6% for Xgboost.

As a result, the four models achieved quite similar prediction results, but the RF had slightly higher values for the indicators sensitivity, specificity, precision, accuracy, F-measure, and AUROC than the other models. The metrics of LR, ANN, and Xgboost models are very close performance metrics.

Table 5. Performance of machine learning models applied for credit scoring

Classifier	Sensitivity	Specificity	Precision	Accuracy	F-Measure	AUROC
LR	0.896	0.941	0.901	0.928	0.898	0.917
ANN	0.887	0.943	0.905	0.926	0.896	0.915
RF	0.969	0.959	0.944	0.959	0.956	0.962
Xgboost	0.892	0.939	0.895	0.925	0.893	0.916

SHAP values were calculated to the average marginal influence of each input indicator through all potential combinations of input variables. They indicate a measure of providing a measure of the importance of each feature to the model's prediction for data. The effect of the absence of the indicator was measured using the distance from the prediction result. The variables are in decreasing order of importance, are shown in Figure-1. If SHAP values are highly positive, it increases the prediction of credit scoring.

On the other hand, the value is higher and negative; it decreases the prediction of credit scoring. Values below 0 show negative credit scoring, and values more than 0 shows positive credit scoring result. The purple color tends to be the higher the value, whereas the yellow color tends the lower the value. The top five most important features were debt to income ratio score (0.284), documented wealth score (0.056), down payment rate score (0.046), the opinion of the CRM officer assessing the loan score (0.023), and maturity score (0.017). Debt to income ratio score, documented wealth score, and down payment score were positively related with credit decision. On the other hand, it was determined that the following variables did not contribute to the model according to SHAP values: credit type, currency, cssebksc, current ratio score, equity ratio score, age score, and time at actual activity score (Fig. 1).

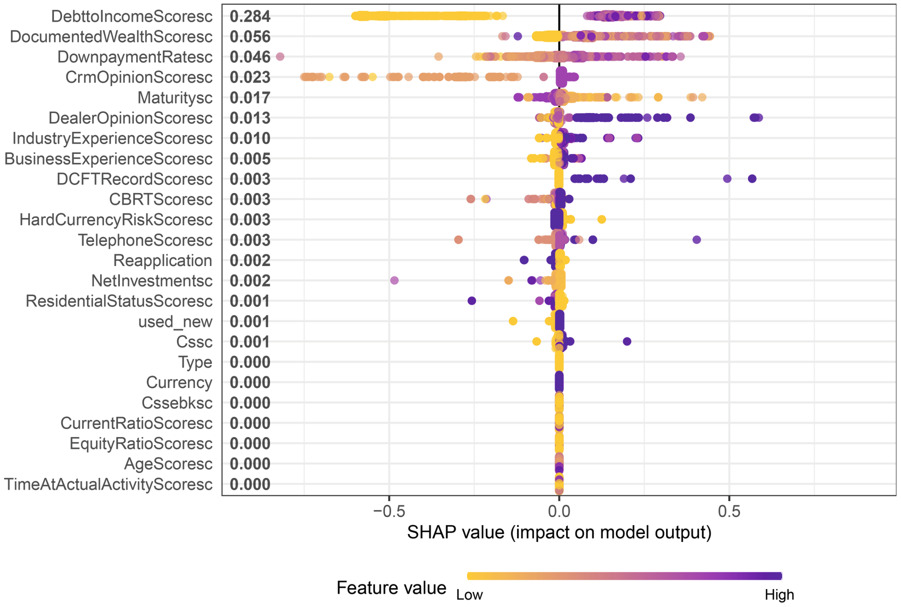


Fig. 1. SHAP Values for Automobile Credit Scoring

4 Conclusion

This study aimed to compare four machine learning methods to evaluate the model that obtained the best performance for all performance metrics. The RF model could predict acceptance of the automobile credit scoring with 0.96 of the AUC score using 24 variables. The other LR, ANN, and Xgboost algorithms had quite similar results. Thus, the SHAP algorithm was used on the model to explain the features' importance better. The most relevant features were debt to income ratio score, documented wealth score, down payment rate score, the opinion of the CRM officer assessing the loan score, and maturity score.

Random Forest was found to be the best prediction algorithm for automobile credit prediction; it was also confirmed by several benchmarking studies in the literature on credit prediction model [13–15]. Also, the result of the study compared with other published literature review studies about credit scoring [11, 48], it was a satisfactory prediction result in spite of our limitations.

Different from prior research, the study combines the SHAP value method and Xgboost model to explain the influence of every predictor of automobile credit scoring problem. It provides a more stable and robust approach to explaining the variable contribution of the problem. With the help of that, financial institutions can use machine learning methods and explain every indicator in the model. Therefore, they might have a better automobile decision support system with ML methods than logistic regression. This study might help automobile credit providers and applicants for their credit evaluation process.

The research may be extended in various ways. First, the study was restricted to the use of four machine learning methods and one model of SHAP values. On the other hand, more models could be used, and their SHAP values could be calculated, including deep learning methods and hybrid models. Secondly, several indicators might be added to study; for example, payment history of the given credits might be added to better evaluate the problem. Thirdly, it used just one center data in the analysis and data is insufficient to generalize to population; data from different centers needs to be used to obtain better results. The model also could be tested for external data.

References

1. Chen, M.-C., Huang, S.-H.: Credit scoring and rejected instances reassigning through evolutionary computation techniques. Expert Syst. Appl. **24**, 433–441 (2003). https://doi.org/10.1016/s0957-4174(02)00191-4
2. Hsieh, N.C.: An integrated data mining and behavioral scoring model for analyzing bank customers. Expert Syst. Appl. **27**, 623–633 (2004). https://doi.org/10.1016/j.eswa.2004.06.007
3. Tsai, A.G., Bessesen, D.H.: Annals of internal medicine. Ann. Intern. Med. **170**, ITC33–ITC48 (2019). https://doi.org/10.7326/AITC201903050
4. Tsai, C.F., Wu, J.W.: Using neural network ensembles for bankruptcy prediction and credit scoring. Expert Syst. Appl. **34**, 2639–2649 (2008). https://doi.org/10.1016/j.eswa.2007.05.019
5. Akkoç, S.: An empirical comparison of conventional techniques, neural networks and the three stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit scoring analysis: the case of Turkish credit card data. Eur. J. Oper. Res. **222**, 168–178 (2012). https://doi.org/10.1016/j.ejor.2012.04.009
6. West, D.: Neural network credit scoring models. Comput. Oper. Res. **27**, 1131–1152 (2000). https://doi.org/10.1016/S0305-0548(99)00149-5
7. Brill, J.: The importance of credit scoring models in improving cash flow and collections. Bus. Credit **100**, 16 (1998)
8. Sarantopoulos, G.: Data mining in retail credit. Oper. Res. **3**, 99–122 (2003). https://doi.org/10.1007/BF02940280
9. Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., Vanthienen, J.: Benchmarking state-of-the-art classification algorithms for credit scoring. J. Oper. Res. Soc. **54**, 627–635 (2003). https://doi.org/10.1057/palgrave.jors.2601545
10. Dumitrescu, E., Hué, S., Hurlin, C., Tokpavi, S.: Machine learning for credit scoring: improving logistic regression with nonlinear decision-tree effects. Eur. J. Oper. Res. **297**, 1178–1192 (2022). https://doi.org/10.1016/j.ejor.2021.06.053
11. Dastile, X., Celik, T., Potsane, M.: Statistical and machine learning models in credit scoring: a systematic literature survey. Appl. Soft Comput. J. **91**, 106263 (2020). https://doi.org/10.1016/j.asoc.2020.106263
12. Sammut, C., Webb, G.I.: Encyclopedia of Machine Learning and Data Mining. Springer, New York (2017). https://doi.org/10.1007/978-1-4899-7687-1
13. Lessmann, S., Baesens, B., Seow, H.V., Thomas, L.C.: Benchmarking state-of-the-art classification algorithms for credit scoring: an update of research. Eur. J. Oper. Res. **247**, 124–136 (2015). https://doi.org/10.1016/j.ejor.2015.05.030
14. Brown, I., Mues, C.: An experimental comparison of classification algorithms for imbalanced credit scoring data sets. Expert Syst. Appl. **39**, 3446–3453 (2012). https://doi.org/10.1016/j.eswa.2011.09.033

15. Li, Y., Chen, W.: A comparative performance assessment of ensemble learning for credit scoring. Mathematics **8**, 1–19 (2020). https://doi.org/10.3390/math8101756
16. Grennepois, N., Alvirescu, M.A., Bombail, M.: Using random forest for credit risk models machine (2019)
17. EC: On artificial intelligence - a European approach to excellence and trust (2020)
18. EBA: Big data and advanced analytics EBA report on big data and advanced analytics (2020)
19. Dupont, L., Fliche, O., Yang, S.: Governance of artificial intelligence in finance discussion document (APCR, Discussion Document) (2020)
20. Belle, V., Papantonis, I.: Principles and practice of explainable machine learning. Front. Big Data **4**, 1–25 (2021). https://doi.org/10.3389/fdata.2021.688969
21. Lundberg, S.M., Lee, S.-I.: A unified approach to interpreting model predictions. In: 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA (2017). https://doi.org/10.1016/j.ophtha.2018.11.016
22. Štrumbelj, E., Kononenko, I.: An efficient explanation of individual classifications using game theory. J. Mach. Learn. Res. **11**, 1–18 (2010)
23. Shapley, L.S.: A value for n-person games. Contribution to theory games, pp. 307–317 (1953)
24. Giudici, P., Raffinetti, E.: Shapley-Lorenz eXplainable artificial intelligence. Expert Syst. Appl. **167**, 114104 (2021). https://doi.org/10.1016/j.eswa.2020.114104
25. Joseph, A.: Parametric inference with universal function approximators (2020)
26. Bracke, P., Datta, A., Jung, C., Sen, S.: Machine learning explainability in finance: an application to default risk analysis. SSRN Electron. J. (2019). https://doi.org/10.2139/ssrn.3435104
27. Bücker, M., Szepannek, G., Gosiewska, A., Biecek, P.: Transparency, auditability, and explainability of machine learning models in credit scoring. J. Oper. Res. Soc. **73**, 70–90 (2022). https://doi.org/10.1080/01605682.2021.1922098
28. Saritaş, H., Kar, A., Pazarci, Ş.: Türkiye'de Doğrudan Yabancı Yatırımlar ile CDS, VIX Endeksi ve Kredi Derecelendirmeleri İlişkisi. Yönetim ve Ekon. Derg. **30**, 21–39 (2023). https://doi.org/10.18657/yonveek.1180755
29. Automotive Distributors' and Mobility Association: Passenger Car and Light Commercial Vehicle Market Evaluation. https://www.odmd.org.tr/web_2837_2/neuralnetwork.aspx?type=35. Accessed 26 July 2023
30. Chen, Y., Zhang, R.: Default prediction of automobile credit based on support vector machine. J. Inf. Process. Syst. **17**, 75–88 (2021). https://doi.org/10.3745/JIPS.04.0207
31. Lim, H.-E., Yeok, S.G.: Estimating the determinants of vehicle loan default in Malaysia: an exploratory study. Int. J. Manag. Stud. **24**, 73–90 (2017). https://doi.org/10.32890/ijms.24.1.2017.10477
32. Kursa, B.M., Rudnicki, W.R.: Package 'Boruta'. J. Stat. Softw. (2022)
33. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer, New York (2008). https://doi.org/10.1007/978-0-387-84858-7
34. Levy, J.J., O'Malley, A.J.: Don't dismiss logistic regression: the case for sensible extraction of interactions in the era of machine learning. BMC Med. Res. Methodol. **20**, 1–15 (2020). https://doi.org/10.1186/s12874-020-01046-3
35. Nusinovici, S., et al.: Logistic regression was as good as machine learning for predicting major chronic diseases. J. Clin. Epidemiol. **122**, 56–69 (2020). https://doi.org/10.1016/j.jclinepi.2020.03.002
36. Christodoulou, E., Ma, J., Collins, G.S., Steyerberg, E.W., Verbakel, J.Y., Van Calster, B.: A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J. Clin. Epidemiol. **110**, 12–22 (2019). https://doi.org/10.1016/j.jclinepi.2019.02.004
37. Haykin, S.: Neural networks and learning machines. (2008). https://doi.org/978-0131471399

38. Zhong, H., et al.: The application of machine learning algorithms in predicting the length of stay following femoral neck fracture. Int. J. Med. Inform. **155**, 1–7 (2021). https://doi.org/10.1016/j.ijmedinf.2021.104572
39. Fritsch, S., Guenther, F., Wright, M.N., Suling, M., Mueller, S.M.: Package "neuralnet": training of neural networks. The R J. **2**, 30–38 (2022)
40. Breiman, L.: Random forests. Mach. Learn. **45**, 5–32 (2001)
41. Breiman, L.: Bagging predictors. Mach. Learn. **24**, 123–140 (1996). https://doi.org/10.1023/A:1018054314350
42. Ho, T.K.: The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. **20**, 832–844 (1998). https://doi.org/10.1109/34.709601
43. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery Data Mining, 13–17 August 2016, pp. 785–794 (2016). https://doi.org/10.1145/2939672.2939785
44. Xi, Q., et al.: Individualized embryo selection strategy developed by stacking machine learning model for better in vitro fertilization outcomes: an application study. Reprod. Biol. Endocrinol. **19**, 1–10 (2021). https://doi.org/10.1186/s12958-021-00734-z
45. Quan Tran, V., Quoc Dang, V., Si Ho, L.: Evaluating compressive strength of concrete made with recycled concrete aggregates using machine learning approach. Constr. Build. Mater. **323**, 126578 (2022). https://doi.org/10.1016/j.conbuildmat.2022.126578
46. Chen, T., et al.: Package 'xgboost' extreme gradient boost. R Package version 1.7.5.1 (2023). https://doi.org/10.1145/2939672.2939785
47. Liu, Y., Just, A., Mayer, M.: Package 'SHAPforxgboost' SHAP plots for "XGBoost." R Package version V 0.1.3 (2023). https://doi.org/10.5281/zenodo.3568449
48. Louzada, F., Ara, A., Fernandes, G.B.: Classification methods applied to credit scoring: systematic review and overall comparison. Surv. Oper. Res. Manag. Sci. **21**, 117–134 (2016). https://doi.org/10.1016/j.sorms.2016.10.001

Exploring Lightweight Blockchain Solutions for Internet of Things: Review

Omar Ayad Ismael[1](✉), Mohammed Majid Abdulrazzaq[2], Nehad T. A. Ramaha[3], Yasir Adil Mukhlif[4], and Mustafa Ali Sahib Al Zakitat[5]

[1] Department of Computer, Diyala University, Baqubah, Diyala, Iraq
omarismail@uodiyala.edu.iq

[2] Department of Computer Engineering, Anbar University, Ramadi, Anbar, Iraq
moh.majid@uoanbar.edu.iq

[3] Department of Computer Engineering, Karabük University, Demir Celik Campus, 78050 Karabük, Turkey
nehadramaha@karabuk.edu.tr

[4] Department of Computer Engineering, Karabük University, Karabük, Turkey
2038166018@ogrenci.karabuk.edu.tr

[5] Electrical and Computer Engineering, Altinbas University, Istanbul, Turkey

Abstract. The world is witnessing a major digital transformation and is moving towards more interaction, connectivity, ease, and intelligence through the Internet of Things (IoT). The IoT offers these advantages to the world by linking necessary devices with each other, making it easier to manage and deal with those devices. However, the IoT faces many challenges, such as authentication, privacy, security, and access management. The application of blockchain technology may provide a solution to these challenges. Nevertheless, applying blockchain technology may face limitations, such as the limited resources of the IoT devices used and the resource-intensive requirements of the blockchain. Therefore, to overcome these limitations, several studies have proposed using a lightweight blockchain; this blockchain is specifically designed for resource-limited IoT devices. In this paper, a comprehensive review has been made on the uses of lightweight blockchain in the IoT. Moreover, we identified some of the challenges facing the application of blockchain technologies in the IoT and the future directions.

Keywords: blockchain · IoT · proof of work · privacy · security

1 Introduction

IoT is constantly expanding in all joints of life and works to transform the world into an intelligent world capable of management, control, direction, and decision-making. IoT allows the entire environment to be connected to the Internet, providing a smart life. Reports indicate that Internet of Things devices are growing continuously, reaching more than 26 billion devices in 2020 compared to 2009, as it was noted that the number of devices has increased by 30 times, as it was only 7 billion devices, including smartphones, computers, and other devices. These devices depend mainly on central process servers

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 437–455, 2024.
https://doi.org/10.1007/978-3-031-56728-5_37

and connect to the computer servers via the Internet, with a decrease in the security nature of work. Now it may be good, but in the future, with the size of the growth and with the importance of hardware, it may be a problem [1]. With the growth of the IoT and its importance, connected devices that are designed to transmit essential and sensitive information have arisen problems related to privacy and security, as most research has proven that the IoT lacks privacy and security and that the safety methods used cannot be applied on the Internet due to the nature of the devices are cheap with weak computing power. Some companies may be providers of services that use this essential and sensitive information illegally, with the aim of providing protection and privacy for the Internet of Things [2].

This research focuses on appropriate solutions for the future of the IoT by reviewing research concerned with the IoT. The blockchain enables the communication of all devices with their parts, peer-to-peer. Also, it provides the ability to track all devices, coordinate work with each other, and diagnose errors and risks. Moreover, one of the most important things provided by the blockchain is decentralization, where no third party can obtain important information, encryption, access management, and prevention of tampering by recording all operations that occur on devices. Also, the blockchain can define access and usage policy [3] "Fig. 1". Stages of development of the Internet of Things.

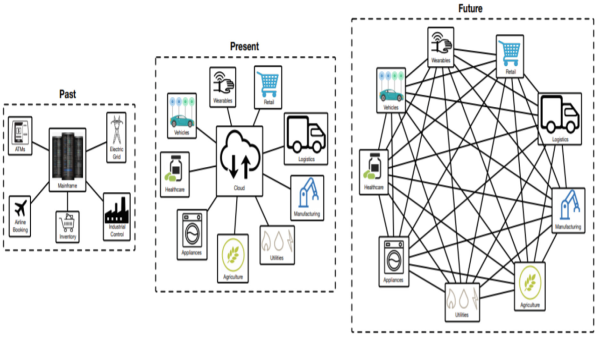


Fig. 1. Stages of development of the IoT [4].

This paper is divided as follows: Following the introduction, the second section presents the previous literature on IoT in general through the complete definition, problems, solutions, and applications of the IoT. The paper's third section will be about blockchain and the difference between lightweight blockchain and blockchain, and most IoT applications in which blockchain is involved. The fourth section concludes the paper, which consists of an extract from the steps of the work and the essential details and works in the future, in which it provides the most critical advice and recommendations for future researchers.

2 Literature Review

Lin, I. C., & Liao, T. C. [5] states that blockchain is more than just a technology of exchange, transfer and transmission. It contains an integrated system in the work of basic algorithms and important economic models. What distinguishes it is decentralization in dealing, continuous updating, transparency and complete confidentiality in the exchange of important information and transfers. It is considered a modern era, but it is not. It is devoid of challenges, the most important of which is legal, as this technology needs a law supported by legislated by governments, and it must be embraced by the establishment of this matter. People must take into account and be careful in dealing with it. Some research focuses on making use of blockchain in secure areas, including in general benefit from. The authors summarized [6] cases that are used in the IoT. The most important uses can be mentioned. Fixing events in fixed records that cannot be modified, and accessing data through decentralized distribution, key management, or symmetric and asymmetric encryption in [7]. The authors focused on the most important challenges facing the IoT, most notably the proof of identity, ownership and access to data through authentication, privacy and security, and also mentioned how the blockchain can solve these problems [8]. The authors proposed an integrated framework for the IoT in the industry based on blockchain technology in addition to communicate with the cloud is a structure where sending to cloud storage is to be analyzed and at the same time sending to blockchain-based devices. And in smart contracts The authors see [9] and how smart contracts facilitate work and communication between devices and share data between them. It is by billing and electronic payment Authors [10] take advantage of blockchain in this function as well as commercial shipping, as it facilitates the buying and selling of energy (Table 1).

Table 1. Comparison of literature review.

Author	Year	Publishing search	Main idea	Primary application domain	Specific security service
Osama A. Khashan, Nour M. lightness [47]	2023	Journal of King Saud University - Computer and Information Sciences	The author proposes a centralized and blockchain-based hybrid authentication architecture for IoT systems. The architecture consists of two layers: the core layer and the blockchain layer. The central layer is responsible for providing authentication services to devices. The blockchain layer is responsible for providing decentralized authentication and verification services. The proposed system is based on lightweight encryption methods implemented in the architecture on the Elliptic Curve Digital Signature Algorithm (ECDSA), which is well suited for IoT devices	smart homes	decentralized authentication
Ramamoorthi S., Muthu Kumar B., Ahilan Appathurai [46]	2023	Computer Communications Supports open access	The author proposes a new energy-saving and secure communication system for a live IoT. The system uses a combination of five chaotic map-based Elgamal authentication, signature-based Enroute filtering, credit-point-based aggregation, and Capuchin search optimization-based packet routing. The paper implemented the system using an NS3 network simulator and evaluated his performance. The results showed the author the superiority of the system over the existing systems in all scales that were evaluated as found in the results mentioned	smart homes	energy-saving and secure communication

(*continued*)

Table 1. *(continued)*

Author	Year	Publishing search	Main idea	Primary application domain	Specific security service
Xiaolong Xu Yu Jiang [43]	2022	IEEE	The author proposes a lightweight blockchain called LBlockchainE is designed for marine mobility systems of ships and ships that contain IoT technology to ensure the security of sensor data in their parts to save the resources of the terminal servers on the parts of the ship The new system has proven to use 1.6% less time and consume 78% less battery power compared to It is compatible with traditional blockchain systems and ensures that data is decentralized and protected from external danger	marine mobility systems of ships	Protection and real time realization
B. D. Deebak [42]	2022	IEEE	In this article, the author proposes a blockchain-based remote mutual authentication (B-RMA) that takes into account AI-enabled devices and cloud networks to enhance security and privacy. The proposed B-RMA could be activated in IoT-based smart homes and cities and make user authentication requests decentralized	smart devices and cloud networks	authentication
Meryem Ammi, Shatha Alarabi, Elhadj Benkhelifa	2021	Information Processing & Management	The author proposes a new system to protect smart homes and make them more secure. This system is based on Blockchain using an integrated hyperledger fabric and hyperledger composer. The proposed system consists of four cloud storage layers, a Hyperledger fabric, a Hyperledger composer and a smart home layer. The most important aspect of the proposed system is mapping the smart home attributes to those own Composer: hyperledger. This designation allows for a purpose-built solution that can provide security for smart homes	smart homes	security

(continued)

Table 1. (*continued*)

Author	Year	Publishing search	Main idea	Primary application domain	Specific security service
Sachi Nandan Mohanty, K.C. Ramya, S. Sheeba Rani	2020	Future Generation Computer Systems	The author proposes A lightweight integrated blockchain model was developed to meet the requirements of the Internet of Things. This model is implemented in a smart home IoT environment. The model is based on three optimizations, including a lightweight consensus algorithm, certificateless y (CC) and a distributed throughput management (DTM) system. The proposed system achieves an overall 50% saving in processing time compared to the baseline method with the lowest energy consumption of 0.07 mJ	smart home	Real time and energy saving
Shafieinejad, A., & Almasian	2022	Computer Standards & Interfaces	A study presenting a model for secure file sharing in the cloud using blockchain technologies and attribute-based description cryptography. The model uses smart contracts to control access and supports the confidentiality of the authoritative information and the credibility of the participants. The key used is contained in the parameters of a mathematical function called access parameters, and users can retrieve the key to decrypt files. The model is scalable and has shown acceptable performance with 20,000 users	smart contracts	secure file sharing in the cloud
Qi, L., Tian, J., Chai, M., & Cai, H. [49]	2023	Computer Networks	The research presents a solution to design a lightweight Proof-of-Work (PoW) mechanism for the Internet of Things (IoT) blockchain, while maintaining a high level of security. This is done by introducing the Heterogeneity Considered Trust Mechanism (HCTM) for the trust contract between IoT nodes, this proposed model has lower power consumption and higher security	IoT NODE	reduce the energy & trust contract

(*continued*)

Table 1. (*continued*)

Author	Year	Publishing search	Main idea	Primary application domain	Specific security service
Tomar, A., Gupta, N., Rani, [50]	2023	Internet of Things	This study aims to provide a new protocol called BIoMTAKE to secure data of Internet devices of important things in the medical field and hospitals. This protocol uses blockchain technology and Hyperledger Fabric to create a distributed and secure environment where communication and data exchange between authenticated devices is secured. In terms of safety and performance, the results showed the strength and efficiency of the proposed protocol	healthcare	prevent unauthorized access
Cunha, J., Duarte, R., Guimarães, T	2022	Procedia Computer Science	This study aims to develop a system based on blockchain technology and the common standard for electronic records in healthcare. The goal is to ensure integrated operation, rapid access, reliability, and security of critical health data. The study revolves around proposing two concepts of architecture in the intensive care unit, and then comparing them and determining the most beneficial scenario for the hospital	healthcare	reliability, and security of critical health data
Friese, J Heuer, N Kong [7]	2014	IEEE	It presents a set of solutions and recommendations that can be used in the future for identity management	Identities of Things	solutions to manage “Identities of Things”
Bahga, A., & Madisetti, V. K [8]	2016	Journal of Software Engineering and Applications	Technically, he did not mention the future addition, but the author advises the application of the Blockchain in the Internet industry to benefit from the benefits of the Blockchain and achieve the most gains	Industrial Internet of Things	decentralized, peer-to-peer platform
K Christidis, M Devetsikiotis [9]	2016	IEEE	The author believes that the blockchain’s connection to the IoT can cause major transformations in the industry	IoT	To improve interaction and verification between connected devices

2.1 The Internet of Thing

The IoT, which is considered the Internet of everything in the present time, in other words, entering the Internet in everything, it expresses the new and modern paradigm, the transformation of things into the world of the net, and the linking of important things in a person's life to the Internet and making them as a network that understands with each other, for example, you do not need to inspect your home. You can know. What is in it through the mobile, and the IoT enters the fields of health, industrial, commercial, transportation, smart homes, etc. [11]. The basis of the work of the IoT is done through the computers that in turn collect data through the sensors, the data is transferred through special platforms to the cloud storage, and this data is analyzed and the problems identified are addressed, so it is via a network of devices and sensors connected to each other via the Internet These devices allow the exchange of information. The Internet allows communication with these devices and users. The infrastructure of the IoT is IoT = Services + Data + Networks + Sensors [12]. As shown in Fig. 2 And no system is without problems, and among the most prominent problems of the IoT are the attacks that happen to the IoT and devices that are widely spread, data leakage and the inability to track data due to the large number of devices.

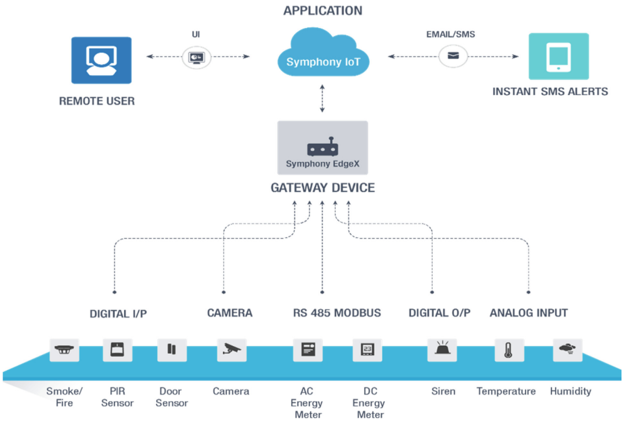


Fig. 2. IoT architecture.

2.2 IoT Safety

The more important things in our world, the greater the risks. This is why must care about the safety of important things in our lives, and one of the most important things that require high protection in our present world is the IoT because of its high importance and weak security used between devices during transmission and reception operations and because the cheap devices used are not Its design is to support high protection and the threat here does not depend on software only Because it has important for the privacy of people, governments and organizations, as the IoT is involved in most things around people and in the simplest details, which may threaten their safety and privacy in the event of lack of concern for security in the IoT. The threat in the IoT is transmitted to the material components and the control of the devices and the taking of important information about the users. This is why must pay attention to the safety of the IoT [13].

2.3 Blockchain

The true definition of blockchain as a distributed ledger technology and blockchain supports a secure platform and database that maintains all users and gives the blockchain high encryption of the transferred data as well as gives the blockchain a decentralized transaction in the exchange of information based on the peer-to-peer principle. The blockchain is distinguished in these characteristics that make it. Suitable in the safety of the IoT, but after the reduction process [14]. The blockchain consists in a short form of two main components, which are the transactions that contain the actions taken by the users of the technology and the blocks that contain the correct records that are related to each other and distributed [15] Fig. 3 Block components.

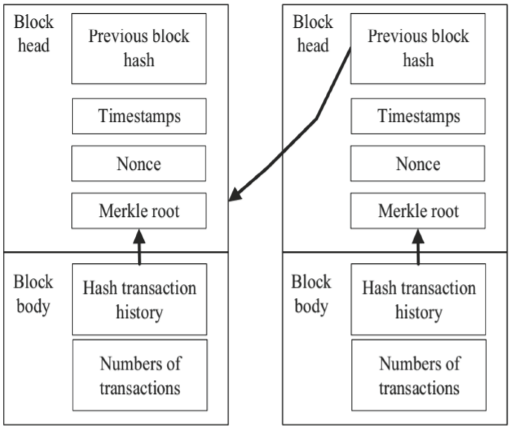


Fig. 3. Blockchain components [15].

2.4 The Benefits that Blockchain Provides

There are many benefits that blockchain provides to the IoT, which is the ideal solution to many IoT problems [16].

1) Distributed ledgers are decentralized: It is difficult to change it because it is widespread and distributed and changing one of them is exposed from the other and therefore it is very difficult to change it. A lot of work and budgets that are allocated to materials are reduced. This enhances confidence among users, and this feature will support the IoT because it provides high security for devices and eliminates failure points [17, 23]

2) Security: it provides the transfer of data between nodes and platforms in a safe and unbreakable manner. It is difficult to spy on it after adding another level of protection, which is encryption that does.
3) The record is not subject to change: This means that the data that is saved cannot be changed at all, and thus the stored data becomes in a high safety and is called stability in the information. Privacy and security [18]. Iot allow the hacker to access the blockchain network. Supports a secure platform for IoT devices [2].
4) Transparency: At every stage on the networks and everything in the IoT will be under the cover of the blockchain, where all operations are recorded and it is also possible to know where the data has been leaked. This feature gives all victims the ability to see what is going on and see all the processes that are taking place. They each have a shared private notebook containing all the processes that are taking place. This feature gives internet users the ability to view all devices [19].
5) Blockchain reduces the cost: You get rid of the third party, which may be a protection company or a maintenance company associated with the central entity and servers through the blockchain, it will only direct you to improve work [9].
6) Preserving the identity: All members of the network use unique addresses and numbers that preserve their identity and not reveal it, despite objecting to this feature in illegal commercial dealings, but it can be used in the IoT through its application in elections. Voting is completely secret [20–22] (Fig. 4).

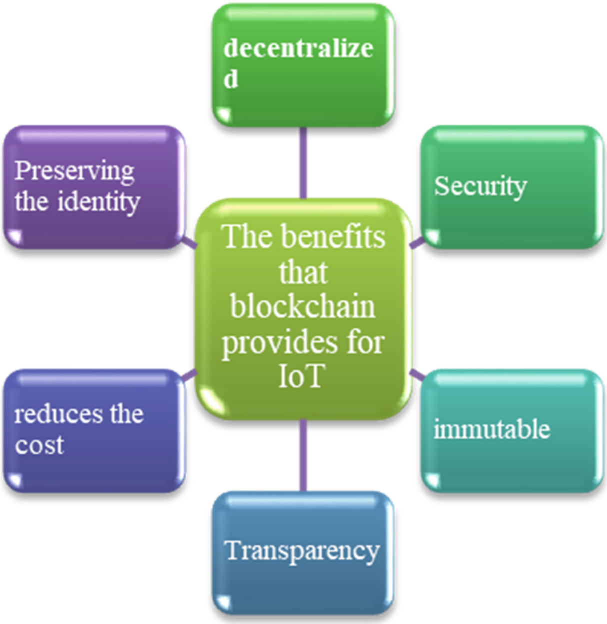


Fig. 4. The benefits that blockchain provides

3 Discussion

By searching for keywords related to the lightweight Blockchain in the Internet of Things, it becomes clear that the best solution to the problem of security and privacy in the Internet of Things is to use the Blockchain to protect the Internet of Things, despite the novelty of the concept of Blockchain, which was introduced 15 years ago as a technology behind the Bitcoin currency. The matter has expanded more after the emergence of the concept of Ethereum, which supports smart contracts, whose use has expanded beyond digital currencies. These security solutions do not change the structure of the network. They only use the features and characteristics of the blockchain, and the most prominent of these features are security, privacy, decentralization, and trust. Through research, we found that smart contracts Blockchain features are also a key factor in the process of improving the security of the Internet of Things and making it more reliable.

IoT security has not been treated to a great extent, and blockchain technology can be used in the IoT, but blockchain implementation in the IoT faces great difficulty because it requires great computing power, including the used keys, Merkle Hash Tree, and Proof of Work (PoW), but it can be effectively applied. The idea of blockchain, but something simpler and lighter, can be described as lightweight to take advantage of the many advantages provided by the blockchain in the IoT. Blockchain can be used to connect a group of devices securely and for a long time [24]. In Table 2 show the most important differences between the traditional and blockchain-based IoT.

Table 2. Comparison between the IoT and blockchain

Features	Blockchain	IoT
Administration	Decentralized	Centralized
Resources	consuming	restricted
Time	It takes a long time to mine	Low execution time
devices used	Less equipment used, more power	The number of devices is very large with low power
Energy consumption	High energy consumption	IoT devices have limited bandwidth and resources
Safety	strong safety	Security is one of the big challenges of IoT
Confidence	Trust	Trust is one of the big challenges IoT
the cost	high cost	low cost
Internet	need big internet packages	Need low internet packages
Legally	Not legally supported	Supported

3.1 Challenges of Implementing Blockchain in the IoT

The performance of devices is of limited source because the devices in nature are designed to be suitable for less computing power, and the blockchain requires great computing power through mathematical operations. It will not be able to run cryptographic algorithms [17]. On the other hand, realizing the real-time response and reception may be a problem in the case of blockchain implementation in the IoT. The great expansion of the IoT may cause the future [25]. The data redundancy model adequate and the storage capacity of the IoT is few and does not match the distributed ledgers. High encryption in the blockchain. Most of the challenges relate to the state of the IoT architecture. Storage in the blockchain is decentralized, unlike the nature of the IoT, which relies on cloud storage, with the lack of the ability of devices to store details, and the blockchain depends on storage in the nodes through the distributed ledger. For users, this technology is still new and few people have the ability to deal with it and it requires full awareness [15]. There are also legal problems because blockchain technology is not subject to government oversight and can be managed from any country. IoT devices can be managed from anywhere without monitoring. It may be a violation in some countries and there are also problems with naming and discovery [26] (Fig. 5).

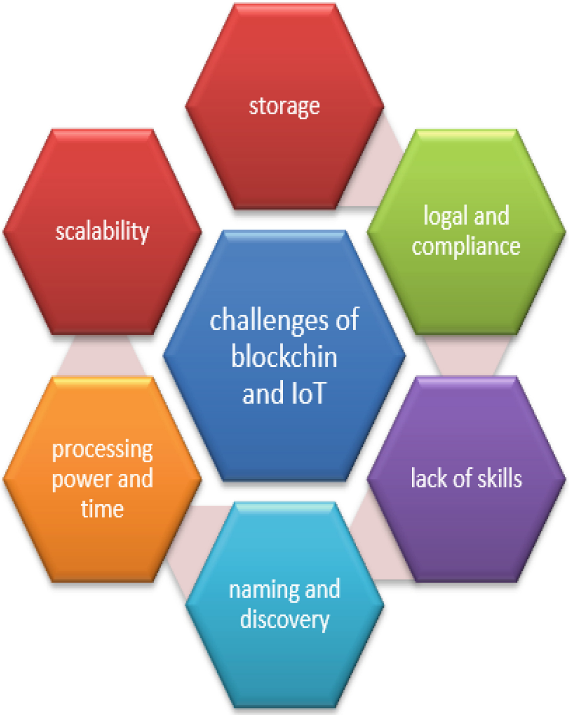


Fig. 5. BlockChain Implementation Challenges in IoT.

3.2 Blockchain Applications in the IoT

- Health care in the study The lightweight blockchain was applied in the medical field of the IoT, with the aim of protection in the process of transmitting necessary medical information, and it has been proven to provide strong security advantages and also in another study that was applied in medicines through the use of blockchain, where the public can access Records for the dates and safety of medicines from heat and humidity [27, 28]
- In agriculture, blockchain has been applied in the agricultural field of the IoT through
- the application of a system that tracks agricultural supplies in China.
- Monitoring public devices: Blockchain is applied in monitoring devices IoT. In this proposal, Ethereum technology has been applied in controlling and managing devices remotely. Public keys are stored in Ethereum and the private key for devices is one of the features of Ethereum allowing the necessary code to be written and this facilitates the maintenance process (34).
- In the energy sector: In the energy sector, the blockchain also contributes effectively to this proposal that has been implemented. Blockchain allows devices to pay each other, and this reduces the human need. It pays electricity wages by relying on Bitcoin [30].
- The energy sector: In the energy sector, the blockchain also contributes effectively to this proposal that has been implemented. Blockchain allows devices to pay each other, and this reduces human need. It pays electricity wages by relying on Bitcoin [30].
- Smart cities: Smart cities have great importance in our daily life and their information is very important, as well as controlling devices that may work completely disrupting this. Blockchain has been proposed.
- secure smart cities and their information by integrating blockchain with special devices to create a secure platform [31].

There are many applications that blockchain enters into with the IoT in all aspects of life. It provides us with an integrated service that supports security, privacy and distribution. In Table 3 a summary of the most important applications.

Table 3. Blockchain applications in the IoT.

The author	the year	platform	BIoT applications	Summarization
M Humayun, NZ Jhanjhi, B Hamid, G Ahmed	2020	Custom	Transportation [33]	The author presented a proposal to integrate the IoT with blockchain in order to provide safe, intelligent and lower costs
T Moura, A Gomes	2017	Custom	Public elections [38]	To support democracy and ensure that election theft and manipulation are counted through blockchain technology that ensures data integrity and stability, high transparency and enhances trust
Lundqvist, T., de Blanche, A	2017	proof-of-concept	Energy [30]	The bill payment process is without the need for a direct third party using blockchain technology and helps in energy consumption
C Lazaroiu, M Roscia -	2017	Ethereum	Smart city [32]	IoT application with blockchain in order to provide a smart zone that preserves the privacy of residents and reduces energy consumption
M Humayun, NZ Jhanjhi, B Hamid, G Ahmed	2020	Custom	Transportation [33]	The author presented a proposal to integrate the IoT with blockchain in order to provide safe, intelligent and lower costs

(*continued*)

Table 3. (*continued*)

The author	the year	platform	BIoT applications	Summarization
N Rožman, M Corn, T Požrl, J Diaci - Procedia CIRP	2019	Distributed logistics	logistics platform [34]	It introduces a new model for modern sourcing by splitting partners into a contract system that allows partners to present their contracts and acts as a bridge between the real and the virtual
D Han, H Kim, J Jang	2017	Ethereum	Smart living [35]	In this study, the author proposes a system to maintain the safety of homes by proposing a model of smart doors that open and close depending on the data sent by relying on the blockchain
M Siddiqi, ST All, V Sivaraman	2017	Hyperledger Fabric	Personal Sensing [36]	The authors propose a model for recording people's data through human-carried devices that help in storing necessary medical data. Information is secured through blockchain technology
T Ahram, A Sargolzaei, S Sargolzaei	2017	Custom	Industry [19]	Development of health industries using blockchain, which is also proof that blockchain serves the industrial sector

(*continued*)

Table 3. (*continued*)

The author	the year	platform	BIoT applications	Summarization
N Kshetri	2017	Custom	Defense & Public Safety [19]	A model for storing data in the cloud and blockchain security effects on the IoT
T Bocek, BB Rodrigues, T Strasser, B Stiller	2017	Ethereum	Healthcare [28]	In this proposal, medication records can be accessed, and storage medication reviewed for temperature and humidity for suitability
Tian, F	2016	Custom	Farming [22]	blockchain has been applied in the agricultural field of the IoT through the application of a system that tracks agricultural supplies in China

4 Conclusion

The importance of the IoT is increasing day by day in the world with the development of the Internet. Also, this development is not a priority for the security and privacy of the IoT. The IoT devices deal with very important information in homes, laboratories, hospitals, universities, and even e-governments. Threats in the IoT are not limited to software things only, but rather have become encroaching on physical things, and this means that some threats may reach homes, factories, and hospitals by judging important devices, and this negatively affects privacy and security in this aspect. These devices are not able to defend themselves against these attacks, so you should pay attention to that.

A review of the literature was conducted to review the most important benefits provided by the uses of the blockchain in order to solve the problems of the IoT, especially the problems related to the security of devices, the retention of information and not to tamper with it, and the treatment of points of failure Through previous literature, the blockchain is considered as what the IoT needs to solve these problems. Problems and future work of the IoT were also discussed, where it was noted that it is necessary to maintain privacy and security, and these features are provided by the blockchain.

References

1. Cisco Visual Networking Index: Global mobile data traffic forecast update, 2016–2021 white paper, vol. 7, p. 180. Cisco, San Jose (2017)

2. Dorri, A., Kanhere, S.S., Jurdak, R.: Blockchain in Internet of Things: challenges and solutions. arXiv preprint arXiv:1608.05187 (2016)
3. Yelowitz, A., Wilson, M.: Characteristics of Bitcoin users: an analysis of Google search data. Appl. Econ. Lett. **22**(13), 1030–1036 (2015)
4. Fernández-Caramés, T.M., Fraga-Lamas, P.: A review on the use of blockchain for the Internet of Things. IEEE Access **6**, 32979–33001 (2018)
5. Lin, I.C., Liao, T.C.: A survey of blockchain security issues and challenges. IJ Netw. Secur. **19**(5), 653–659 (2017)
6. Conoscenti, M., Vetro, A., De Martin, J.C.: Blockchain for the Internet of Things: a systematic literature review. In: 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA), pp. 1–6. IEEE (2016)
7. Friese, I., Heuer, J., Kong, N.: Challenges from the Identities of Things: introduction of the Identities of Things discussion group within Kantara initiative. In: 2014 IEEE World Forum on Internet of Things (WF-IoT), pp. 1–4. IEEE (2014)
8. Bahga, A., Madisetti, V.K.: Blockchain platform for industrial Internet of Things. J. Softw. Eng. Appl.Softw. Eng. Appl. **9**(10), 533–546 (2016)
9. Christidis, K., Devetsikiotis, M.: Blockchains and smart contracts for the Internet of Things. IEEE Access **4**, 2292–2303 (2016). https://doi.org/10.1109/ACCESS.2016.2566339
10. Brody, P., Pureswaran, V.: Device democracy: saving the future of the Internet of Things. IBM **1**(1), 15 (2014)
11. Lee, I., Lee, K.: The Internet of Things (IoT): applications, investments, and challenges for enterprises. Bus. Horiz.Horiz. **58**(4), 431–440 (2015)
12. Wang, Q., Zhu, X., Ni, Y., Gu, L., Zhu, H.: Blockchain for the IoT and industrial IoT: a review. Internet Things **10**, 100081 (2020)
13. Alladi, T., Chamola, V., Sikdar, B., Choo, K.K.R.: Consumer IoT: security vulnerability case studies and solutions. IEEE Consum. Electron. Mag. **9**(2), 17–25 (2020)
14. Stanciu, A.: Blockchain based distributed control system for edge computing. In: 2017 21st International Conference on Control Systems and Computer Science (CSCS), pp. 667–671. IEEE (2017)
15. Banafa, A.: IoT and blockchain convergence: benefits and challenges. IEEE Internet Things (2017)
16. Bandara, E., Tosh, D., Foytik, P., Shetty, S., Ranasinghe, N., De Zoysa, K.: Tikiri-towards a lightweight blockchain for IoT. Future Gener. Comput. Syst. **119**, 154–165 (2021)
17. Samaniego, M., Deters, R.: Zero-trust hierarchical management in IoT. In: 2018 IEEE International Congress on Internet of Things (ICIOT), pp. 88–95. IEEE (2018)
18. Atlam, H.F., Alenezi, A., Walters, R.J., Wills, G.B., Daniel, J.: Developing an adaptive risk-based access control model for the Internet of Things. In: 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), pp. 655–661. IEEE (2017)
19. Ahram, T., Sargolzaei, A., Sargolzaei, S., Daniels, J., Amaba, B.: Blockchain technology innovations. In: 2017 IEEE Technology & Engineering Management Conference (TEMSCON), pp. 137–141. IEEE (2017)
20. Torkaman, A., Seyyedi, M.A.: Analyzing IoT reference architecture models. Int. J. Comput. Sci. Softw. Eng. **5**(8), 154 (2016)
21. Atlam, H.F., Alenezi, A., Alharthi, A., Walters, R.J., Wills, G.B.: Integration of cloud computing with Internet of Things: challenges and open issues. In: 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), pp. 670–675. IEEE (2017)

22. Tian, F.: An agri-food supply chain traceability system for China based on RFID & blockchain technology. In: 2016 13th International Conference on Service Systems and Service Management (ICSSSM), pp. 1–6. IEEE (2016)
23. Atlam, H.F., Alenezi, A., Alassafi, M.O., Wills, G.: Blockchain with Internet of Things: benefits, challenges, and future directions. Int. J. Intell. Syst. Appl. **10**(6), 40–48 (2018)
24. Liu, Y., Wang, K., Lin, Y., Xu, W.: LightChain: a lightweight blockchain system for industrial Internet of Things. IEEE Trans. Ind. Inf. **15**(6), 3571–3581 (2019)
25. Atlam, H.F., Alenezi, A., Hussein, R.K.H., Wills, G.: Validation of an adaptive risk-based access control model for the Internet of Things. Int. J. Comput. Netw. Inf. Secur. **10**(1), 26–35 (2018)
26. Daza, V., Di Pietro, R., Klimek, I., Signorini, M.: CONNECT: CONtextual NamE disCovery for blockchain-based services in the IoT. In: 2017 IEEE International Conference on Communications (ICC), pp. 1–6. IEEE (2017)
27. Seliem, M., Elgazzar, K.: BIoMT: Blockchain for the internet of medical things. In: 2019 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), pp. 1–4. IEEE (2019)
28. Bocek, T., Rodrigues, B.B., Strasser, T., Stiller, B.: Blockchains everywhere-a use-case of blockchains in the pharma supply-chain. In: 2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM), pp. 772–777. IEEE (2017)
29. Huh, S., Cho, S., Kim, S.: Managing IoT devices using blockchain platform. In: 2017 19th International Conference on Advanced Communication Technology (ICACT), pp. 464–467. IEEE (2017)
30. Lundqvist, T., de Blanche, A., Andersson, H.R.H.: Thing-to-thing electricity micro payments using blockchain technology. In: 2017 Global Internet of Things Summit (GIoTS), pp. 1–6. IEEE (2017)
31. Biswas, K., Muthukkumarasamy, V.: Securing smart cities using blockchain technology. In: 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), pp. 1392–1393. IEEE (2016)
32. Lazaroiu, C., Roscia, M.: Smart district through IoT and blockchain. In: 2017 IEEE 6th International Conference on Renewable Energy Research and Applications (ICRERA), pp. 454–461. IEEE (2017)
33. Humayun, M., Jhanjhi, N.Z., Hamid, B., Ahmed, G.: Emerging smart logistics and transportation using IoT and blockchain. IEEE Internet Things Mag. **3**(2), 58–62 (2020)
34. Rožman, N., Corn, M., Požrl, T., Diaci, J.: Distributed logistics platform based on Blockchain and IoT. Procedia CIRP **81**, 826–831 (2019)
35. Han, D., Kim, H., Jang, J.: Blockchain based smart door lock system. In: 2017 International Conference on Information and Communication Technology Convergence (ICTC), pp. 1165–1167. IEEE (2017)
36. Siddiqi, M., All, S.T., Sivaraman, V.: Secure lightweight context-driven data logging for bodyworn sensing devices. In: 2017 5th International Symposium on Digital Forensic and Security (ISDFS), pp. 1–6. IEEE (2017)
37. Kshetri, N.: Blockchain's roles in strengthening cybersecurity and protecting privacy. Telecommun. Policy **41**(10), 1027–1038 (2017)
38. Moura, T., Gomes, A.: Blockchain voting and its effects on election transparency and voter confidence. In: Proceedings of the 18th Annual International Conference on Digital Government Research, pp. 574–575 (2017)
39. Ismael, O.A., Çelik, Ö., Yüksel, Ü.: A new approach to Arabic spam tweet detection in Twitter using machine learning algorithms. In: AIP Conference Proceedings, vol. 2398, no. 1. AIP Publishing (2022)

40. Shaker, A.S., Khaleel, M.F., Ismael, O.A., Majeed, R.S., Ahmed, M.R.: Information retrieval system of Arabic alphabetic characters by using hidden Markov Model. In: 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), pp. 1–6. IEEE (2022)
41. Alrifaie, M.F., Ismael, O.A., Hameed, A.S., Mahmood, M.B.: Pedestrian and objects detection by using learning complexity-aware cascades. In: 2021 2nd Information Technology to Enhance E-Learning and Other Application (IT-ELA), pp. 12–17. IEEE (2021)
42. Deebak, B.D., et al.: A lightweight blockchain-based remote mutual authentication for AI-empowered IoT sustainable computing systems. IEEE Internet Things J. **10**(8), 6652–6660 (2022)
43. Jiang, Y., Xu, X., Gao, H., Rajab, A.D., Xiao, F., Wang, X.: LBlockchainE: a lightweight blockchain for edge IoT-enabled maritime transportation systems. IEEE Trans. Intell. Transp. Syst.Intell. Transp. Syst. **24**(2), 2307–2321 (2022)
44. Mohanty, S.N., et al.: An efficient Lightweight integrated Blockchain (ELIB) model for IoT security and privacy. Future Gener. Comput. Syst. **102**, 1027–1037 (2020)
45. Ammi, M., Alarabi, S., Benkhelifa, E.: Customized blockchain-based architecture for secure smart home for lightweight IoT. Inf. Process. Manag.Manag. **58**(3), 102482 (2021)
46. Ramamoorthi, S., Appathurai, A.: Energy aware Clustered blockchain data for IoT: an end-to-end lightweight secure & Enroute filtering approach. Comput. Commun.. Commun. **202**, 166–182 (2023)
47. Khashan, O.A., Khafajah, N.M.: Efficient hybrid centralized and blockchain-based authentication architecture for heterogeneous IoT systems. J. King Saud Univ.-Comput. Inf. Sci. **35**(2), 726–739 (2023)
48. Shafieinejad, A., Almasian, M.: Secure cloud file sharing scheme using blockchain and attribute-based encryption. Available at SSRN 4252101 (2022)
49. Qi, L., Tian, J., Chai, M., Cai, H.: LightPoW: a trust based time-constrained PoW for blockchain in Internet of Things. Comput. Netw.. Netw. **220**, 109480 (2023)
50. Tomar, A., Gupta, N., Rani, D., Tripathi, S.: Blockchain-assisted authenticated key agreement scheme for IoT-based healthcare system. Internet Things **23**, 100849 (2023)
51. Cunha, J., Duarte, R., Guimarães, T., Santos, M.F.: Permissioned blockchain approach using open data in healthcare. Procedia Comput. Sci. **210**, 242–247 (2022)

Harnessing Advanced Techniques for Image Steganography: Sequential and Random Encoding with Deep Learning Detection

Mustafa Ali Sahib Al Zakitat[1], Mohammed Majid Abdulrazzaq[2](✉), Nehad T. A. Ramaha[3], Yasir Adil Mukhlif[4], and Omar ayad Ismael[5]

[1] Electrical and Computer Engineering, Altinbas University, Istanbul, Turkey
[2] Department of Computer Engineering, Anbar University, Anbar, Iraq
moh.majid@uoanbar.edu.iq
[3] Department of Computer Engineering, Karabük University, Demir Celik Campus, 78050 Karabük, Turkey
nehadramaha@karabuk.edu.tr
[4] Department of Computer Engineering, Karabük University, Karabük, Turkey
[5] Department of Computer, Diyala University, Diyala, Iraq
omarismail@uodiyala.edu.iq

Abstract. This study delves into the intricacies of steganography, a method employed for concealing information within a clandestine medium to enhance data security during transmission. Given that information is often represented in various forms, such as text, audio, video, or images, steganography offers a distinctive advantage over conventional cryptography by focusing on concealing the very existence of the message, rather than merely its content. This research introduces a novel steganographic technique that places equal emphasis on both message concealment and security enhancement. This study highlights two primary steganographic methods: sequential encoding and random encoding. By employing both encryption and image compression, these techniques fortify data security while preserving the visual integrity of cover images. Advanced deep learning models, namely Vgg-16 and Vgg-19, are proposed for the detection of image steganography, with their accuracy and loss rates rigorously evaluated. The significance of steganography extends across various sectors, including the military, government, and online domains, underscoring its pivotal role in contemporary data communication and security.

Keywords: Data mining · machine learning · artificial neural network · network security · cryptograph

1 Introduction

In the era of proliferating data transmission technologies, ensuring secure data transmission has become paramount. Steganography, a method for covertly embedding confidential information within a medium, has gained significant attention. This concealment technique applies to various data formats, including text, audio, video, and image

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 456–470, 2024.
https://doi.org/10.1007/978-3-031-56728-5_38

files. This research introduces an innovative steganographic method, aiming not only to conceal messages within images but also to enhance data security [1].

Steganography, deriving its name from the Greek words “steganos” (meaning “hidden”) and “graphos” (meaning “writing” or “drawing”), fundamentally conceals messages, in contrast to cryptography, which encodes them into a cipher to obscure their meaning. It exploits human perception, making the hidden data imperceptible. While computer programmers may conduct Steganalysis to uncover concealed information, it primarily challenges human perception [2]. Typically, concealed data is encrypted with a password when embedded within a carrier file. This section presents an overview of the problem statement, research objectives, and contextual background. Cryptography and steganography are widely employed to safeguard data privacy and integrity during transmission and storage. Steganography allows data to remain hidden within a file, thus avoiding suspicion. Cryptography, on the other hand, converts the original message into a transmittable format through encryption, which can be reversed into plaintext using a secret key. Symmetric-key and public-key systems are two common cryptographic approaches, with symmetric-key systems generally providing higher security [3].

These combined techniques facilitate the secure transmission of confidential communications across networks, safeguarding information accessibility. As data sharing and electronic information exchange become increasingly prevalent, data security has become paramount. Techniques like Steganography and cryptography are instrumental in concealing or encrypting sensitive data and information. While often used interchangeably, they vary in their ability to hide data effectively. Steganography leverages human perception, making concealed information nearly imperceptible, while encryption ensures data confidentiality. Steganography communications represent a new frontier in the digital age, enabling messages to be hidden within unused bits of digital data, such as audio, video, and image files.

This research focuses on enhancing communication security, particularly in the context of image steganography, utilizing advanced deep learning models such as Vgg-16 and Vgg-19, while evaluating their accuracy and loss rates. This study aims to advance communication security through deep learning models, specifically targeting image steganography, with a focus on Vgg-16 and Vgg-19 models and their performance metrics.

2 Research Background

Attackers can potentially bypass cryptographic safeguards by gaining access to systems responsible for data encryption and decryption or by exploiting vulnerabilities in cryptographic implementations, such as weak default keys. These scenarios are plausible and underscore the significance of encryption in preventing malicious actors from deciphering encrypted data or messages. While steganography and cryptography are distinct concepts, their synergy can significantly enhance content security. Although the carrier medium may no longer be concealed, an encoded message can still remain hidden, offering several advantages that cannot be achieved through encryption alone [4]. Steganography finds applications across various sectors, including the military, government, public sector, and the internet. Diverse encryption and decryption techniques have been developed to preserve the confidentiality of electronic communications. However, in certain situations, merely keeping a conversation private is insufficient; concealing the very existence of the message becomes paramount. Steganography, as its name implies, conceals information within a message, whereas cryptography aims to protect its content.

Steganography's objective is to conceal the communication medium itself. To enhance security, innovative methods have emerged that combine both cryptography and steganography, allowing for sequential encoding or random encoding of information within an image. These techniques remain imperceptible to computers, as they operate within a single image, offering benefits such as data compression, reduced memory usage, and the ability to use RGB pixels without distortion [5]. While both steganography and cryptography contribute to data protection, they are not immune to vulnerabilities. If the existence of hidden information becomes known or suspected, the purpose of steganography may be compromised [6]. Combining steganography with encryption enhances its effectiveness [7].

The safeguarding of intellectual property necessitates unique algorithms in steganography. The subsequent paragraphs delve deeper into the characteristics of an effective steganography algorithm. Watermarks play a vital role in consistently identifying an item, regardless of its location in a document. They are commonly employed to protect intellectual property by obscuring identifying marks, such as signatures [8]. One such method involves fingerprinting, where individual fingerprints are embedded in multiple copies of a carrier object, enabling the tracking of unauthorized distribution of licensed materials [9].

3 Image Steganography

In the realm of computing, an image is essentially a numerical representation that encapsulates the range of luminosity found within it. Each element in this representation is referred to as a pixel, and the collective arrangement of these pixels forms a grid, commonly known as a pixel grid. These pixels, or encoded bits, compose an image in a rectangular layout. Each pixel's position and color are individually marked, and the image's resolution is indicated by another marker. When displayed on a computer screen, these pixels are organized horizontally in rows. The bit depth of a color palette signifies the number of bits allocated per pixel. To minimize bit depth, contemporary color schemes can employ as few as 8 bits to specify the color or shade of each pixel. For monochrome or grayscale images, an eight-bit value is assigned to each pixel, allowing for a total of 256 possible colors or tonalities. In the case of digital color photographs, a 24-bit format is commonly used, following the RGB color model, informally referred to as "true color." In this model, only red, green, and blue are employed to create a 24-bit image, with each color represented by eight bits. Consequently, a single pixel can exhibit 256 different levels of red, green, and blue, out of a potential 16 million. As a file's size increases, so does its capacity to display a broader spectrum of colors [8]. Picture steganography encompasses two primary techniques: those that operate in the Image Domain and those functioning within the Transform Domain.

Transform-domain techniques, also known as frequency-domain techniques, initially transform images before embedding the message into the altered image. Conversely, Image Domain techniques encompass methods that involve bit insertion and noise adjustment directly in the image domain. For image-domain steganography, the ideal choice is an image format unaffected by compression, as the methods used for encryption or decryption often depend on the specific image format employed. Steganography can employ algorithms, visual transformations, or information manipulations to conceal data. In these approaches, information is hidden within more prominent areas of the cover photo, making it more resistant to alterations. Many transform domain methods can preserve encoded information even when the image undergoes compression via lossy or lossless methods, as they are not reliant on the image format used. Sections discussing stenographic methods are typically categorized based on the image file formats utilized and the domains in which they are applied. Figure 1 illustrates the possibility of utilizing images as cover media [9].

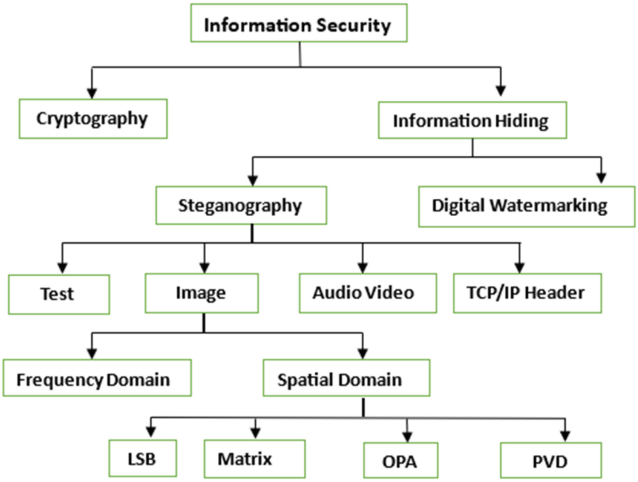


Fig. 1. Image processing Stenography [3]

4 Related Work

In the context of online privacy concerns, deep learning plays a pivotal role in real-time image data hiding. High-level semantic feature extraction through DenseNet enhances precision compared to low-level approaches. This method categorizes internet photos based on user needs, extracting high-level semantic properties. Combining feature sequences, DC, and geographic coordinates with DCT yields a robust hash sequence. An inverted index structure based on the hash sequence enables rapid image matching. Once matching stego-images are found, they are transmitted to the receiver for feature retrieval. The secret picture is reconstructed using received stego-images and transmitter-provided position information. Experimental results demonstrate that this method offers superior robustness, retrieval accuracy, and capacity without detectable modification traces compared to existing techniques for concealing information within images [10]. Coverless information concealment has gained prominence for its resistance to steganography. A novel approach leverages the creation of anime characters as a foundation for concealing information. Attribute labels derived from secret information transform into character attributes. Generative adversarial networks (GANs) are then used to generate anime characters, addressing prior GAN challenges. Test results indicate superior image quality and concealment capacity, with the proposed method offering approximately 60 times more concealed capacity compared to traditional methods [11]. Image interpolation for steganography presents a security challenge. A novel data-concealing strategy combines OPAP and LSB substitution, aiming to surpass previous methods. However, it remains susceptible to RS detection, necessitating a new approach. By merging GEMD and RGEMD with picture interpolation, a secure and efficient information-hiding strategy is introduced, improving steganographic image quality [12]. DNA-based steganography introduces high capacity, randomness, and low modification rates. Various DNA concealment techniques are evaluated, with replacement yielding optimal Bit per Nucleotide (BPN) for short hidden signals while maintaining low visibility. Insertion proves more challenging but is equally secure compared to replacement [13]. Steganography conceals confidential messages within image least significant bits. The original image is restored by substituting new bytes using a mathematical approach. Following encryption and image digitization, statistical testing evaluates image quality. Specific procedures ensure accurate concealment and extraction sequences. The study culminates in data decryption, utilizing both private and public keys [13]. Covert information concealment provides a secure and durable method for hiding data within images. Deep convolution features connect visual cues with confidential data using convolutional neural networks. Depth features are generated, transformed into binary sequences, and hashed. The method offers high detection accuracy and reduces feature usage [14]. A method for efficient feature selection in steganalysis is presented, reducing classification training time by 50%. Reparability values guide feature component selection, evaluating their impact on image classification. This approach outperforms existing methods in terms of feature selection time, detection accuracy, and dimension reduction [14].

This study suggests several important findings and implications:

- Steganography's Effectiveness: The study underscores the effectiveness of steganography as a method for concealing secret information within digital media without arousing suspicion. It highlights that this technique can hide large volumes of data within cover files without any alteration to the cover file itself, offering a powerful means of covert communication.
- Deep Learning Enhancements: The study demonstrates how deep neural networks, particularly VGG-16 and VGG-19 models, can significantly enhance the capacity and security of steganography. By utilizing these networks for feature extraction and embedding, it becomes possible to hide information more efficiently.
- Steganography Detection: The research showcases the use of CNNs for steganography detection. The CNNs exhibit high accuracy, achieving a 100% detection rate, which implies that advanced machine learning techniques can effectively identify hidden information.
- Security and Privacy: With the improved techniques presented in the study, individuals and organizations may employ steganography for secure data transmission,

copyright protection, and privacy preservation. It could have applications in fields where confidentiality and data integrity are paramount.

5 Methodology

Steganography is a type of encryption in which digital data serves as the carrier, and networks allow for the rapid delivery of the message that is being concealed. Steganography operates in conjunction with networks. The sequence of the events For the aim of storing and transferring a wide variety of communication data, there are a number of different encoding and decoding methods that are utilized at every given point in time in the development of information technology [16–21].

These methods are used for their ability to encode and decode information. The pixel in the upper left corner is utilized rather frequently during the process of encoding and decoding data, which ultimately results in signals that are immediately recognizable and constant. Encoding or decoding pixels that are positioned in close proximity to one another is a typical application of this technique, despite the fact that it is more difficult to generate. The construction of an initialized concept is exceptionally difficult owing to the fact that it is often utilized without a specific starting point in mind. This makes the construction of those thoughts extremely difficult.

When it comes to this particular scenario, there is no preconceived pattern; the selection is decided exclusively on the basis of the pixel location that is provided by the Random Number Generator after the process of encoding or decoding has been finished. However, there is no predetermined strategy for encoding or decoding that may be applied in the process of doing histogram analysis in order to uncover a faster recovery rate. This is because there is no predetermined plan. As a result of the fact that doing so speeds up the recovery process, pre-defining an encoding scheme is a typical practice. To put it into perspective, determining the significance of the message is not an easy task. It is stated in this section that an outline of the research methodologies that were applied in the study is presented. In this section, not only is an overview of the research concept presented, but also the methodology and data sets that were utilized in the study are detailed.

The military, the government, and the internet are all finding more and more uses for steganography. Cryptography was developed and many different methods of encrypting and decrypting data have been developed to ensure the confidentiality of communications. Sometimes it's not enough to merely conceal a message's contents in order to prevent its disclosure. Also, secrecy surrounding the existence of the communication might be mandatory. A method known as steganography is used to achieve this goal. When compared to cryptography, its primary goal is not message security but rather message opacity. However, steganography differs from encryption in that its primary goal is to conceal the message itself rather than its content. To beef up security, creative uses of cryptography and steganography were implemented. For the most part, two methods are employed: The conventional approach is by far the most common. In the first, data is encoded sequentially from left to right, while in the latter, data is encoded randomly across the image in an effort to make it less noticeable to the observer. Analysis of an image typically masks the fact that both approaches were used.

Incorporating both image compression and encryption/decryption into a single process, these techniques strengthen the safety of sensitive information. In addition to reducing the amount of data that needs to be delivered and the amount of storage space required, this technology can encode or store information using the RGB pixels in a cover image without altering its visual appearance (Fig. 2).

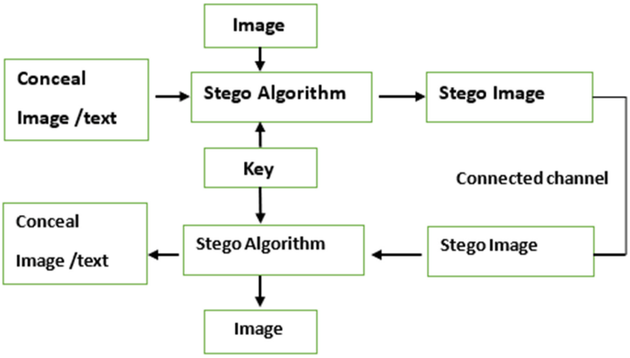


Fig. 2. The Least Significant Bit (LSB)

6 Image Detection

Machine learning solved the problem by coming up with a different approach that didn't require any human coders to do any of the heavy lifting. Characteristics of the image were generated by a computer using machine learning. This movement was pioneered by smaller applications that could identify distinct visual patterns. Recognizing patterns, classifying images, and identifying objects within them all require the use of statistical learning algorithms like linear regression, logistic regression, and support vector machine. Predicting cancerous cells is just one example of the many challenging problems that machine learning has the potential to address.

However, a lot of effort was required to develop the features. Traditional methods of machine learning involved extensive work and the input of many different types of experts, including but not limited to computer scientists, mathematicians, and engineers. Deep learning is an approach to machine learning that differs greatly from more conventional approaches. This is achieved by feeding examples into a neural network trained to solve similar problems[15]. When given the input, the neural network can recognize commonalities among the samples and generalize about them. After identifying these patterns, a mathematical equation is developed to aid in the forecasting process. With just a pre-made algorithm and some face examples for training, a deep learning network can recognize human faces without any additional parameters being specified. This also applies to the earlier described instance of facial recognition. Deep learning is considered one of the most efficient approaches to computer vision. In order to retrain the model for new purposes, a good deep learning algorithm takes into account a large number of datasets that have already been trained. Parameters in this context can refer to many different things, including the number of hidden layers, the type of layers, the number of training epochs, and many others. When compared to more conventional machine learning methods, deep learning can be developed and implemented at a much quicker pace. Self-driving cars, cancer detection, and facial recognition are three prominent applications of deep learning. When it comes to processing massive amounts of complex data, deep learning is the machine learning technique of choice. Geoff Hinton, Yann Lecun, Andrew Ng, Andrej Karpathy, and Yoshua Bengio are just some of the well-known researchers who are concentrating on deep learning right now.

Many large corporations, including Google, Apple, NVIDIA, Toyota, and many more, are actively engaged in deep learning. One of the main goals of deep learning is to simulate the functioning of the human brain. Image processing, natural language processing, biology, autonomous driving, artificial intelligence, and countless other fields have all seen incredible advancements thanks to deep learning in recent years.

Convolution layer:

$$\mathrm{z}^{1} = \mathrm{h}^{1-1} * \mathrm{W}^{1} \tag{1}$$

Maxpooling layer:

$$h_{xy}^{1} = \max_{i=0..,j=0..s} h^{1-1}(x+i)(y+j) \tag{2}$$

Fully-connected layer:

$$\mathrm{z}_{l} = W_{l} * \mathrm{h}_{l-1} \tag{3}$$

Relu layer:

$$\mathrm{ReLU}(\mathrm{z_i}) = max(0, \mathrm{z_i}) \tag{4}$$

Softmax layer (Fig. 3):

$$\mathrm{softmax}(z_i) = \frac{e^{zi}}{\sum_j e^{ij}} \tag{5}$$

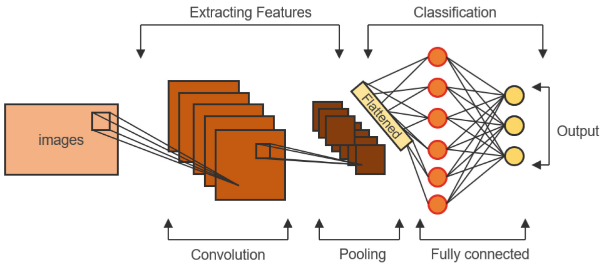


Fig. 3. Convolution neural network architecture

7 Results and Discussion

The LSB technique, which relies on optical illusions, is commonly used to disguise the message or picture. You can choose to select the pixels either randomly or sequentially. Steganography is enhanced by encrypting data and hiding it within the least significant bit (LSB). Steganography is a method that alters the pixel values of a picture represented in the spatial domain. The information that is believed to be confidential is concealed by altering the Least Significant Bits. The probability of the image distortion becoming apparent due to alterations in the least significant bits (LSB) is highly improbable. The utilization of MSBs for concealing information depends on the intensity value of the data. An intruder cannot extract the key from the original image as it is one of the components included within the shot.

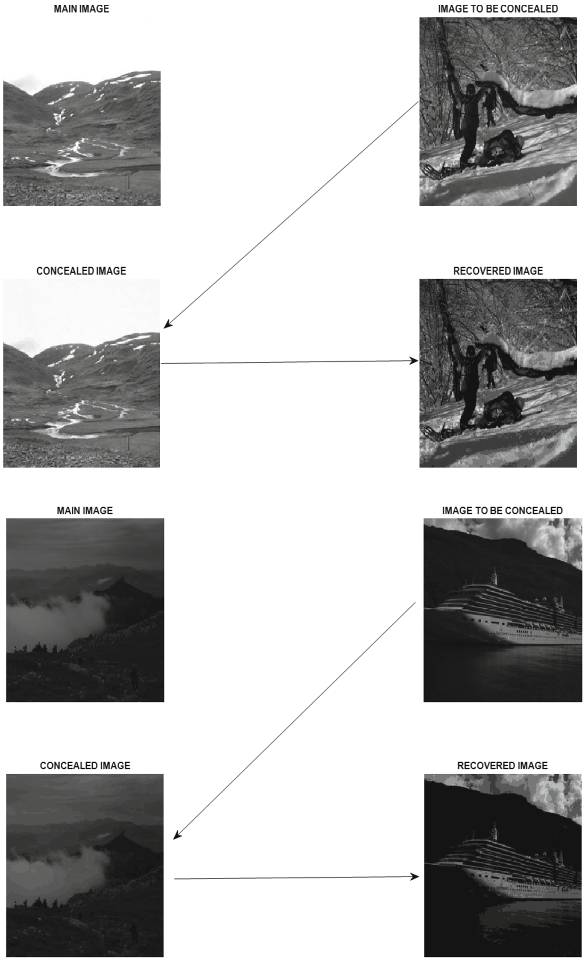


Fig. 4. LSB image steganography

In this thesis, Fig. 4 provides a detailed demonstration of the process of embedding one image within another image. The explanation is derived from the discoveries of the investigation. Figure 5 illustrates the presence of both regular and steganographic images. The training results of the Vgg16 model demonstrate a consistent process of data training and rapid model learning, which can be attributed to its particular morphologies. This is achieved by extracting and classifying the characteristics of the pictures. Although I first encountered some delay during the verification process, I managed to overcome this issue by utilizing the improved likelihood of the (softmax) layer. The confusion matrix presents the model's results, showcasing the findings with utmost precision and accuracy. Compared to its predecessor, the Vgg-19 model has more layers, leading to a larger difference in features and a potential little delay in the verification process. The model test obtained a perfect accuracy rate of 100%, as evidenced by the confusion matrix (Figs. 6 and 7).

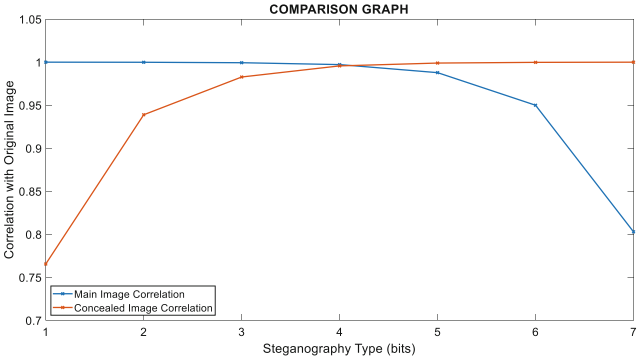


Fig. 5. Comparison graph for first image

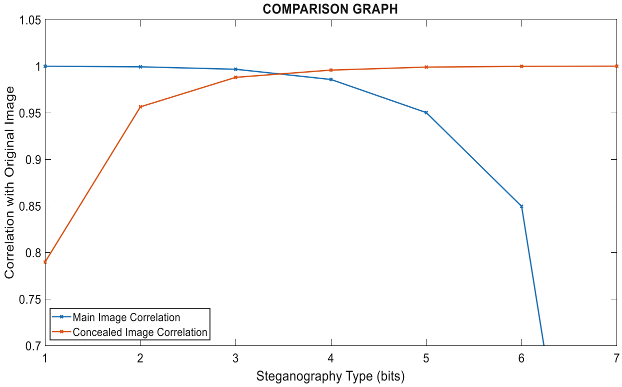


Fig. 6. Comparison graph for second image

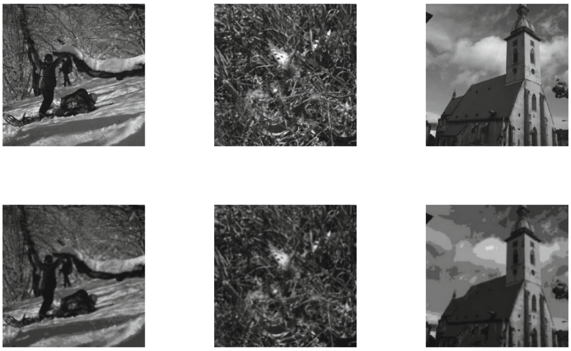


Fig. 7. Samples of normal and stego dataset in this study

For the purpose of resolving the supervised learning problem of picture classification, it is necessary to first construct a collection of target classes, which are all of the different things that can be recognized in photographs. After that, it is necessary to train a model to recognize these target classes by making use of sample images that have been labeled. It was during the early phases of computer vision when raw pixel data was considered to be an essential component. An undertaking that is both wonderful and incredibly worthwhile is the endeavor of attempting to classify an image in order to have a better understanding of the picture. It is essential to provide the photograph with a name that is significant in order to guarantee that it is positioned in the correct folder. When most people talk about image categorization, they are particularly talking to the process of evaluating monocular images.

This is the case in the majority of cases. Element detection, on the other hand, is a method that takes into account more practical scenarios in which a picture may contain a number of different components. This method is a technique that is used to identify elements. In order to successfully accomplish this stage, it is essential to make use of both classification and localization practices (Figs. 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 and Table 1).

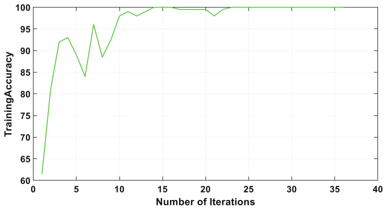


Fig. 8. Vgg-16 training *accuracy.*

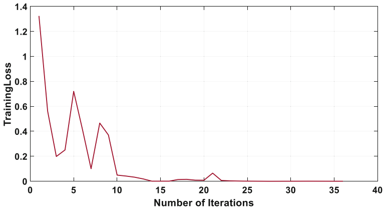


Fig. 9. Vgg-16 training losses.

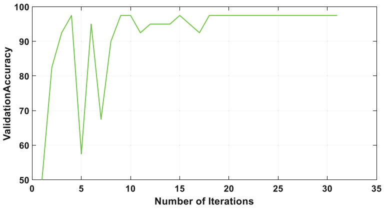


Fig. 10. Vgg-16 validation accuracy

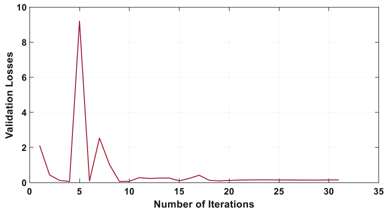


Fig. 11. Vgg-16 validation losses

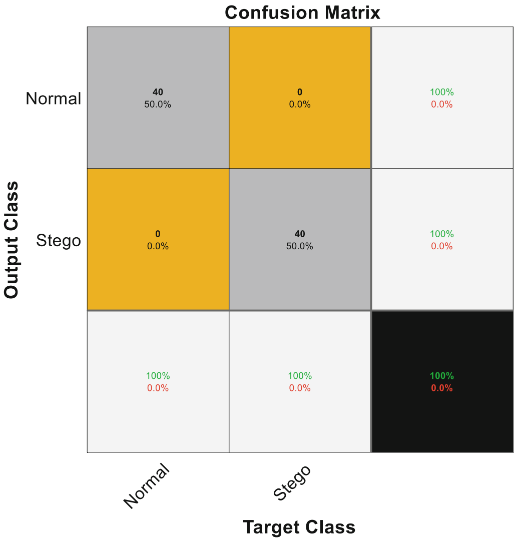


Fig. 12. Vgg-16 confusion matrix

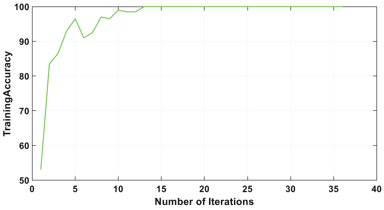


Fig. 13. Vgg-19 training accuracy

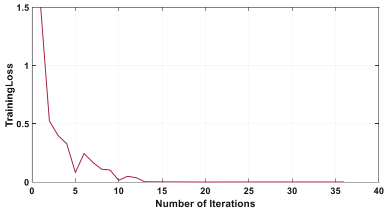


Fig. 14. Vgg-19 training losses

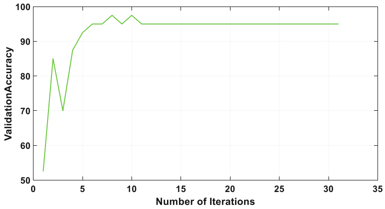


Fig. 15. Vgg-19 validation accuracy

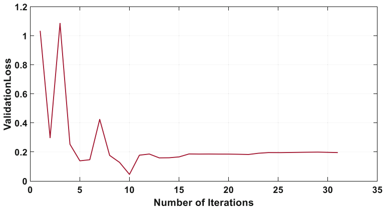


Fig. 16. Vgg-19 validation losses

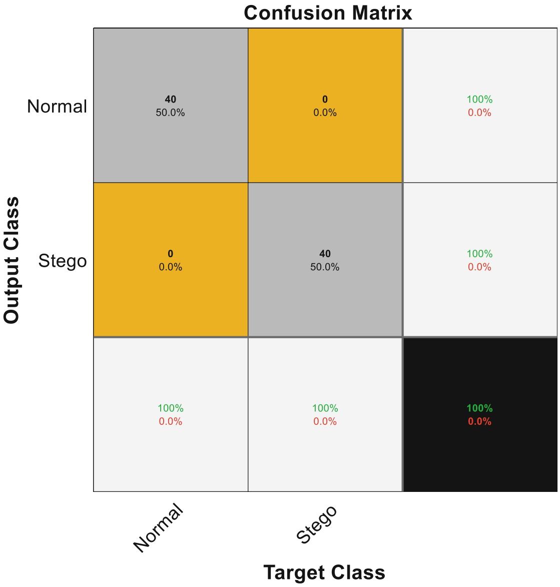


Fig. 17. Confusion matrix

Table 1. Training metricses for models

Metrics	Vgg-16	Vgg-19
min	61.5	53
max	100	100
mean	96.31	96.82
median	99.5	100
mode	100	100

8 Conclusion

Cryptography employs encryption techniques to ensure the confidentiality of data during transmission. Nevertheless, imposters frequently employ this encrypted iteration, potentially resulting in harm. Watermarking can also safeguard, confirm, and validate users. This ensures the protection of intellectual property rights. A carrier, without any disguise, holds a marker during the procedure. Therefore, digital watermarks can be recognized. Steganography is the practice of hiding messages within apparently harmless digital content. Steganography is derived from the Greek terms stegos, meaning "cover," and grafia, meaning "writing." The research commences with the observation that steganography is executed covertly, without altering the cover file. The method can conceal secret messages of any magnitude by comparing the ASCII codes of the secret characters to those of the picture cover file. Utilizing a character from a previous appearance to regain a position is acceptable. Therefore, the size of the hidden text message (picture) will not impose any limitations. Steganography hides confidential data within an imperceptible medium without any mistakes. Deep neural networks can utilize a hidden network and a decoding network to disguise one image within another, surpassing the limitations of conventional information hiding and steganography techniques.

This study employs the residual block within the hidden network and determines that Vgg-16 and Vgg-19 enhance the process of feature extraction. Initially, the covert network embedded the confidential image within the concealment image. The steganography picture we have created is meticulously crafted to closely resemble the cover image during the encoding process. In order to reveal the hidden image, the recipient transmits the steganographic image to the decoding network. The experimental findings demonstrate that the technique will enhance the integration of information and resolve issues related to visual signals.

The stego picture was obtained by the utilization of LSB image steganography. The recognition was achieved by the utilization of deep learning, namely convolutional neural networks. The findings demonstrated that the models effectively detected steganographic images. The integration of cutting-edge techniques, such as convolutional neural networks and innovative approaches like the one proposed here, will likely play a pivotal role in enhancing the concealment and detection of hidden information within digital media. Furthermore, the fusion of cryptography, watermarking, and steganography will contribute to more robust and versatile solutions for securing data and verifying authenticity in an increasingly digital world.

References

1. Dalal, M., Juneja, M.: A survey on information hiding using video steganography. Artif. Intell. Rev. 1–65 (2021)
2. Duan, X., Li, B., Xie, Z., Yue, D., Ma, Y.: High-capacity information hiding based on residual network. IETE Tech. Rev. **38**(1), 172–183 (2021)
3. Elshoush, H.T., Ali, I.A., Mahmoud, M.M., Altigani, A.: A novel approach to information hiding technique using ASCII mapping based image steganography. J. Inf. Hiding Multimedia Signal Process. **12**(2), 65–82 (2021)

4. Kshirsagar, A., Shah, M.: Anatomized study of security solutions for multimedia: deep learning-enabled authentication, cryptography and information hiding. Adv. Secur. Solut. Multimedia (2021)
5. Qin, C., Qian, Z., Li, X., Wang, J.: Artificial intelligence oriented information hiding and multimedia forensics. IETE Tech. Rev. **38**(1), 1–4 (2021)
6. Praghash, K., Vidyadhari, C., NirmalaPriya, G., Cristin, R.: Secure information hiding using LSB features in an image. Mater. Today Proc. (2021)
7. Kumar, S., Singh, V.: Information hiding techniques for cryptography and steganography. In: Computational Methods and Data Engineering, pp. 511–527. Springer, Singapore (2021)
8. Cabaj, K., Caviglione, L., Mazurczyk, W., Wendzel, S., Woodward, A., Zander, S.: The new threats of information hiding: the road ahead. IT Prof. **20**(3), 31–39 (2018)
9. Taha, M.S., Rahim, M.S.M., Hashim, M.M., Khalid, H.N.: Information hiding: a tools for securing biometric information. Tech. Rep. Kansai Univ. **62**(04), 1383–1394 (2020)
10. Meng, R., Zhou, Z., Cui, Q., Sun, X., Yuan, C.: A novel steganography scheme combining coverless information hiding and steganography. J. Inf. Hiding Priv. Prot. **1**(1), 43 (2019)
11. Hashim, M., Mohd Rahim, M.S., Alwan, A.A.: A review and open issues of multifarious image steganography techniques in spatial domain. J. Theor. Appl. Inf. Technol. **96**(4), (2018)
12. Ghosal, S.K., Mukhopadhyay, S., Hossain, S., Sarkar, R.: Application of Lah transform for security and privacy of data through information hiding in telecommunication. Trans. Emerging Telecommun. Technol. **32**(2), e3984 (2021)
13. Majeed, M.A., Sulaiman, R., Shukur, Z., Hasan, M.K.: A review on text steganography techniques. Mathematics **9**(21), 2829 (2021)
14. Mohsin, A.H., et al.: PSO–Blockchain-based image steganography: towards a new method to secure updating and sharing COVID-19 data in decentralised hospitals intelligence architecture. Multimedia Tools Appl. **80**(9), 14137–14161 (2021)
15. Wang, Z., Yin, Z., Zhang, X.: Distortion function for JPEG steganography based on image texture and correlation in DCT domain. IETE Tech. Rev. 1–8 (2017)
16. Pandey, S., Parganiha, V.: Hiding secret image in video. Int. J. Res. Sci. Eng. **3**, 1–9 (2017). e-ISSN 2394-8299. p-ISSN 2394-8280
17. Yang, C.-N., Kim, C., Lo, Y.-H.: Adaptive real-time reversible data hiding for JPEG images. J. Real-Time Image Process. 1–11. Springer (2016)
18. Pelosi, M.J., Kessler, G., Brown, M.S.S.: One-time pad encryption steganography system. In: Annual Conference on Digital Forensics, Security and Law. 4. CDFSL Proceedings 2016 (2016)
19. Jiang, N., Zhao, N., Wang, L.: LSB based quantum image steganography algorithm. Int. J. Theor. Phys. **55**(1), 107–123 (2016)
20. Dogan, S.: A new approach for data hiding based on pixel pairs and chaotic map. Int. J. Comput. Netw. Inf. Secur. (IJCNIS) **10**(1), 1–9 (2018). https://doi.org/10.5815/ijcnis.2018.01.01
21. Sharifzadeh, M., Agarwal, C., Salarian, M., Schonfeld, D.: A new parallel message distribution technique for cost-based steganography (2017). arXiv preprint arXiv:1705.08616

Performance Analysis for Web Scraping Tools: Case Studies on Beautifulsoup, Scrapy, Htmlunit and Jsoup

Yılmaz Dikilitaş[1](✉), Çoşkun Çakal[2], Ahmet Can Okumuş[1], Halime Nur Yalçın[1], Emine Yıldırım[1], Ömer Faruk Ulusoy[1], Bilal Macit[2], Aslı Ece Kırkaya[2], Özkan Yalçın[2], Ekin Erdoğmuş[2], and Ahmet Sayar[2]

[1] Kocaeli University, 41001 Kocaeli, Turkey
yilmazdikilitas91@gmail.com
[2] Haratres Teknoloji, Kocaeli, Turkey

Abstract. Web scraping has become an indispensable technique for extracting valuable data from websites. With the growing demand for efficient and reliable web scraping tools, it is crucial to assess their performance to guide developers and researchers in selecting the most suitable tool for their needs. In this paper, we present a comprehensive performance analysis of four popular web scraping tools: BeautifulSoup, Scrapy, HtmlUnit, and Jsoup. Our study focuses on evaluating these tools based on metrics such as execution time, memory usage, and scalability. We conducted experiments using various websites and datasets to provide a comprehensive evaluation of the tools' performance. The results highlight the strengths and limitations of each tool, allowing users to make informed decisions when choosing a web scraping tool based on performance requirements. Additionally, we discuss real-world use cases and the impact of website structures on tool performance. This paper aims to assist developers and researchers in selecting the most appropriate web scraping tool for their specific needs, and it also identifies avenues for future research to further enhance the performance of these tools.

Keywords: Web Scraping · Scrapy · Beautifulsoup · Jsoup · HtmlUnit

1 Introduction

Web scraping, the process of extracting data from websites, has gained significant importance in various domains such as data analytics, research, and business intelligence. It enables the automated collection of valuable information from a wide range of online sources. As web scraping continues to evolve, numerous tools have emerged to facilitate this task, each offering different functionalities, features, and performance characteristics.

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 471–480, 2024.
https://doi.org/10.1007/978-3-031-56728-5_39

The performance of web scraping tools plays a critical role in determining their effectiveness and efficiency in data extraction. The selection of an appropriate tool depends on factors such as execution time, memory usage, scalability, and adaptability to different website structures. However, the lack of comprehensive performance analysis often leaves developers and researchers uncertain about which tool to choose for their specific requirements.

In this paper, our aim is to bridge this gap by conducting a thorough performance analysis of four popular web scraping tools. BeautifulSoup, Scrapy, HtmlUnit, and Jsoup. These tools were chosen based on their popularity, widespread usage, and diverse functionalities. Our study focuses on evaluating its performance on various metrics to provide a comprehensive assessment of its strengths and limitations.

Web scraping tools such as Scrapy, BeautifulSoup, HtmlUnit, and Jsoup have been widely used in various research studies and applications. These tools have been used in different domains, including monitoring cyber trafficking, sentiment analysis, job market analysis, online reviews analysis, and disease detection. The following studies highlight the use and performance analysis of these web scraping tools in various contexts.

In a study by (Acerado, 2023), a web application for cyber monitoring and trafficking integrated with a web scraper was developed using Beautiful Soup. The sensitivity analysis conducted in the study determined that Beautiful Soup was the most suitable tool for developing web scraping algorithms based on performance, portability, and accuracy rates compared to Scrapy and Selenium [1].

Moro et al. (2019) conducted experiments using web scraped data from online sources to analyze hotel online reviews. They highlighted the advantages of web scraping, such as the retrieval of freely written opinions and the collection of a large volume of information at high speed [2].

Han and Anderson (2020) discussed popular Python libraries for web scraping, including BeautifulSoup and Scrapy, in the context of hospitality research. They provided instructions and insights on the use of these libraries for online data collection [3].

Wooldridge and King (2018) mentioned the use of web scraping tools and APIs to track the attention of online media in the context of alternative metrics (altmetrics) for the evaluation of research impact [4].

Zucco et al. (2019) conducted a review of sentiment analysis methods and tools, including web scraping tools, to mine text and social networks data. They compared and analyzed 24 tools based on criteria such as usability, flexibility, and type of analysis performed [5].

Pellert et al. (2020) built a self-updating monitor of emotion dynamics during the COVID-19 pandemic using web scraping and API access. They used Web scraping to retrieve data from news platforms, Twitter, and chat platforms for sentiment analysis [6].

Alrusaini (2023) used web scraping with BeautifulSoup, SERP API, and request libraries to obtain skin images for the detection of Monkeypox disease.

Deep learning models were trained on the scraped data for accurate disease detection [7].

Two points are very important in terms of performance in web scraping. The first is the time it takes to transfer the data and the second is the time it takes to parse the data. There have been several studies in the literature on data transfer acceleration. Some of them are related to changing the format of the data ([8–11]), some are related to data dilution ([12] and [13]) and some are related to parallel transfer of data ([14]and [15]). Since we will pull the data directly with the HTML protocol over port 8080, such approaches do not provide us with a solution. Analyzes in parsing the data rather than pulling the data will be done. The work presented in this paper will evaluate the performance of web scraping tools in parsing HTML pages and other identified features.

These studies demonstrate the diverse applications of web scraping tools in different research domains. Performance analysis of these tools highlights their suitability for specific tasks based on factors such as performance, accuracy, and ease of use.

The objectives of this research are two-fold. First, we aim to assess the execution time of each tool, considering factors such as parsing speed and network latency. Second, we analyze memory usage to understand the impact of each tool on system resources.

By conducting rigorous experiments and benchmarking against real-world websites and datasets, we provide an in-depth analysis of the performance characteristics of these web scraping tools. The insights gained from this study will help developers and researchers make informed decisions when selecting a tool based on their performance requirements.

The remainder of this paper is organized as follows. Section 2 provides a state-of-the-art on web scraping and performance analysis. Section 3 describes the methodology employed for our performance analysis, including the experimental setup and the metrics used for evaluation, presents the results of our performance analysis, highlighting the strengths and limitations of each tool. Finally, Sect. 4 concludes the paper by summarizing the key findings and providing recommendations for selecting a web scraping tool based on performance requirements.

2 State of Art

Web scraping is a technique that is used to extract data from websites automatically. It involves targeting specific webpages, extracting the underlying HTML code, parsing the relevant data, and saving it to a file system or database for further analysis (Darmawan et al., 2022). Web scraping has become an essential tool in various domains, including healthcare, psychology, website monitoring, entrepreneurship research, film analysis, and credibility analysis on social media platforms [16].

In the healthcare field, web scraping has been used to collect digital images of skin lesions for diseases such as Monkeypox, Chickenpox, Smallpox, Cowpox, and Measles (Islam et al., 2022). However, the lack of publicly available and reliable digital image databases for certain diseases, such as monkeypox, has led researchers to utilize web scraping to collect the necessary data (Islam et al., 2022) [17].

Psychological research has also benefited from web scraping, particularly in the extraction of big data from the Internet for use in studies (Landers et al., 2016). Researchers are encouraged to determine their research questions and hypotheses before using web scraping to address those questions, to avoid the pitfalls associated with hypothesizing after the results are known (Landers et al., 2016) [18].

Web scraping has proven to be valuable in website monitoring systems, allowing automatic checks of website availability (Arhandi et al., 2021). By using web scraping techniques and tools such as Raspberry Pi, researchers have achieved high accuracy in monitoring website availability (Arhandi et al., 2021) [19].

Entrepreneurship research has also used web scraping to collect large-scale data on entrepreneurial activities (Quinn et al., 2022). These data can be used to develop and test theories in entrepreneurial research (Quinn et al., 2022) [20].

In film analysis, web scraping has been used to study cultural phenomena such as hipster culture. By analyzing data collected through web scraping, researchers can gain insight into the prevalence and impact of hipster culture in various contexts.

Web scraping has also been used for credibility analysis on social media platforms such as Twitter. It has been compared to other extraction methods, such as API methods, to analyze credibility on social media platforms.

Web scraping techniques have been evaluated and compared in terms of their performance and effectiveness. For example, a study evaluated the performance of web scraping techniques using XPath, CSS Selector, Regular Expression, and HTML DOM with multiprocessing technical applications (Darmawan et al., 2022). The study found that web scraping is an effective and efficient technique for extracting and storing data from websites (Darmawan et al., 2022) [16].

In general, web scraping has become an important tool in various fields for data collection, analysis, and research purposes. It offers advantages such as efficiency, scalability, and the ability to collect data from diverse sources on the Internet. However, it is important for researchers to carefully consider their research questions and hypotheses before using web scraping to ensure its suitability for their specific research needs. Furthermore, researchers should be aware of the challenges associated with web scraping, such as changes in website structure and availability of data.

3 Methodology

In the performance analysis of web scraping tools, the selection of appropriate websites or data sets for testing and benchmarking plays a crucial role in obtaining reliable and meaningful results. Researchers must carefully consider several factors to ensure a comprehensive evaluation of the performance of the tools. First, the choice of website should encompass a diverse range of structure and complexity. Moreover, websites featuring dynamic content generated through JavaScript should be included to assess the tools' handling of such scenarios. Real-world websites are essential for ensuring the evaluation's relevance to practical use cases. These websites often exhibit variations in coding practices, responsiveness, and content presentation, providing a more accurate representation of the challenges faced during actual web scraping tasks.

As a method, information of 24 products was taken on the homepage of an e-Commerce site. This process was repeated 100 times for all tools. The Table 2 was created by taking the average of these epochs. Specifications of the computer used in this experiment; AMD Ryzen 5600X, 16 GB 3600 DDR5 RAM, 6600XT 8 GB, 1 TB SSD (1 GB Read/1 GB Write).

The size and scale of the website used for testing are critical factors in covering a broad spectrum of scenarios. This includes large-scale website with extensive content to evaluate the tools' performance under varying data volumes. Additionally, the data set should contain diverse data types and formats, such as structured data such as tables, unstructured text, images, and multimedia content, to test the tools' parsing and extraction capabilities across different data representations. We choose very large e-commerce website for these reasons.

Ethical considerations are paramount that must obtain permission from website owners before conducting scraping activities to respect the websites' terms of service and legal constraints. Adherence to the website's robots.txt file is also essential to avoid overloading the servers and maintaining ethical scraping practices (Table 1).

Table 1. Comparison of Web Scraping Libraries

Library	Language	DOM Parsing	JavaScript Execution
Scrapy	Python	No	Yes
BeautifulSoup	Python	Yes	No
Jsoup	Java	Yes	No
HtmlUnit	Java	Yes	Yes

Table 2. Performance Comparison of Web Scraping Libraries

Library	Memory Usage	CPU Usage	Working Time
Scrapy	2400 MB	2.8	8.1 s
BeautifulSoup	8500 MB	3.7	8.6 s
Jsoup	2150 MB	1.5	7.3 s
HtmlUnit	2600 MB	4.1	9.7 s

Memory Usage: This metric refers to the amount of memory consumed by each web scraping library during the scraping process. It provides an indication of the library's efficiency in managing memory resources. Libraries with low memory usage are more efficient in terms of memory consumption, which can be advantageous when dealing with large-scale scraping tasks or limited system resources. On the other hand, libraries with moderate or high memory usage may be more suitable for projects that require more advanced features or handling of complex data structures.

CPU Usage: This metric measures the CPU utilization of each web scraping library. Reflects the amount of computational resources the library requires to perform scraping tasks. Libraries with low CPU usage are more efficient in terms of utilizing system resources, which can result in faster scraping times and lower overall CPU load. However, libraries with moderate or high CPU usage may be necessary for projects that involve complex data processing, JavaScript execution, or other computationally intensive operations.

Working time: This metric represents the time it takes each web scraping library to complete a scraping task. Indicates the efficiency and speed of the library in retrieving and processing web data. Libraries with fast working times are advantageous when quick results are desired or when dealing with time-sensitive data. Moderate working times may be acceptable for most scraping tasks, while slower working times might be justified for projects that require extensive data processing or interaction with dynamic web elements.

By evaluating these metrics, developers can assess the performance characteristics of each library and choose the one that best suits their specific scraping requirements. It is important to consider the trade-offs between memory usage, CPU utilization, and working time based on the project's constraints and priorities.

The Table 2 provides a comprehensive comparison of four popular web scraping libraries: Scrapy, BeautifulSoup, Jsoup, and HtmlUnit. The evaluation is based on three key variables: memory usage, CPU usage, and working time. Scrapy, a powerful Python framework, offers moderate memory usage and CPU utilization, resulting in a balanced working time. It provides a high-level interface for handling complex scraping tasks and is particularly useful for projects that require advanced features or support for JavaScript execution.

BeautifulSoup, a Python library focused on HTML parsing, stands out with its medium memory and low CPU usage. This makes it an efficient choice for scraping tasks that prioritize speed and resource efficiency. Excels at extracting data from HTML documents and offers a simple and intuitive API, making it suitable for smaller projects or cases where a lightweight solution is desired.

Jsoup, a Java library designed for HTML parsing and manipulation, exhibits characteristics similar to those of BeautifulSoup. It also boasts low memory and CPU usage, enabling fast scraping operations. The strength of Jsoup lies in its ability to navigate and manipulate the HTML structure, making it ideal for extracting specific data elements from HTML documents.

HtmlUnit, a Java-based library, takes a different approach by simulating a Web browser environment. While it requires a moderate amount of memory, it utilizes higher CPU resources due to its JavaScript execution capabilities. HtmlUnit is suitable for scraping tasks that involve dynamic web content or require interaction with JavaScript-based elements.

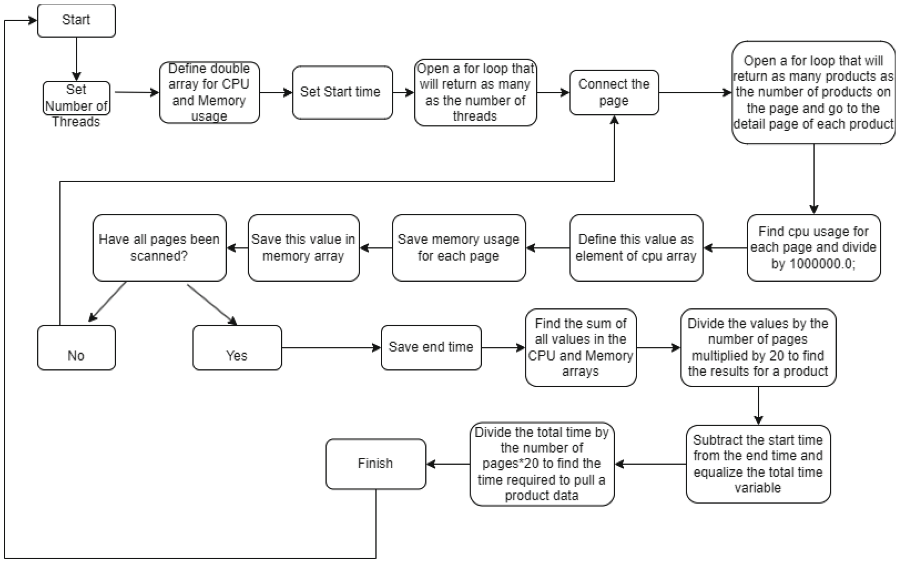

Start
Set Number of Threads
Define double array for CPU and Memory usage
Set Start time
Open a for loop that will return as many as the number of threads
Connect the page
Open a for loop that will return as many products as the number of products on the page and go to the detail page of each product
Find cpu usage for each page and divide by 1000000.0;
Define this value as element of cpu array
Save memory usage for each page
Save this value in memory array
Have all pages been scanned?
No
Yes
Save end time
Find the sum of all values in the CPU and Memory arrays
Divide the values by the number of pages multiplied by 20 to find the results for a product
Subtract the start time from the end time and equalize the total time variable
Divide the total time by the number of pages*20 to find the time required to pull a product data
Finish

Fig. 1. Process of Used System

Performance analysis of web scraping libraries reveals several interesting findings. Figure 1 shows in detail how the process works and how the runtime is measured.

Firstly, Scrapy, a Python-based framework, demonstrates a balance between memory usage, CPU usage, and work time. This makes it a versatile choice for projects that require advanced features, support for JavaScript execution, and moderate resource consumption. However, for projects where speed and resource efficiency are prioritized over advanced functionality, BeautifulSoup and Jsoup emerge as strong contenders. Both libraries exhibit low memory and CPU usage, resulting in faster scraping operations and efficient resource utilization.

Another noteworthy finding is the impact of JavaScript execution on resource requirements. HtmlUnit, the Java-based library with JavaScript execution capabilities, demonstrates moderate memory usage and higher CPU utilization. This indicates that projects requiring dynamic web content or JavaScript interaction may benefit from HtmlUnit's capabilities, albeit with a trade-off in terms of resource consumption and working time.

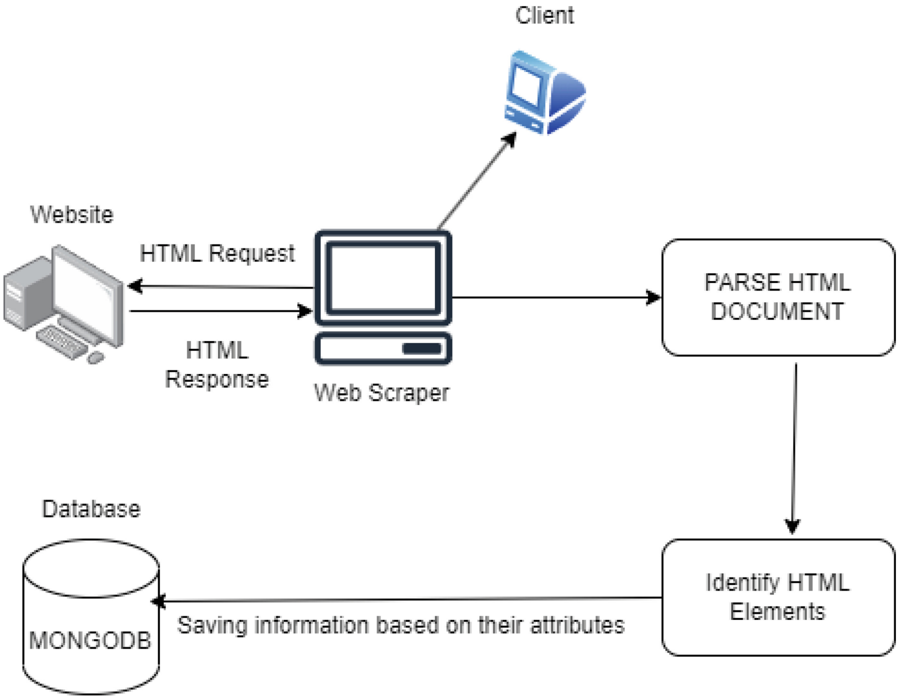


Fig. 2. Architecture of System

In general, the choice of a web scraping library depends on the specific requirements and constraints of the project. Figure 2 shows how the system works and is set up. For simpler scraping tasks that prioritize speed and resource efficiency, BeautifulSoup and Jsoup provide lightweight options. Scrapy, on the other hand, offers a comprehensive solution for more complex scraping scenarios, especially when advanced features and JavaScript execution are needed. HtmlUnit serves as a specialized option for projects with specific requirements for JavaScript handling.

It is crucial for developers to carefully evaluate the trade-offs between memory usage, CPU utilization, and working time based on the specific needs of their project. By considering these factors, developers can select the web scraping library that best aligns with their desired performance characteristics and resource constraints.

4 Conclusion

In this analysis, we compared the performance of four popular web scraping libraries: Scrapy, BeautifulSoup, Jsoup, and HtmlUnit. By evaluating metrics such as memory usage, CPU utilization, and working time, we gained insight into the strengths and trade-offs of each library.

Scrapy stood out as a powerful framework that offers advanced features and JavaScript execution capabilities. Its moderate memory usage and CPU utilization make it suitable for complex scraping tasks that require extensive data processing. BeautifulSoup and Jsoup, on the other hand, excelled in efficiency with low memory and CPU usage, enabling fast scraping operations for simpler tasks.

The presence of JavaScript execution played a significant role in resource requirements. HtmlUnit, which supports JavaScript interaction, exhibited moderate memory usage and higher CPU utilization. This makes it a valuable choice for projects involving dynamic web content, but may come with a trade-off in terms of resource consumption and working time.

The selection of a web scraping library should be based on the specific requirements of the project. Factors such as the complexity of scraping tasks, the availability of system resources, the desired scraping speed, and the need for JavaScript execution should be considered. By carefully evaluating these factors, developers can make informed decisions to optimize memory usage, CPU utilization, and working time for efficient and effective Web data extraction.

The study's contributions lie in providing a comprehensive understanding of the performance characteristics of these web scraping tools, enabling developers and researchers to make informed decisions when selecting the most suitable tool for their specific scraping needs. By evaluating a diverse set of tools and metrics, the research broadened the scope of the existing literature on web scraping performance analysis and identified strengths and weaknesses that can aid users in optimizing their scraping processes. Additionally, the paper highlights the importance of considering website complexity, scalability, and JavaScript handling when choosing an appropriate web scraping tool for different use cases.

In conclusion, our analysis provides valuable insight into the performance characteristics of different web scraping libraries. It highlights the importance of considering trade-offs between memory usage, CPU utilization, and working time when selecting a library. Ultimately, the choice should align with the project's requirements, balancing efficiency, speed, and resource constraints to achieve successful web scraping outcomes.

References

1. Acerado, R.: CMATA: cyber trafficking monitoring and tracking prototype. IJFCC **12**, 19–22 (2023). https://doi.org/10.18178/ijfcc.2023.12.1.598
2. Moro, S., Esmerado, J., Jalali, S.M.J.: Can we trace back hotel online reviews' characteristics using gamification features? Int. J. Inf. Manage. **44**, 88–95 (2019). https://doi.org/10.1016/j.ijinfomgt.2018.09.015
3. Han, S., Anderson, C.: Web scraping for hospitality research: overview, opportunities, and implications. Cornell Hosp. Q. **62**, 89–104 (2020). https://doi.org/10.1177/1938965520973587
4. Wooldridge, J., King, M.: Altmetric scores: an early indicator of research impact. J. Assoc. Inf. Sci. Technol. **70**, 271–282 (2018). https://doi.org/10.1002/asi.24122

5. Zucco, C., et al.: Sentiment analysis for mining texts and social networks data: methods and tools. WIREs Data Mining Knowl. Discov. **10**, e1333 (2019). https://doi.org/10.1002/widm.1333
6. Pellert, M., et al.: Dashboard of sentiment in Austrian social media during Covid-19. Front. Big Data **3**, 32 (2020). https://doi.org/10.3389/fdata.2020.00032
7. Alrusaini, O.: Deep learning models for the detection of Monkeypox skin lesion on digital skin images. IJACSA **14**, 637–644 (2023). https://doi.org/10.14569/ijacsa.2023.0140170
8. Eken, S., Sayar, A.: Performance evaluations of vector-raster satellite image transfers through web services. In: IEEE 36th Annual Computer Software and Applications Conference, pp. 346–347. IEEE (2012)
9. Eken, S., Sayar, A.: Vectorization and spatial query architecture on island satellite images. Procedia Comput. Sci. J. **2**, 37–43 (2012)
10. Eken, S., Aydin, E., Sayar, A.: Vectorization of large amounts of raster satellite images in a distributed architecture using HIPI. In: International Artificial Intelligence and Data Processing Symposium (IDAP), pp. 1–4. IEEE (2017)
11. Eken, S., Sayar, A.: Uydu Görüntülerinin Yüksek Performansta İşlenmesi Üzerine Bir İnceme: Vektör Tabanlı Mozaik Örme Durum Çalışması (2016)
12. Sayar, A.: Adaptive proxy map server for efficient vector spatial data rendering. J. Appl. Remote Sens. **7**(1), 073498 (2013)
13. Eken, S., Sayar, A.: Vector modelling of island satellite images for spatial databases. In: Proceedings of International Science and Technology Conference (ISTEC 2011), pp. 25–30 (2011)
14. Ozel, A., et al.: Web Servisler ile Paralel Görüntü İşleme Mimarisi: Raster İmgelerde Kenar Belirleme Uygulanması (2012)
15. Fox, G.C., et al.: Grids for real time data applications. In: Wyrzykowski, R., Dongarra, J., Meyer, N., Waśniewski, J. (eds.) PPAM 2005. LNCS, vol. 3911, pp. 320–332. Springer, Heidelberg (2006). https://doi.org/10.1007/11752578_39
16. Darmawan, I., et al.: Evaluating web scraping performance using XPath, CSS selector, regular expression, and HTML DOM with multiprocessing technical applications. JOIV Int. J. Inform. Vis. **6**, 904 (2022). https://doi.org/10.30630/joiv.6.4.1525
17. Islam, T., et al.: Can artificial intelligence detect Monkeypox from digital skin images? (2022). https://doi.org/10.1101/2022.08.08.503193
18. Landers, R., et al.: A primer on theory-driven web scraping: automatic extraction of big data from the internet for use in psychological research. Psychol. Methods **21**, 475–492 (2016). https://doi.org/10.1037/met0000081
19. Arhandi, P., Mashudi, I., Nugroho, F.: Automated website monitoring system using web scraping and Raspberry Pi. Telematika **18**, 222 (2021). https://doi.org/10.31315/telematika.v18i2.5506
20. Quinn, L., et al.: Explaining offenders' longitudinal product-specific target selection through changes in disposability, availability, and value: an open-source intelligence web-scraping approach. Crime Sci. **11**, 2 (2022). https://doi.org/10.1186/s40163-022-00164-1

Exploring Spreaders in a Retweet Network: A Case from the 2023 Kahramanmaraş Earthquake Sequence

Zeynep Adak[1](✉) and Ahmet Çetinkaya[2]

[1] İstanbul 29 Mayıs University, 34764 Ümraniye, Istanbul, Turkey
zadak@29mayis.edu.tr
[2] Marmara University, 34722 Kadıköy, Istanbul, Turkey
ahmet@marmara.edu.tr

Abstract. Two massive earthquakes struck Kahramanmaraş district of Türkiye on 6 February 2023, leaving loss of life and damage in a catastrophic scale. Many blamed the government for its inefficiency in dealing with the disaster. #devletyok (*there is no government*) was a hashtag used in the aftermath in social networking sites. We analyze the retweet network around the hashtag on 24th February, two weeks after the disaster, and aim to extract topological characteristics of the network, the influential spreaders in the network and the source of the diffusion. We make use of centrality measures, the HITS algorithm, PageRank algorithm and the k-shell decomposition in order to detect the influential spreaders. The social network analysis here is different from much of the previous research in that we explore the central roles in an information diffusion on a network, where all nodes are active, representing an already diffused information. In-degree centrality, betweenness centrality and HITS algorithm provide useful results in detecting spreaders in our network, while closeness centrality, PageRank and k-shell decomposition supply no additional knowledge. We figure out three nodes in the network with central roles in the diffusion, one being the source node. Checking the account of this source node reveals an anonymous user, who does not declare his/her identity. The study here has useful future implications for political and governmental studies. Moreover, the procedure applied to detect influential spreaders has many potential use cases in other fields such as marketing and sociology.

Keywords: Influential Spreaders · Kahramanmaraş Earthquake · Information Diffusion · Twitter · Retweet Network · Social Network Analysis

1 Introduction

Social networks model complex interactions between individuals and studies try to reveal the underlying mechanism behind important processes that take place among the actors in the network. Information diffusion in social networks have been researched in a large extent and several models are proposed to explain how information propagates in a network [1]. One crucial element of a diffusion process is the influential spreaders that take

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 481–492, 2024.
https://doi.org/10.1007/978-3-031-56728-5_40

key roles in the spread of information. Detecting these spreaders may help marketing professionals improve product promotion, sociologists understand and analyze the society, and government officials ensure safety and peace in a country. Centrality measures are one option that may reveal clues about critical actors in a network [2]. Authorities and hubs are also considered essential in an information diffusion process [3]. There are circumstances, however, where nodes with high centrality scores or hub nodes may have little effect in a given spreading process. Reference [4] proposed to utilize the k-core information to identify best individual spreaders when the spreading originates in a single node. Nodes in a network may also be ranked based on the well-known PageRank algorithm. Reference [5] used the algorithm to calculate the authority scores of nodes. Extensions of PageRank algorithm were developed later [6] to incorporate other characteristics of networks into influence determination process. Recently, influential spreaders in a network were seen to constitute part of a collective entity, a minimal set whose removal would dismantle the network in many disconnected and non-extensive components [7, 8]. Collective Influence Algorithm was proposed by [8] to capture this idea, and [7] showed an implementation of the algorithm in nearly linear time. Different from previous research, [9] emphasized the topological characteristics of a network, apart from any algorithm, to have significant role in figuring out the influential nodes. A good amount of recent research proposed to integrate multiple characteristics of nodes to detect vital nodes more reliably [28]. Reference [29], for example, incorporates the number of neighbors, the influence of neighbors, the location of nodes, and the path information between nodes into model to assess the importance of a node in diffusion process.

Retweet is an information diffusion mechanism allowed in Twitter. Three topics generally studied in retweet literature are predicting the spread in a retweeting case, factors affecting popularity of tweets, and reasons behind retweeting. Reference [10] studied factors effecting retweetability, and found URL and hashtag content, number of followers, and age of the account to predict retweetability. Reference [11] explored the topological characteristics of Twitter and the information diffusion by retweets. Apart from their findings about Twitter, they showed that retweets reach a large network of users no matter how many followers the source has, and half of this amount of retweeting occurs within an hour. Users retweet only from a few people and being retweeted by a small number of their followers. Reference [12] studied political hashtags and found that, contrary to [10], number of followers and followees has little effect on message replication, while user activity and mention network have considerable importance in message diffusion. In contrast to voluntary retweeting behaviour, [13] explored why people retweet at a stranger's request, and trustworthiness of the content, content relevance and value of the information in the tweet were three reasons they found. Reference [14] explored the effect of community structure in the retweeting process, and their study revealed that diffusions of retweets are typically trapped within a community, while inter-community diffusion would reach to a higher future popularity. Reference [15] investigated Twitter activity during an online political protest and found that a small group of very active users take the lead while the remaining large majority participates minimally.

On 6 February 2023, an earthquake of magnitude 7.8 struck southern and central Türkiye, and parts of Syria as well, followed, 9 h later, by a new 7.6-magnitude earthquake in the same region. The earthquake doublet is referred as the Kahramanmaraş earthquake series. The quake series generated more than 38,000 aftershocks, up to time of this writing, 8 of them being over magnitude 5.5. The earthquake doublet caused severe damage across 11 provinces in Türkiye, and according to Disaster and Emergency Management Authority (AFAD), it resulted in an official death toll of 50,096, while 107,204 people were injured. In this study, we analyze the retweet network around the #devletyok (*there is no government*) hashtag used in the aftermath of the Kahramanmaraş Earthquake sequence. It was used to claim that the state was not filling out its responsibility in the process of response and recovery. The hashtag is important since it conveys criticism towards the government response to the earthquake. We aim to extract topological characteristics of the retweet network, and detect influential spreaders in the network as it may provide helpful insights about the trustworthiness of the accounts spreading the notion. We apply centrality measures, PageRank, HITS and k-shell decomposition algorithms, as these are the fundamental algorithms in detecting influential spreaders. Different from the previous studies, we investigate a network representing an already diffused information. This approach is similar, in a way, to information source detection research [1], but different in that we detect not only the source of the information but also the influential spreaders in the network. Analysis of this kind of topic-based retweet networks would convey essential information to many professionals from fields as divergent as state governance and marketing.

2 Methodology

Tweets with #devletyok (*there is no government*) hashtag are collected and the retweeting interaction is represented as a social network. Extracted data covers tweets posted on February 24, 2023, two weeks after the disaster.

An edge in the network represents that node A retweeted node B, as shown in Fig. 1. Nodes are the users that had the hashtag in their tweets at that time.

A visualization of the hashtag network reveals two opposing groups, one in the favor of the government, and the other against (see the Results and Discussion section). Although, the meaning of the hashtag states an opposing position to the government, the hashtag was also used by government supporters to disprove the claims. While detecting the nodes with central roles in the spread, we carry out a comparative assessment of the diffusion in the two groups. Since the diffusion was much larger in the group against the government, we analyze the influential spreaders in that group, and aim to understand why the negative idea became spread at that extent.

Aggregate network metrics, such as number of connected components, network diameter and density, are computed to examine topological features of the network. In order to detect the central nodes in the diffusion, we consider fundamental approaches and algorithms. We calculate centrality measures, and use PageRank algorithm to locate influential spreaders in the network. We also use HITS algorithm to figure out authorities and hubs in the network. Lastly, we apply k-core decomposition algorithm for a potential further assessment of the impact of the nodes.

3 Results and Discussion

3.1 Aggregate Network Metrics

The network is directed and unweighted, and includes 2378 nodes, 2081 edges and 434 connected components. Connected components metric measures the number of isolated clusters in a network where a cluster has no link with the remaining part of the network. The high number here is due to the "middle region" group that are typical of social media interaction graphs [16]. These are small groups who interact with one another but not with the Giant Component of the network.

Network diameter (maximum geodesic distance) is 3 with an average distance of 1.04. That means two nodes are at a maximum of 3 hops and an average of 1 hop from each other. This indicates that most of the time retweeting nodes are not retweeted from someone else. Certainly, we are only dealing with the tweets having the hashtag in the body.

Graph density is virtually zero. Since the network edges are due to retweet activity, we do not expect to have a dense network. That is, a user would not retweet to much of other users having tweets with the same hashtag.

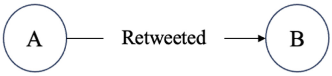


Fig. 1. Illustration of an edge in the network

A visualization of the network graph is given in Fig. 2. The arrangement of the graph is based on the ForceAtlas2 graph layout algorithm [17]. The label size as well as the coloring are based on the in-degree score of the node, see the following subsection for indegree distribution. Two separate groups, one in favor of the government -named group 1- and the other against – named group 2-, can be distinguished in the network. This fact is revealed after checking the tweets of the most retweeted nodes, those having larger labels in the figure. These groups can be observed better through a closer look in the network by omitting the singletons at the boundaries: see Fig. 3. Group 1 is clearly larger than Group 2, meaning that the opposing view is being spread to a wider audience. Nodes with central roles behind this spread is explored in the following section.

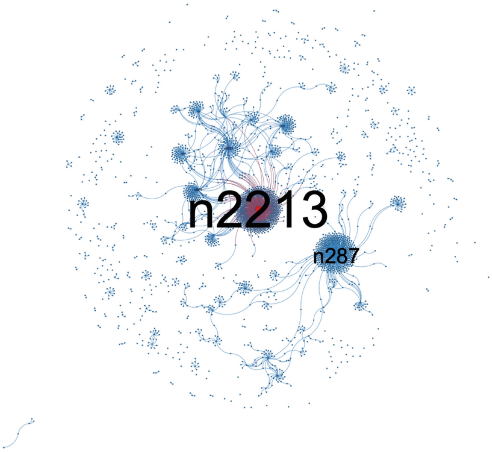


Fig. 2. Retweet network graph of #devletyok hashtag

3.2 Node-Specific Metrics

Several centrality measures are calculated to reveal out the important nodes in the network. Importance may be defined from many different aspects [2]. A node can be considered important if it has too much connections with the other nodes, or it can be important if it has a bridging role in the otherwise disconnected segments of the network. Moreover, having connections with central nodes may make a node itself important.

We calculate the in-degree centralities of the nodes. In-degree score is the number of incoming edges to a node. Indegree distribution of the nodes are represented in Fig. 4 together with Pareto Lognormal and Power Law models. Pareto lognormal model provides a better fit, a result previously proposed by [18]. The Pareto Lognormal shape implies that very few nodes has very high indegree values. In our retweet network, a node with high in-degree centrality indicates a user retweeted many times. That is, the spread of information initiates from these high-indegree-nodes, making them central in information diffusion process.

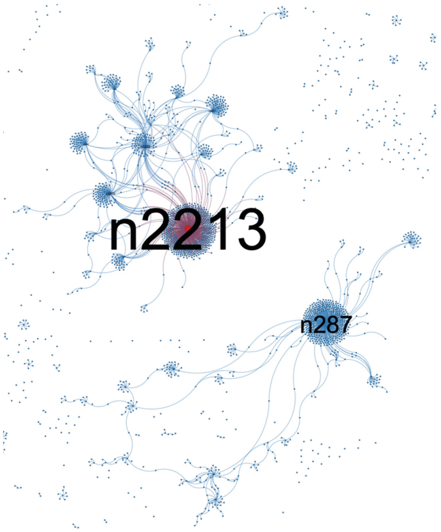


Fig. 3. Two groups with opposing views in the retweet network

The top 10 nodes with the highest indegrees are given in Table 1. The group information is also supplied in the table. n2213, the node with the highest indegree score, is part of group 1, and only two of the top 10 highest degree nodes are from Group 2. This is one reason why information spread is higher in the first group. Indegree distribution is visible in Fig. 3, where red color and label size indicates nodes with higher indegree. The high retweet rate of node n2213 can be associated with both its high number of followers, over 690 thousand, and the age of the account, over 7 years, [10]. What is interesting is that the account owner does not declare his/her identity even the name. How such an account can attract many followers? Homophily is the reason behind this behavior, as users are disproportionately exposed to like-minded information in social media [19]. Node n2213 is also the source node of the information diffusion process. This is revealed by a simple eye-check on the tweets posted by the node on the day of the data collection.

Betweenness centrality is another critical measure to detect nodes that contribute to the spread. It measures how many times a node lies on the shortest paths between other nodes. A node with high betweenness centrality takes a bridging role between otherwise separate parts of the network. The node with the highest betweenness score is node n1549, which is part of the first group, another contributing factor to the wider spread. The bridging role of the node is visible in Fig. 5. Edges are colored the same as the target node, and the text size is proportional to the betweenness score. Betweenness scores of some other several nodes are negligible, while the remaining have zero scores.

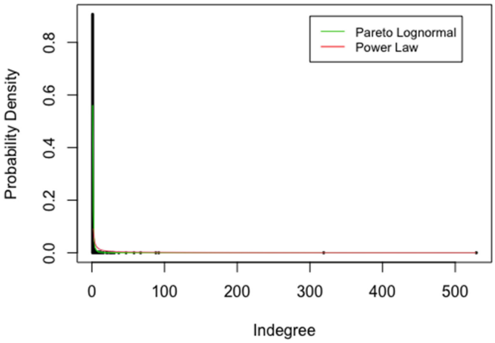


Fig. 4. Indegree distribution of the nodes in the network

Table 1. Top 10 in-degree centrality and group information

Node Id	In-degree	Group no
n2213	528	1
n287	318	2
n2222	91	1
n2210	87	1
n2217	66	1
n2238	57	1
n1549	46	1
n2211	36	1
n2223	30	1
n1335	28	2

Length of a path between two nodes in a network is the number of edges that needs to be traversed to reach from one to another. People with shorter paths to other individuals are expected to receive information rapidly in a network. Closeness centrality is a score to measure this aspect of nodes and it is considered as another factor in information diffusion [20]. It is calculated for every node as the inverse of average distance to all other nodes. In our directed retweet network, most nodes are connected only with a single node through a retweet activity, while the singletons on the edges of the network have no connection. This renders the closeness values to be either 1 or 0 most of the time, making closeness centrality measure not conveying significant information in our network.

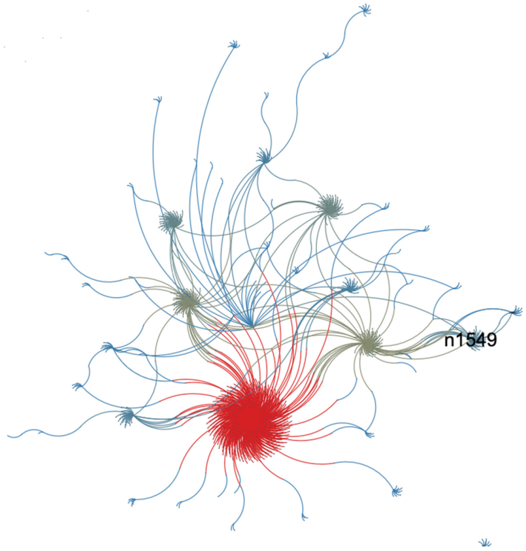


Fig. 5. Node with higher betweenness centrality

PageRank also was used to detect influential nodes in a social network [5]. Originally developed for web page ranking, PageRank Algorithm assigns an importance score to each node based on the number and quality of incoming edges. In the current retweet network, quality of nodes retweeting a node are not very much different, hence the PageRank ranking of nodes becomes similar to in-degree ranking of them. Thus, no additional information is supplied with PageRank scores.

Another aspect we analyze in the network is the existence of authorities and hubs. Hubs and authorities are first defined for World Wide Web networks [3] and the concept is later applied to many other types of networks including online social networks [21–23]. Authorities are defined as the leading sources of information, while hubs are nodes with many connections. Hubs are regarded as the key players who are responsible from the largest scale of the spreading process [24]. Good hubs are nodes that point to many good authorities, and good authorities are nodes pointed by numerous good hubs. The relationship is a mutually reinforcing one, and reveals critical roles in information dissemination process. We identify authorities and hubs in our retweet network using the HITS algorithm [25]. There is a single node, n2213, with very high authority score, while the scores of the other nodes are nearly zero. n2213 is also the node at the center of group 1 with the highest indegree score, see Fig. 3. The top ten highest hub scores are given in Table 2. The node with the highest hub score, n398, and all other nodes with relatively high hub scores are part of group 1. Indeed, there are no hub nodes in group 2. Figure 6 illustrates the hub score distribution, where the nodes with red colors and larger in size are hub nodes. It is clearly visible that the core of the network in group 1 is surrounded with hub nodes that facilitate the distribution process. Hubs are known to be good spreaders when they are not part of the periphery of the network [4]. The extensive presence and central location of hub nodes in group 1 is another factor leading to a wider spread of information in group 1.

Table 2. Top 10 hub score distribution

Node Id	Hub score
n398	0.048999
n567	0.045837
n856	0.045766
n228	0.045702
n2045	0.045381
n951	0.045378
n286	0.044756
n965	0.044756
n1037	0.044756
n1198	0.044756

In their seminal work, [4] separated the core and the periphery of the network to assess the impact of a node. They showed that when spreading starts from a single node, nodes with a high coreness value are the best spreaders. Coreness of a node represents the k-shell it belongs to. k-shell consists of nodes belonging to k-core and not to (k + 1)-core, while k-core is the maximal subgraph where every node has at least k number of connections to other nodes in the subgraph [26]. Coreness values of nodes are calculated using the k-core decomposition algorithm [27]. In the current retweet network each node has one or two outgoing edges, while some few nodes have many incoming edges. This special feature of the retweet network renders coreness values of either 0 or 1, not identifying a distinguishing factor in the dissemination process.

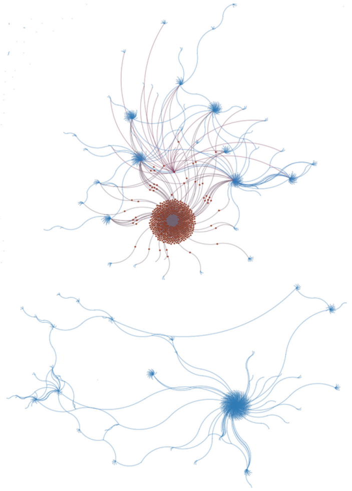


Fig. 6. Hub nodes in Group 1

4 Conclusion and Future Work

Retweeting is an important means of information dissemination in Twitter, one of the largest social networking sites. Following the catastrophic earthquake sequence in Kahramanmaraş district of Türkiye on 6th February 2023, #devletyok (*there is no government*) hashtag was used to claim that the government was not filling out its responsibility in the process of response and recovery. We have explored the retweet network around the hashtag on the 24th February, two weeks after the disaster. We aimed to explore topological characteristics of the network, identify influential spreaders and detect the source node. The network has shown to consist of two opposing groups, one in the favor of the government and the other against. The information spread in the disfavoring group was wider. The reasons behind the intensive spread has been explored by examining influential spreaders in the network.

In-degree centrality and betweenness centrality measures have revealed two central nodes in the diffusion, one being the source node. The source has been discovered to be an anonymous user, raising concerns about the trustworthiness of the information content. Closeness centrality has provided no useful knowledge in the current network. Authorities and hubs have been detected using the HITS algorithm, and we have shown these to possess critical roles in the spread. Besides many hubs surrounding an authority node in the opposing group, we have identified one node with higher hub score compared to others. PageRank and k-shell decomposition algorithms have also been applied to discover any additional knowledge about the diffusion process. However, because of the special structure of the retweet network, these algorithms have not supplied useful information.

Our study here has potential implications for political and governmental studies. Practical results are also immediate. Use of social network analysis to detect the source and the influential spreaders in a retweet network proves a convenient and practical way for officials in case of a widespread public confusion. The approach is also beneficial when applied in other contexts such as screening the activity and the key users around a brand hashtag in marketing, or exploring the spreading patterns of a rumor-related hashtag among a community.

As a future research, it would be interesting to investigate the spreading process in a follower or friendship network by activating the nodes found as spreaders in the retweet network, and applying an information diffusion model such as Linear Threshold or Independent Cascade. This way of information diffusion analysis would require integration of two or three interdependent networks, which may yield promising results in many research areas such as influence maximization or minimizing spread of misinformation.

References

1. Chang, B., Xu, T., Liu, Q., Chen, E.-H.: Study on information diffusion analysis in social networks and its applications. Int. J. Autom. Comput. **15**(4), 377–401 (2018). https://doi.org/10.1007/s11633-018-1124-0
2. Hansen, D.L., Shneiderman, B., Smith, M.A., Himelboim, I.: Social network analysis: measuring, mapping, and modeling collections of connections. In: Hansen, D.L., Shneiderman, B., Smith, M.A., Himelboim, I. (eds.) Analyzing Social Media Networks with NodeXL, 2nd edn, ch. 3, pp. 31–51. Morgan Kaufmann (2020)
3. Kleinberg, J.M.: Hubs, authorities, and communities. ACM Comput. Surv. **31**(4es), (1999)
4. Kitsak, M., et al.: Identification of influential spreaders in complex networks. Nat. Phys. **6**(11), 888–893 (2010). https://doi.org/10.1038/nphys1746
5. Cataldi, M., Di Caro, L., Schifanella, C.: Emerging topic detection on Twitter based on temporal and social terms evaluation. In: MDMKDD 2010, Washington, DC, USA, 25 July 2010. ACM (2010). https://doi.org/10.1145/1814245.1814249
6. Al-Garadi, M.A., et al.: Analysis of online social network connections for identification of influential users. ACM Comput. Surv. **51**(1), 1–37 (2019). https://doi.org/10.1145/3155897
7. Morone, F., Min, B., Bo, L., Mari, R., Makse, H.A.: Collective influence algorithm to find influencers via optimal percolation in massively large social media. Sci. Rep. **6**(1), 30062 (2016). https://doi.org/10.1038/srep30062
8. Morone, F., Makse, H.A.: Influence maximization in complex networks through optimal percolation. Nature **524**(7563), 65–68 (2015). https://doi.org/10.1038/nature14604
9. Al-Garadi, M.A., Varathan, K.D., Ravana, S.D., Ahmed, E., Chang, V.: Identifying the influential spreaders in multilayer interactions of online social networks. J. Intell. Fuzzy Syst. **31**(5), 2721–2735 (2016). https://doi.org/10.3233/jifs-169112
10. Suh, B., Hong, L., Pirolli, P., Chi, E.H.: Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. In: IEEE Second International Conference on Social Computing. IEEE (2010). https://doi.org/10.1109/socialcom.2010.33
11. Kwak, H., Lee, C., Park, H., Moon, S.: What is Twitter, a social network or a news media?. ACM (2010). https://doi.org/10.1145/1772690.1772751. https://doi.org/10.1145/1772690.1772751
12. Bastos, M.T., Raimundo, R.L.G., Travitzki, R.: Gatekeeping Twitter: message diffusion in political hashtags. Media Cult. Soc. **35**(2), 260–270 (2013). https://doi.org/10.1177/0163443712467594
13. Lee, K., Mahmud, J., Chen, J., Zhou, M., Nichols, J.: Who will retweet this?. ACM (2014). https://doi.org/10.1145/2557500.2557502

14. Tsugawa, S.: Empirical analysis of the relation between community structure and cascading retweet diffusion. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 13, pp. 493–504 (2019).https://doi.org/10.1609/icwsm.v13i01.3247
15. Morales, A.J., Borondo, J., Losada, J.C., Benito, R.M.: Efficiency of human activity on information spreading on Twitter. Soc. Netw. **39**, 1–11 (2014). https://doi.org/10.1016/j.socnet.2014.03.007
16. Kumar, R., Novak, J., Tomkins, A.: Structure and evolution of online social networks. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 20 August 2006. ACM (2006). https://doi.org/10.1145/1150402.1150476
17. Jacomy, M., Venturini, T., Heymann, S., Bastian, M.: ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PLoS ONE **9**(6), (2014)
18. Sala, A., Zheng, H., Zhao, B.Y., Gaito, S., Rossi, G.P.: Brief announcement. In: Proceedings of the 29th ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing, 25 July 2010. ACM (2010). https://doi.org/10.1145/1835698.1835791
19. Halberstam, Y., Knight, B.: Homophily, group size, and the diffusion of political information in social networks: evidence from Twitter. J. Public Econ. **143**, 73–88 (2016)
20. Mochalova, A., Nanopoulos, A.: On the role of centrality in information diffusion in social networks. In: ECIS 2013 Completed Research (2013)
21. Java, A., Song, X., Finin, T., Tseng, B.: Why we twitter. In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis, 12 August 2007. ACM (2007). https://doi.org/10.1145/1348549.1348556
22. Deborah, A., Michela, A., Anna, C.: How to quantify social media influencers: an empirical application at the Teatro alla Scala. Heliyon **5**(5), e01677 (2019). https://doi.org/10.1016/j.heliyon.2019.e01677
23. Agichtein, E., Castillo, C., Donato, D., Gionis, A., Mishne, G.: Finding high-quality content in social media. Presented at the Proceedings of the 2008 International Conference on Web Search and Data Mining, Palo Alto, California, USA (2008)
24. Boster, F.J., Kotowski, M.R., Andrews, K.R., Serota, K.: Identifying influence: development and validation of the connectivity, persuasiveness, and maven scales. J. Commun. **61**(1), 178–196 (2011). https://doi.org/10.1111/j.1460-2466.2010.01531.x
25. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. J. ACM **46**(5), 604–632 (1999). https://doi.org/10.1145/324133.324140
26. Kong, Y.-X., Shi, G.-Y., Wu, R.-J., Zhang, Y.-C.: k-core: theories and applications. Phys. Rep. **832**, 1–32 (2019)
27. Vladimir, B., Matjaz, Z.: An O (m) algorithm for cores decomposition of networks. arXiv preprint arXiv:cs/0310049 (2003)
28. Lü, L., Chen, D., Ren, X.-L., Zhang, Q.-M., Zhang, Y.-C., Zhou, T.: Vital nodes identification in complex networks. Phys. Rep. **650**, 1–63 (2016). https://doi.org/10.1016/j.physrep.2016.06.007
29. Li, Z., Huang, X.: Identifying influential spreaders by gravity model considering multi-characteristics of nodes. Sci. Rep. **12**(1), (2022). https://doi.org/10.1038/s41598-022-14005-3

Verifying the Facial Kinship Evidence to Assist Forensic Investigation Based on Deep Neural Networks

Ruaa Kadhim Khalaf[1](✉) and Noor D. Al-Shakarchy[2]

[1] University of Kerbala, Kerbala, Iraq
ruaa.k@uokerbala.edu.iq
[2] College of Computer Science and Information Technology, Kerbala, Iraq
noor.d@uokerbala.edu.iq

Abstract. The criminal incident evidence can be considered as the substance of the search for the crime perpetrator. Even if the offender received the deserved punishment. However, when the requisite time for catching the culprit was shorter, the society's confidence of security and justice will be higher. Thus, justice agencies should custom any new technology which contributes to this process as soon as possible. Kinship Verification can be regarded an interesting and difficult area of study in computer vision and computing forensics. Facial Kinship Verification has the capability to predicts whether two people are related in kinship or not depending on the facial images or videos. Facial Kinship Verification has a diversity of real-world practices, including forensic investigations, contributing to the resolution of missing person cases, social media analysis, and genealogy research. The proposed approach involves the Verification of the relationship which exists between the provided facial images using a Three-Dimensional Convolution Neural Network. This approach involves of following stages: face preprocessing, deep features extraction and Classification. Extensive experiments revealed promising results compared with many state-of-the-art approaches. The accuracy of proposed system reached to 89.25% in KinFaceW-I dataset.

Keywords: Facial Kinship Verification · Three-Dimensional Convolution Neural Network · Deep learning

1 Introduction

Forensic science entails using natural sciences such as computer, biology, chemistry, physics, and humanities such as psychology and sociology to collect and analyze evidence left inside and outside the crime scene, The results of these analyses are then used to describe the perpetrator, the victim, and the crime, which can be used in court to convict or clear the accused.

Facial image analysis is now a core research area of image processing, computer vision, and pattern recognition. This is because the human face contains huge social data including gender, age, and emotional state, in addition to identifying characteristics

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 493–504, 2024.
https://doi.org/10.1007/978-3-031-56728-5_41

that may be used to ascertain an individual's identity. In recent years, face recognition, facial expression recognition, and age estimate have been the topic of intensive research.

Kinship verification is a process of determining the biological relatedness or familial relationships between individuals. These relationships may be "Parent_Child", "Sibling_Sibling", "Grandparent_Child", etc., as seen in Fig. 1. More than fifty percent of a parent's genes are passed on to their offspring. As genes overlap, children inherit various characteristics from their parents, including likeness in look, behavior, and voice [1]. Face Kinship Verification uses face images or videos to identify automatically whether two individuals are related [2].

While Deoxyribonucleic Acid (DNA) tests are a valuable tool for kinship verification, they do have certain limitations such as unavailable or inaccessible reference samples and ethical and privacy considerations. These limitations make DNA tests unfeasible solutions in some life scenarios related to forensic applications and video surveillance. In addition, DNA needs many days to process, Hence, its use is limited in scenarios necessitating real-time processing or involving difficult people. Owing to the increasing expansion of multimedia, facial kinship verification has a substantial impact on a number of fields [3].

Human sensory perception is inadequate for detecting similarities between photos taken by various people [4]. The difficulties of efficiently recognizing face characteristics including the size, shape, and color of facial components cause low accuracy rates. The most challenging and important phase in the verification process, as well as the system's core, is feature extraction, because the salient features made accessible for recognition have a powerful effect on the precision of the kinship verification and recognition tasks [5].

Fundamentally, there are two types of challenges to acknowledging kinship that can impact the precision of verifying facial kinship: directly challenging (associated to the kinship itself), which includes variations in gender, age, and feature likeness among relatives, and indirectly challenging (related to the database's environment), which can include lighting, noise, occlusion, facial expressions, blue, position variation, clutter, and lower picture quality [6].

Deep learning techniques, such as deep neural networks, are powerful AI methods that have shown promise in various fields, including kinship verification. These models are capable of learning complex representations and patterns from genetic data, enabling more accurate predictions, which outperform various shallow techniques, and obtained quality on important visual recognition functions [5]. Deep learning produces more informative representations for classification problems, leading to increased accuracy.

The facial kinship verification method is made up of a number of stages, each of which has a number of steps that serve various purposes. The preprocessing stage (which covers all aspects of image preprocessing), classification predictor stage (which covers feature extraction, feature selection, and other activities that might result in salient features, and classification task) [1].

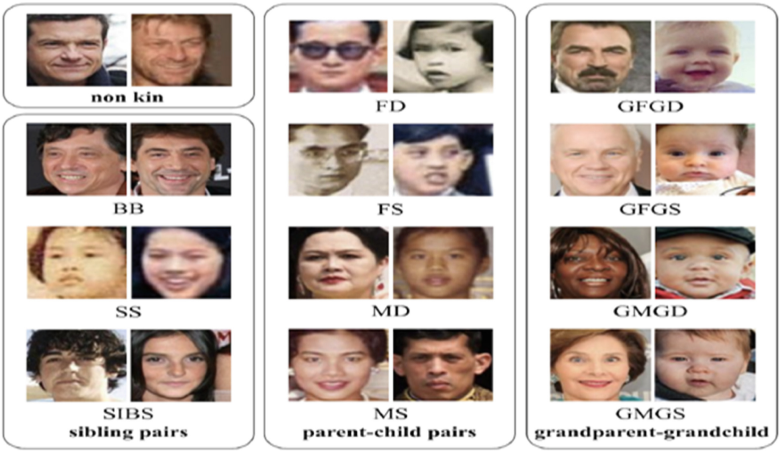


Fig. 1. Common examples of kin relationships

2 Related Work

Facial kinship verification is active area of research, with many studies focusing on the most effective techniques for extracting distinguishing characteristics used in existing kinship verification methods.

Qingyan Duan et al. [7] in 2017 used a deep transfer model called Coarse-to-Fine Transfer (CFT) to apply subspace transfer learning on some datasets (KinFaceW-I, KinFaceW-II, Cornell KinFace, and UB KinFace databases). CFT is built on a deep CNN architecture and consists of a Coarse CNN (cCNN) and a Fine CNN (fCNN). The cCNN uses a pre-trained mechanism on a large-scale facial picture database for face recognition. The fCNN model is then retrained using a multi-class learning rule on a small-scale kinship database based on cCNN (two kinship images per class). cCNN is utilized for generalized face features in each facial picture, whereas fCNN is used for kin-relation specific characteristic features. CFT we are going to study the learning problem of small-scale data by utilizing data from another large-scale area. Extensive experiments show that the proposed CFT performs as well as or better than Modern approaches to kinship verification.

Xiaoting Wu et al. [8] in (2019) proposed a Similarity Metric-based Convolution Neural Network (SMCNN) technique on the KinFaceW-I dataset and KinFaceW-II dataset to verify kinship. The SMCNN structure utilizes two identical Convolution Neural Networks, each with 8 layers. The L1 norm between two CNN outputs was calculated, and a decision was made using a learned threshold. The superior results are obtained with KinFaceW-II due to the cropped image sharing a comparable environment, such as chrominance and brightness. The verification accuracies were 72.7% in the KinFaceW-I, 79.25% in the KinFaceW-II database.

Chergui et al. [9] 2019 use (ResNet) for the feature extraction stage, in addition to our suggested pair feature extraction function and Rank Features (T test) to decrease the number of features through feature selection, and then uses the support vector machine to verify kinship decision. The verification accuracy was 87.16% on the Cornell Kin

Face, 83.68% on the UBKin, 82.07% on the Familly101, 79.76% on the KinFaceW-I, and 76.89% on KinFaceW-II datasets respectively.

Chergui et al. [10] in 2019 offered a technique for extracting features that are based on combining several descriptors (Local Binary Patterns (LBP), Local Phase Quantization (LPQ), and Binarized Statistical Image Feature (BSIF)). The Multi-Block (MB) representation approach was used, and the effect of alternative feature representations for verifying kinship was examined to minimize the number of features selected using the TTest function. For kinship classification, a Support Vector Machine (SVM) was used. This technique achieves kinship verification accuracy of 84.74% On Cornell KinFace, 82.74% on UBKin, 81.69% on KinFace-I, 80.12% on KinFace-II, and 78.16% on Familly 101, respectively.

Nandy and Mondal [11] in 2019 proposed a Deep learning technique using Siamese Convolutional Neural Network Architecture for facial Kinship Verification. And combine the two networks into a single output using a similarity metric and fully connected networks. Several similarity metrics were employed, including L1 norm, L2 norm, and Cosine Similarity, but the cosine similarity metrics outperforms than L1 and L2 metrics concerning accuracy because of effective and simple learning of the objective function. This network gives good accuracy. The verification accuracy was 67.66% on the FIW datasets.

Yan and Wang [12] in (2019) use an attention network for facial Kinship Verification in 2019, attention network is designed to extract information about the local parts and guide the learning by adding a mask to five facial feature portions of each face to assist the network in focusing on extracting more discriminative information in these areas. The attention network performs well on KinFaceW-I and KinFaceW-II. They outperformed basic CNN with 82.6% and 92.0% accuracy.

Van and Hoang [13] in 2019 uses Local Binary Pattern (LBP) for Kinship on different color space. Then, they calculated features based on (Chi-Square) distance and applied the Fisher score to discover important features. A Support Vector Machine is utilized for model training and prediction. The accuracy for the KinFaceW-I and the KinFaceW-II databases was 72.6% and 81.8%, respectively.

Zhang et al. [14] in 2019 uses a Deep learning method to verify facial kinship. Using shape and appearance complementary information. Both are necessary when determining kinship from face photos. To train this model with limited Kinship data, the researchers used an adaptation-based two-phase training approach using large-scale face recognition data, with the verification accuracy (78.3%) on the KinFaceW-I dataset.

Goyal & Meenpal [15] in 2019 used two descriptors (Local Binary Pattern (LBP) and Histogram of Gradient (HOG)) to identify salient features. Then, used a Support Vector Machine classifier to obtain an understanding of the retrieved face characteristics. The results showed that the (LBP_SVM) technique performed better than the (HOG_SVM) technique. On the KinFaceW-I dataset, the LBP_SVM approach's mean accuracy was 75.57%. On the KinFaceW-I dataset, the HOG_SVM technique had an average accuracy of 73.35%.

Mukherjee & Meenpal [16] in (2019) provided a method to increase the accuracy of verifying kinship that relies on a compound Local Binary Pattern (CLBP) and local feature-based discriminate analysis (LFDA). Long feature vectors were generated using these two techniques. The only methodology that accelerated the process and selected the best features was the entire feature vector-based LFDA feature selection method. A KNN classifier was used. The accuracy of verification was 82.82 on the KinFaceW-I, and 89.36 on the KinFaceW-II datasets, respectively.

Chergui et al. [17] proposed a strategy depend on examining two images to determine kinship in 2020. The deep features are extracted using the VGG-face model after the face preprocessing stage. Then, using Fisher Score (FS), feature pairs are represented and normalized to determine the salient features. Support Vector Machine (SVM) is used to make the final decision in kinship verification (classification stage). The accuracy was 92.89% on Cornell KinFace, 90.59% on UB KinFace, 86.65% on KinFace W-I, 81.11% on KinFace W-II, and 84.82% on Familly 101, respectively.

Felipe Crispim et al. [18] provided a technique of kinship verification using RGB-D Face Data in 2020. First, a database including 3D information and kinship annotations was collected, and depth information from normalized 3D reconstructions was combined with 2D images to create RGBD data. Then, employ a Convolutional Neural Network and SVM for classification and comparison. The model was evaluated on two a commonly used 2D database to verifying kinship (KinFaceW-I and KinFaceW-II) and the Kin3D dataset. The results suggest that adding depth information enhances the performance of the system, increasing verification accuracy to 90%.

Zhou et al. [19] use a deep learning network model for facial kinship verification in 2020. The network's first half computes the input image pairs' feature vectors using two or more channels, and its second half determines the distance between various feature vectors before outputting the result based on the distance value. In the KinFaceW datasets, the verification accuracy was 91%, while in the TSKinFace datasets, it was around 89.5%.

Wu et al. [20] in 2021 presented a technique for localizing multiple facial feature points by employing a facial feature detector to extract SIFT descriptors around each facial feature point in a face picture. In conclusion, Two methods, feature combination and distance metric learning, are employed to verifying kinship between two images, the verification accuracy was 73.8% on the kinFaceW-I and 78.23% on the kinFaceW-II datasets respectively.

Zekrini et al. [21] presented a technique to verify kinship depend on the combination of two descriptors in 2022. The Gradient Local Binary Patterns (GLBP) is first descriptor, which links gradient and textural information. The Histogram of Templates (HOT) is a shape descriptor. These features are used to define kinship ties in face photos, and SVM is employed to classify kinship, their verification accuracy was 76.99 on the KinfaceW-II and 90.49 on the Cornell datasets, respectively.

Liu et al. [22] in 2022 used an Age-Invariant Adversarial Feature Learning method to verifying the kinship, Which include two modules: an Identity Feature.

Weighted module (IFW) and an Age-Invariant Adversarial Feature learning module (AIAF). Their verification accuracy was 85.08% on the KinFaceW-I, 85.25 on the KinFaceW-II, and 76.80 FIW datasets respectively.

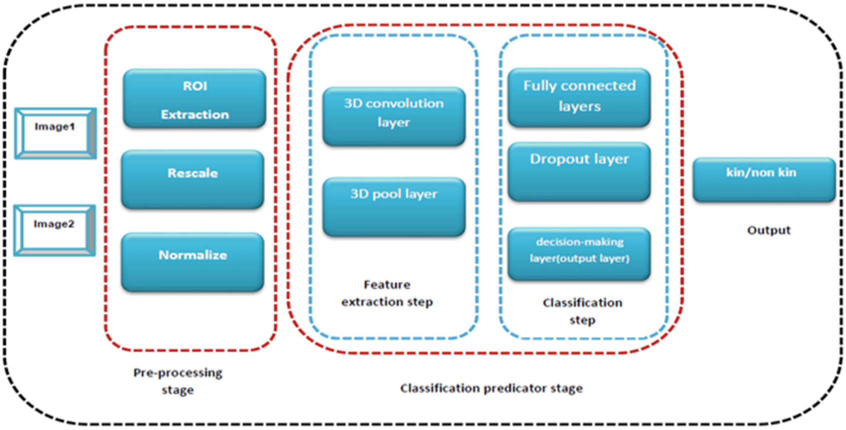


Fig. 2. Block Diagram of the proposed system

3 Research Method

This paper proposes a Three-Dimensional Convolution Neural Network (3D CNN) model with a new architecture for facial kinship verification. The proposed model employs multiple face images as inputs and is designed to learn their shared features. This model consists of two major phases: the pre-processing phase and the classification predictor phase. The stages of the proposed system make use the KinFaceW-I dataset.

3.1 Dataset

The KinFaceW-I dataset was created by Lu et al. The images contained in this dataset were found on the Internet and were taken in situations that were not controlled in terms of motions, demographics, lighting, backgrounds, faces, and partial occlusions. This database has 116 (M–S), 134 (F–D), 127 (M–D), and 156 (F–S) kinship pairs. The pictures in this data were aligned and edited by hand.

3.2 Proposed System Stages

The model that has been suggested is composed of two major stages, as illustrated in Fig. 2: The preprocessing stage and the classification predictor phase. The data set was separated into two training datasets (85% of the overall database) and the testing dataset (15%), was utilized for testing the model on new data. The training data was split into 85% real training data and 15% validation data.

- Pre-processing stage

The first stage of the system is preparing the input images in order to provide generalization by extracting ROI, scaling, and normalizing.

A region of interest (often abbreviated ROI): A part of an image that represents the limits of the object under study. Based on the features used by the suggested method (facial features), the face area is the most interesting area implemented with the suggested method. This step implemented Multi-Task cascaded Convolutional Neural Networks (MTCNN) to detect accurate faces and then extracted them.

Resize (scale): Image scaling means resizing an image, which involves rebuilding it from one pixel grid to another, this is done in this step by decreasing or increasing the sum of all the pixels contained in an image, which brings it to (64 × 64). It is a more important step to make sure that the results are valid for all data and that the deep learning model can use them.

Normalization: By dividing all pixel values by 255, the process modifies the range of pixel intensity values so that each pixel value has a value range between 0 and 1.

- Classification predictor stage

Using a 3D CNN model, the next stage implements a Classification predictor for verifying facial kinship. This stage includes two major steps (feature extraction step and then kinship verification step), both of them is constructed up of multiple layers that perform diverse functions depending on the goal of each layer. These layers are 3D convolutional, non-linear, dropout, pooling, and fully connected layers. The saved weights are used to compute the output of each layer and fed to the next layer until the decision-making (model's output) in the last layer in order to determine whether the two input images have kinship or not. Table 1 provides an overview of the proposed model architecture, as well as the output shape and parameters for each layer.

The 3D convolutional layers employ 3D kernels (filters) on two face images to extract the related salient features in these two images and determine the kin-feature maps. These filters; which represent the layer's depth; have sizes (32, 64, and 128) respectively for the three Convolution layers with strides (3, 3, and 3). The kernel coefficient values; which represent the weights; were set during the training process.

A "Rectified Linear Units (ReLUs)" function is used on all non-linear Layers except the final non-linear layer which used the "sigmoid" function on the output layer. The ReLUs allow the model to learn and represent more complex relationships between inputs and outputs by strengthening strong features and weakening weak ones.

The Pooling layers provide resilience against noise by decreasing the resolution of the features by passing a single neuron with maximum value in one layer from the clustered of several neurons in the previous layer using a max-pool function of clustered neurons.

Finally, the Dense layers (fully connected layers) Flatten the output of preceding layers and classify a sample of face images by using the activation function. The decision-making layer or output layer, which employs a sigmoid function, is the final fully connected layer.

Dropout layers are included in the proposed system to avoid overfitting and making generalizations on unseen data. During the training process, it chooses 50% of the neurons at random and sets their weights to zero. It is an easy way to reduce the model's sensitivity to noise while it is being trained, while keeping the required level of complexity for the architecture of the proposed model.

Table 1. Proposed system summary

Layer (type)	Output Shape	Param #
conv3d (Conv3D)	(None, 2, 64, 64, 32)	2624
batch_normalization (BatchNormalization)	(None, 2, 64, 64, 32)	128
max_pooling3d (MaxPooling3D)	(None, 2, 32, 32, 32)	0
dropout (Dropout)	(None, 2, 32, 32, 32)	0
conv3d_1 (Conv3D)	(None, 2, 32, 32, 64)	55360
batch_normalization_1 (BatchNormalization)	(None, 2, 32, 32, 64)	256
max_pooling3d_1 (MaxPooling3D)	(None, 2, 16, 16, 64)	0
dropout_1 (Dropout)	(None, 2, 16, 16, 64)	0
conv3d_2 (Conv3D)	(None, 2, 16, 16, 128)	221312
batch_normalization_2 (BatchNormalization)	(None, 2, 16, 16, 128)	512
max_pooling3d_2 (MaxPooling3D)	(None, 1, 8, 8, 128)	0
dropout_2 (Dropout)	(None, 1, 8, 8, 128)	0
flatten (Flatten)	(None, 8192)	0
dense (Dense)	(None, 300)	2457900
batch_normalization_3 (BatchNormalization)	(None, 300)	1200
dropout_3 (Dropout)	(None, 300)	0
dense_1 (Dense)	(None, 1)	301
Total params: 2,739,593 Trainable params: 2,738,545 Non-trainable params: 1,048		

3.3 Experiments and Results

The loss, accuracy, and mean square error functions are employed as metrics in the proposed model. These functions are used throughout the whole training and validation phases. The suggested model is implemented in two phases: training stage and then testing stage. In the First phase of the proposed system is trained using all of the training data provided. During the training phase, each layer's weights are updated until the network converges on the lowest error, as measured by the mean square error. Following the model's stability, the validation process is implemented on validation datasets with Labels and utilized the kept weights from the training phase to evaluate performance of model and Accuracy and decide if it is ready for predicting the label of unseen dataset.

The second phase illustrates system's application which predicts the class (output) of unseen data using trained proposed model. Figure 3 illustrates the proposed model's learning behavior on training and validate datasets in all epochs by showing the results of metrics employed in all epoch. Table 2 illustrates how the hyperparameter values are utilized in the proposed model. Use 10-fold cross-validation, As seen in Table 3.

Table 2. The utilized 3D CNN Model's hyper-parameter values

hyper-parameter	values
Learning Rate	0.001
Batch Size	32
Metrics	Accuracy
Loss Function	binary_crossentropy
Optimizer	Adam

3.4 Evaluation Model

In evaluation phase, the suggested model is evaluated by employing the test data. Using Performance Metrics like Precision, Accuracy, F1-Measure, confusion matrix, and Recall, the proposed model is evaluated. Table 4 shows the system's performance metrics to verify the facial kinship. Table 5. The comparative with other state of the art methods in KinFaceW-I.

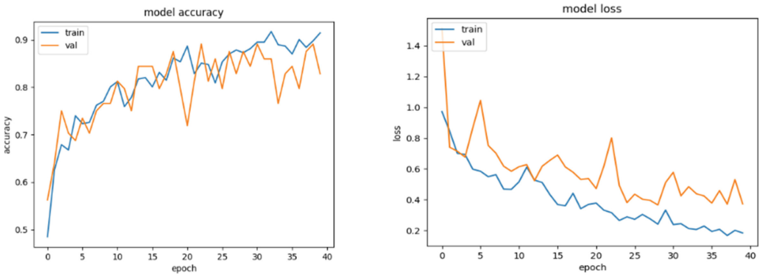


Fig. 3. The Learning Curve of proposed model (Accuracy and Loss function)

Table 3. The 10-fold cross-validation of proposed system in KinFaceW-I

Accuracy of each fold											model accuracy
Kinship	1	2	3	4	5	6	7	8	9	10	
F-S	68.75	65.62	77.41	93.54	87.09	93.54	100.0	93.54	93.54	100.0	87.31% (± 11.79%)
F-D	70.25%	74.07%	62.96%	66.66%	92.59%	88.88%	95.29%	92.59%	90.0%	100.0%	83.33% (± 15.11%)
M-S	70.23%	80.76%	61.53%	87.46%	83.99%	72.00%	100.0%	92.00%	83.99%	95.99%	82.80% (± 11.60%)
M-D	65.38	88.46	73.07	96.15	100.0	83.99	95.99	92.00	100.0	92.00	88.71% (± 10.94%)

Table 4. The Evaluation Measures Values

Measure	F-S	F-D	M-S	M-D
Precision	87.5%	85.5%	92.5%	95%
Recall	84.5%	85%	89%	95%
F1-Score	85.5%	85.5%	91.5%	95%

Table 5. Comparing the proposed system to state-of-the-art approaches in KinFaceW-I

Method	F_S	F_D	M_S	M_D	Average
HOG,LBP + SVM [15]	75.6	77.8	73.3	75.6	75.57
AdvKin [23]	75.70	78.30	77.60	83.10	78.68
AIAF + IFW [22]	88.70	80.80	82.60	88.20	85.08
3DCNN	87%	85%	91%	94%	89.25%

4 Conclusion

This paper presents a proposed system to performs the verifies facial kinship model based on determining feature maps. Regarding the accuracy and loss functions, the deep neural network-based verification system is considered to be better than other traditional methods. As demonstrated in the results above, the classification system based on 3DCNN deep neural networks is the best method to get high accuracy and providing superior results compared to other traditional methods concerning the accuracy and loss functions. The proposed model can handle various illumination conditions. The ReLU activation function is removing all the black elements from the two images at the same time and keeps only those carry a positive value which leads to extracting salient related features in the sample and neglecting the weak features which leads to dealing with different lighting conditions. Reduction in the memory requirements as well as the computation complexity requirements are presented by using the same coefficients across all images in a sample. Face detection is the most challenging part of the effectiveness of the proposed model, all analysis and examination works of the sample are totally based on it (face images). So, facial detection based on MTCNN can find faces in pictures with variables that can't be controlled, such as lighting that isn't uniform, position variation, face rotation, etc., which other methods, like Haarcascade, can't handle.

In the future, the model that was proposed can be developed to determine the level or degree of kinship of the input images, and then the Map Reduce concept can be used to attempt to apply the model in real time.

References

1. Nader, N., El-Gamal, F.E.Z., El-Sappagh, S., Kwak, K.S., Elmogy, M.: Kinship verification and recognition based on handcrafted and deep learning feature-based techniques. PeerJ Comput. Sci. **7** (2021). https://doi.org/10.7717/PEERJ-CS.735
2. Wu, X., et al.: Facial kinship verification: a comprehensive review and outlook. Int. J. Comput. Vis. **130**, 1494–1525 (2022). https://doi.org/10.1007/s11263-022-01605-9
3. Kohli, N.: Automatic kinship verification in unconstrained faces using deep learning. Lane Department of Computer Science and Electrical Engineering, West Virginia University (2019)
4. Bordallo Lopez, M., Hadid, A., Boutellaa, E., Goncalves, J., Kostakos, V., Hosio, S.: Kinship verification from facial images and videos: human versus machine. Mach. Vis. Appl. **29**(5), 873–890 (2018). https://doi.org/10.1007/s00138-018-0943-x
5. Almuashi, M., Mohd Hashim, S.Z., Mohamad, D., Alkawaz, M.H., Ali, A.: Automated kinship verification and identification through human facial images: a survey. Multimed. Tools Appl. **76**(1), 265–307 (2017). https://doi.org/10.1007/s11042-015-3007-5
6. Fang, R., Tang, K.D., Snavely, N., Chen, T.: Towards computational models of kinship verification. In: Proceedings of the International Conference on Image Processing, ICIP, pp. 1577–1580 (2010). https://doi.org/10.1109/ICIP.2010.5652590
7. Duan, Q., Zhang, L., Zuo, W.: From face recognition to kinship verification: an adaptation approach (2017)
8. Wu, X., Feng, X., Li, L., Boutellaa, E., Hadid, A.: Kinship verification based on deep learning. In: Deep Learning in Object Detection and Recognition, pp. 113–132 (2019). https://doi.org/10.1007/978-981-10-5152-4_5
9. Chergui, A., et al.: Investigating deep CNNs models applied in kinship verification through facial images. To cite this version: HAL Id: hal-02400686 Investigating Deep CNNs Models Applied in Kinship Verification through Facial Images (2020)
10. Chergui, A., Ouchtati, S., Mavromatis, S., Eddine Bekhouche, S., Sequeira, J., Zerrari, H.: Kinship verification using mixed descriptors and multi block face representation. In: Proceedings of the ICNAS 2019 4th International Conference on Networking and Advanced Systems (2019).https://doi.org/10.1109/ICNAS.2019.8807875
11. Nandy, A., Mondal, S.S.: Kinship verification using deep siamese convolutional neural network. In: Proceedings of the 14th IEEE International Conference on Automatic Face & Gesture Recognition, FG 2019, August 2019. https://doi.org/10.1109/FG.2019.8756528
12. Yan, H., Wang, S.: Learning part-aware attention networks for kinship verification. Pattern Recognit. Lett. **128**, 169–175 (2019). https://doi.org/10.1016/j.patrec.2019.08.023
13. Van, T.N., Hoang, V.T.: Kinship verification based on local binary pattern features coding in different color space. In: 2019 26th International Conference on Telecommunications, ICT 2019, no. 2, pp. 376–380 (2019).https://doi.org/10.1109/ICT.2019.8798781
14. Zhang, H., Wang, X., Kuo, C.C.J.: Deep kinship verification VIA appearance-shape joint prediction and adaptation-based approach. In: Proceedings of the International Conference on Image Processing, ICIP, vol. September 2019, pp. 3856–3860 (2019). https://doi.org/10.1109/ICIP.2019.8803647
15. Goyal, A., Meenpal, T.: Kinship verification from facial images using feature descriptors. In: Advances in Intelligent Systems and Computing, vol. 768, pp. 371–380 (2019). https://doi.org/10.1007/978-981-13-0617-4_37
16. Mukherjee, M., Meenpal, T.: Kinship verification using compound local binary pattern and local feature discriminant analysis. In: 2019 10th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2019, no. July (2019). https://doi.org/10.1109/ICCCNT45670.2019.8944489

17. Chergui, A., Ouchtati, S., Mavromatis, S., Bekhouche, S.E., Lashab, M., Sequeira, J.: Kinship verification through facial images using CNN-based features **37**(1), (2020). https://doi.org/10.18280/ts.370101
18. Crispim, F., Vieira, T., Lima, B.: Verifying kinship from RGB-D face data. In: Blanc-Talon, J., Delmas, P., Philips, W., Popescu, D., Scheunders, P. (eds.) ACIVS 2020. LNCS, vol. 12002, pp. 215–226. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-40605-9_19
19. Zhou, H., et al.: A method for facial kinship verification based on deep learning. In: Advances in Intelligent Systems and Computing, AISC, vol. 1031, pp. 131–139 (2020). https://doi.org/10.1007/978-981-13-9406-5_17
20. Wu, H., Chen, J., Liu, X., Hu, J.: Component-based metric learning for fully automatic kinship verification. J. Vis. Commun. Image Represent. **79**, 103265 (2021). https://doi.org/10.1016/j.jvcir.2021.103265
21. Zekrini, F., Nemmour, H., Chibani, Y.: Feature fusion for kinship verification based on face image analysis. In: Lejdel, B., Clementini, E., Alarabi, L. (eds.) Artificial Intelligence and Its Applications. AIAP 2021. Lecture Notes in Networks and Systems, vol. 413, pp. 486–94. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-96311-8_45
22. Liu, F., Li, Z., Yang, W., Xu, F.: Age-invariant adversarial feature learning for kinship verification. Mathematics **10**(3), (2022). https://doi.org/10.3390/math10030480
23. Zhang, L., Duan, Q., Zhang, D., Jia, W., Wang, X.: AdvKin: adversarial convolutional network for kinship verification. IEEE Trans. Cybern. **51**(12), 5883–5896 (2021). https://doi.org/10.1109/TCYB.2019.2959403

Software Defects Detection in Explainable Machine Learning Approach

Muayad Khaleel Al-Isawi and Hasan Abdulkader(✉)

Altinbas University, 34217 Istanbul, Turkey
Hasan.abdulkader@altinbas.edu.tr, Hasan.abdulkader@yahoo.com

Abstract. In the era of ubiquitous software systems, the complexity and urgency in software production have often led to compromises in quality. Traditional testing methods are increasingly inadequate, demanding more automated solutions. This research explores the application of machine learning (ML) for Software Defect Prediction (SDP), specifically focusing on binary classification of defective and non-defective software components. Leveraging state-of-the-art ML models such as Random Forest, Artificial Neural Network (ANN), and XGBoost, the study rigorously evaluates their effectiveness on the Promise CM1 dataset. Moreover, the paper addresses the "black box" challenge by employing Explainable AI (XAI) techniques; SHapley Additive exPlanations (SHAP) is used to elucidate the models' decision-making processes. This approach balances predictive accuracy with interpretability, fostering trust, and promoting responsible usage of automated defect prediction. The research findings offer significant advancements in software quality assurance and provide an insightful perspective on the alignment between prediction capabilities and comprehensible models.

Keywords: Software Defect Prediction · Machine Learning · Explainable AI · SHapley Additive exPlanations

1 Introduction

In the modern era, software has become the central pillar of our interconnected world, permeating nearly every facet of our daily lives, from business operations and e-commerce to medical instruments and social networking platforms. The recent global shifts, such as the work-from-home culture imposed by the Covid-19 pandemic, have only accelerated the need for robust and reliable software systems [1, 2]. However, this exponential growth has brought with it an accompanying surge in complexity and a consequent risk to software quality. The intricate network structures and vast amounts of traffic data necessitate sophisticated systems that often defy traditional testing methods like manual testing and code review [3]. These conventional approaches often become infeasible in the face of the ever-growing complexity, leading to high costs and potential compromises in quality [4]. Furthermore, the urgency in software production to meet current demands has led to the overlooking of critical quality considerations. These issues can be especially detrimental in sensitive sectors like medical instruments, where

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 505–519, 2024.
https://doi.org/10.1007/978-3-031-56728-5_42

software faults can have serious consequences [5]. Software testing, while essential, is no longer sufficient in the modern software development life cycle [3, 6]. The manual approach, involving a testing team that tests cases, compares results, and identifies bugs, has become increasingly inadequate [7]. The rapid generation of vast amounts of data, introduction of new programming languages, and the evolving requirements for reliability and robustness all demand more automated and sophisticated solutions. Enter the emerging field of Software Defect Prediction (SDP) powered by machine learning (ML). In the field of software systems, the utilization of ML as a method to identify and predict defects has seen a significant uptick in recent years [8]. This increase in adoption can be traced back to the widespread availability of copious amounts of labeled data. At its core, ML equips a computer with the ability to analyze and learn from data [9], and it typically falls into one of three primary categories [10]: Supervised learning, where both attributes and labels are used; Unsupervised learning, where only the attributes are considered [11]; and Semi-supervised learning, a hybrid approach that includes a limited set of labeled data along with a more substantial collection of unlabeled data. Among these methodologies, supervised learning is often the most frequently applied, although unsupervised learning also finds considerable use in software flaw detection. Supervised learning itself can be subdivided into two distinct methods: classification, where labels are discrete variables, and regression, where labels are continuous variables [12]. In the context of the research presented in this paper, the emphasis will be on classification tasks, given that the datasets under examination are identified with discrete variables. Leveraging artificial neural networks (ANN), modern supercomputers equipped with GPUs, and AI-accelerating modules, the application of ML in predicting software defects has become a promising new frontier [13]. These automatic approaches not only reduce costs but also enhance software quality by employing specific algorithms to predict defects at an early stage [14]. This proactive strategy can prevent significant losses in money, time, and effort and protect against system collapses or degradation in valuation and reputation.

While the application of ML models in predicting software defects has been burgeoning, a significant challenge that has emerged is the “black box” nature of many of these models [15]. This term refers to the often opaque and complex internal workings of the algorithms, which can make it exceedingly difficult for users to understand or trust the decisions being made. Such lack of transparency can lead to skepticism and reluctance to adopt these models, especially in critical domains where understanding the decision-making process is paramount. There is, therefore, an increasing emphasis on explainable AI, where models are designed or accompanied by tools that elucidate how and why particular decisions are reached [16]. The ability to interpret and explain the model’s decisions not only enhances trust and confidence but also facilitates regulatory compliance and enables more effective collaboration between human experts and automated systems [17]. In the ongoing evolution of ML for software defect prediction, striking the right balance between predictive accuracy and interpretability will be a key consideration, ensuring that these powerful tools can be used responsibly and effectively. In this research, we delve into the domain of software defect prediction, focusing on the binary classification of defective and non-defective software components. Utilizing three state-of-the-art ML models—Random Forest, Artificial Neural Network (ANN), and

XGBoost—we rigorously test and evaluate their efficacy on the Promise CM1 dataset, a widely acknowledged benchmark in the field of software engineering. Our exploration extends beyond mere predictive accuracy, as we also strive to unravel the underlying mechanics of these models. Employing Explainable AI (XAI) techniques, specifically SHapley Additive exPlanations (SHAP), we provide insights into the decision-making processes of these models, elucidating why particular predictions are made. This commitment to transparency, facilitated by the combination of Random Forest, ANN, XGBoost, and SHAP, not only builds trust but also encourages wider adoption of automated defect prediction in the software industry. Our study offers a comprehensive perspective on software quality assurance, contributing significant advancements in prediction capabilities and enhancing the comprehensibility of these powerful models. The rest of this research is divided as follow: Sect. 2 represent the recent studies related to our research context, Sect. 3 detail the proposed framework for software defect detection, Sect. 4 illustrate the finding results and finally Sect. 5 illustrate the results interpretation and explanation.

2 Literature Review

The research paper introduces a unique technique to tackle class imbalance by using K-nearest neighbors (KNN) filtering with a stacked ensemble classifier [18]. By filtering overlapping data through KNN and reducing the imbalance ratio, the method incorporates static code metrics into an ensemble that includes five primary classifiers. The study employs a thorough analysis across five NASA datasets and 30 different classifiers, leading to 150 unique prediction models. The evaluation, conducted through various performance measures, revealed that the proposed stacked ensemble with KNN filtering stands out as the most effective approach across all examined datasets, providing a significant contribution to the field of software defect prediction. Jin et al. introduced a software fault-proneness model (SFPM) designed to predict faults in software [19]. The SFPM is a unique combination of three main components: an artificial neural network (ANN) for selecting appropriate metrics, a function to evaluate each metric's contribution, and an SVM learner responsible for making the actual predictions. The proposed model was put to the test with four datasets from the PROMISE repository, where it was benchmarked against five other learning methods, including SVM, LR, k-NN, NB, and the decision tree classifier (DT). The paper does not provide specific details about the comparative results, so it is not clear how SFPM performed relative to these other models. The authors of this study present a hybrid machine learning approach, blending Principal Component Analysis (PCA) and Support Vector Machines (SVM), to address a persistent problem in the field [20]. They utilized NASA's PROMISE data, specifically CM1 and KC1 datasets, to conduct their research. By applying PCA, they identified principal components that optimized the features, reducing time complexity. Subsequently, they used SVM for classification, citing its advantages over traditional methods, and employed GridSearchCV for hyperparameter tuning. The proposed hybrid model achieved better accuracy (CM1: 95.2%, KC1: 86.6%) compared to other methods and excelled in other evaluation criteria. However, the study also highlighted a limitation in the model, namely the lack of a probabilistic explanation for classification within SVM, which could make the model rigid towards classifications. In the study [21], the authors examined the use of a stacking ensemble comprising of adjustable tree-based ensembles, including Random Forest (RF), extra trees, AdaBoost, Gradient Boosting (GB), histogram-based GB, XGBoost, and CatBoost, for software defect prediction. They used grid search to fine-tune the hyperparameters of these models and subsequently constructed a stacking ensemble. The approach was tested on 21 publicly available defective datasets, and results indicated that RF ensembles and extra trees benefited significantly from hyperparameter tuning. Importantly, the stacking ensemble outperformed each of the individually fine-tuned tree-based ensembles, underscoring the potential of ensemble methods in conjunction with hyperparameter optimization to enhance software defect prediction. The method referenced in the cited work [22] focuses on balancing the sample categories to enhance the accuracy of software defect prediction. By utilizing the SMOTE technique in conjunction with XGBoost, the method was shown to outperform more traditional machine learning models, such as logistic regression, decision trees (DTs), Random Forests (RFs), and AdaBoost. The results validate the approach as a feasible solution to the software defect prediction problem, and the method's adaptability indicates its potential applicability in real-world software development tasks. In the research paper denoted as [23], the authors evaluated various machine learning models for software defect prediction (SDP) using four datasets from the NASA archive (KC2, PC3, JM1, and CM1). The models studied included Logistic Regression, Decision Trees (DT), Random Forests (RF), AdaBoost, and XGBoost. The researchers then introduced a new model that was specifically crafted by fine-tuning the XGBoost model, altering parameters such as the number of estimators, learning rate, maximum depth, and subsample. This innovative approach led to the new model outperforming all other contemporary models across the datasets, highlighting the potential efficacy of parameter optimization in XGBoost for SDP.

3 The Proposed Architecture

3.1 Dataset

In this study, we have utilized an open-source dataset related to NASA's software projects [24]. The CM1 dataset is composed of 327 samples, amongst which 42 are defective and 285 are non-defective (Fig. 1).

3.2 Preprocessing

Data Balancing. Our methodology centers around using the Synthetic Minority Oversampling Technique (SMOTE) to address the imbalance between defective and non-defective instances in the dataset [24].

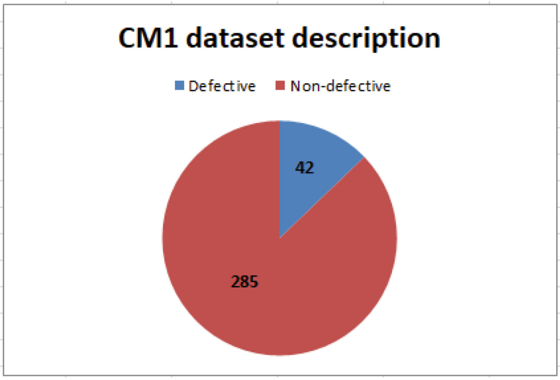


Fig. 1. CM1 data distribution.

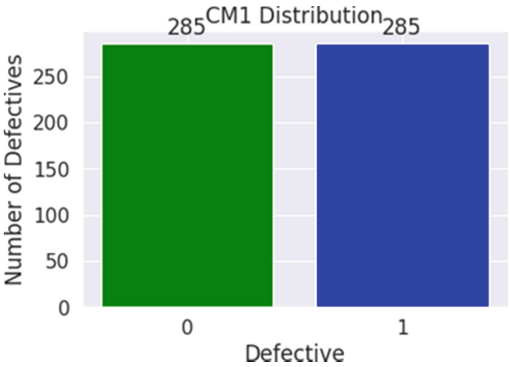


Fig. 2. CM1 data distribution after the application of balancing.

In the context of our study, where instances symbolize software defects, non-defective instances constitute the majority class, and defective ones represent the minority. SMOTE generates synthetic defective instances through a process called interpolation, injecting them into the dataset to create a balanced distribution [24]. This approach avoids biased model and overfitting, a frequent issue when merely replicating the minority class instances [24]. The distribution of the balanced data set is shown in Fig. 2.

Feature Selection. Gain Ratio (GR) is applied for feature selection. This process is essential to minimize overfitting, enhance accuracy, and reduce training time by identifying attributes that significantly contribute to the prediction. GR, an adaptation of Information Gain [25], normalizes Information Gain by evaluating the number and size of dataset branches for each attribute. This normalization lessens bias towards multi-valued attributes, resulting in more balanced selection and improved decision tree classifier performance [26]. With GR, we have increased our predictive model's accuracy and optimized computational resource usage, demonstrating this feature selection method's effectiveness in software defect prediction models.

Proposed Models.
XGBoost Classifier. XGBoost, standing for Extreme Gradient Boosting, represents an enhancement of Gradient Boosting ensembles. Unlike traditional Gradient Boosting, which relies on the first-order gradients of the loss function to train a new base classifier, XGBoost goes a step further. It utilizes the second-order derivatives of the loss function, allowing for a more precise and efficient determination of the optimal base classifier [27]. Model optimization was conducted through the synergistic employment of Extreme Gradient Boosting (XGBoost) and the grid search methodology, a com prehensive technique to fine-tune hyperparameters. The grid search, realized through Scikit-Learn's GridSearchCV functionality, was devised to explore a well-defined grid of hyperparameters including learning rate, number of estimators, maximum depth of trees, minimum child weight, gamma, subsample ratio, colsample by tree ratio, and the objective function, which was specifically set to 'binary: logistic' to align with the binary classification task.

This exhaustive search process allowed for an in-depth exploration of the hyperparameter space, meticulously evaluating each combination through a 3-fold cross-validation process. The best parameters were determined as 'colsample bytree': 0.6, *gamma*: 0.2, *learning rate*: 0.1, *max depth*: 10, *min-child-weight*: 1, *n estimators*: 300, *objective*: *binary: logistic*, *subsample*: 0.8, reflecting an optimal balance that guided the learning algorithm. These parameters hold distinct significance:

- Learning rate: Controls the contribution of each tree, with 0.1 providing a suitable compromise between learning speed and model performance.
- Number of estimators: The chosen 300 trees ensured adequate model complexity without overfitting.
- Maximum depth: The depth of 10 allowed the model to learn intricate patterns.
- Minimum child weight, gamma, and subsample ratio: These parameters aided in controlling overfitting.
- Colsample by tree ratio: The 0.6 value helped in adding randomness, making the model robust.

The best-performing model was then extracted, fully calibrated with these optimal parameters, highlighting the efficiency of grid search in systematically determining the most effective configuration. This rigorous hyperparameter tuning process demonstrated the importance of aligning model complexity with the underlying data structure, consequently enhancing the model's generalization capability on unseen data. By systematically exploring the hyperparameter space, grid search facilitated the discovery of a model configuration that harmonized complexity and performance, contributing significantly to the overall success of the predictive model.

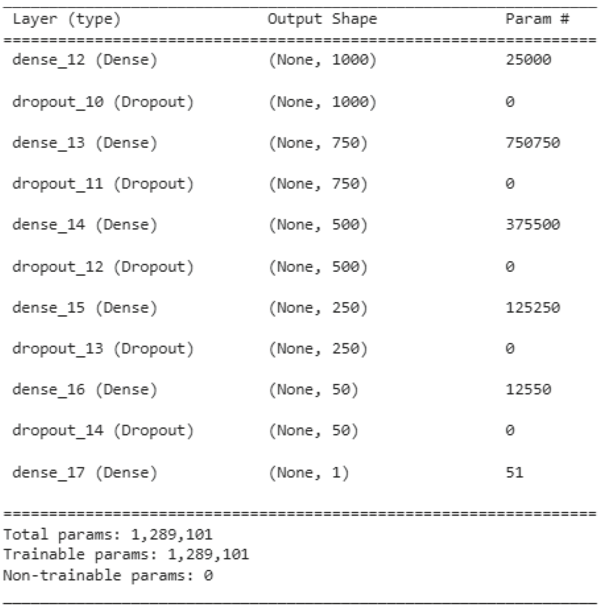
Random Forest Classifier. In our approach, we employed the Random Forest (RF) classifier, a robust ensemble learning method that operates by constructing multiple decision trees during training and outputting the class that is the mode of the classes from individual trees for classification problems, or the mean prediction of the individual trees for regression problems. To find the optimal configuration for the RF model, we leveraged the GridSearchCV method from the Scikit-Learn library, enabling us to perform an exhaustive search over a specified hyperparameter grid. The hyperparameters tuned during the grid search include:

- n estimators: The number of trees in the forest, tested with values [50, 100, 200].
- Max depth: The maximum depth of the tree, with possibilities of [None, 10, 20, 30].
- Min samples split: The minimum number of samples required to split an internal node, tested with [3, 6, 11].
- Min samples leaf: The minimum number of samples required to be at a leaf node, with options [2, 3, 5].
- Max features: The number of features to consider when looking for the best split, using either 'auto' or 'sqrt'.
- criterion: The function to measure the quality of a split, either 'gini' or 'entropy'.

Upon fitting the grid search to the training data, the best parameters obtained were criterion as 'gini', max_depth at 20, max features using 'sqrt', min _samples_leaf of 1, min _amples_split of 2, and n estimators of 200. These hyperparameters were determined to provide an optimal balance between model complexity and predictive performance, thereby ensuring a robust and efficient RF model tailored to our dataset.

Artificial Neural Network. Artificial Neural Networks (ANNs) have emerged as a powerful and flexible class of machine learning models, inspired by the structure and functionality of the human brain. ANNs are particularly well-suited for complex pattern recognition tasks, making them a popular choice for various applications, including software defect prediction. In this context, we have developed an ANN model for binary classification, aiming to accurately identify defective and non-defective instances in software systems. The architecture of our ANN model is designed to handle a dataset with 24 features, each representing specific attributes of the software. The model begins with an input layer of 1000 neurons, each using the Rectified Linear Unit (ReLU) activation function to introduce non-linearity to the network. ReLU is commonly used in deep learning models due to its ability to handle vanishing gradient problems and accelerate convergence. To improve the generalization capability of the model and prevent overfitting, we have added Dropout layers after each dense layer with a dropout rate of 0.5. Dropout is a regularization technique that randomly drops a fraction of the neurons during training, which forces the model to learn more robust and general representations of the data. The model then continues with three more dense layers, each with a decreasing number of neurons, i.e., 750, 500, and 250, respectively. These layers continue to use the ReLU activation function and are followed by Dropout layers to further enhance the model's resilience to overfitting. The last dense layer, consisting of 50 neurons, also utilizes the ReLU activation function and is followed by a final Dropout layer. Finally, to obtain the binary classification output, we include an output layer with a single neuron and a Sigmoid activation function.

The Sigmoid activation function squashes the output values between 0 and 1, representing the probability of the input belonging to the positive class (defective) in our case. This ANN architecture is designed to efficiently capture intricate patterns and relationships present in the input data while mitigating the risk of overfitting through the application of Dropout regularization. The model's depth and width were chosen to strike a balance between complexity and computational efficiency, ultimately delivering an effective binary classification model for software defect prediction. The optimizer used for training is Adam, a popular optimization algorithm known for its efficiency and adaptive learning rates (Fig. 3).



```
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense_12 (Dense)            (None, 1000)              25000

 dropout_10 (Dropout)        (None, 1000)              0

 dense_13 (Dense)            (None, 750)               750750

 dropout_11 (Dropout)        (None, 750)               0

 dense_14 (Dense)            (None, 500)               375500

 dropout_12 (Dropout)        (None, 500)               0

 dense_15 (Dense)            (None, 250)               125250

 dropout_13 (Dropout)        (None, 250)               0

 dense_16 (Dense)            (None, 50)                12550

 dropout_14 (Dropout)        (None, 50)                0

 dense_17 (Dense)            (None, 1)                 51

=================================================================
Total params: 1,289,101
Trainable params: 1,289,101
Non-trainable params: 0
_________________________________________________________________
```

Fig. 3. The ANN architecture.

The BinaryCrossentropy loss function is employed, which is suitable for binary classification tasks like software defect prediction, where the goal is to distinguish between defective and non-defective instances. During the training process, the model undergoes 500 epochs, which represents the number of times the entire training dataset is processed by the model. The training dataset is further divided into a validation set with a 20% split.

4 Results and Discussion

To evaluate the models' results we use the following metrics (Precision, Recall, F1-Score and Accuracy). These metrics are given as follows:

1. **Precision**: It is defined as the ratio of true positives (TP) to the sum of true positives and false positives (FP):

$$Precision = \frac{TP}{TP + FP}$$

2. **Recall**: Also known as Sensitivity, it is defined as the ratio of true positives (TP) to the sum of true positives and false negatives (FN):

$$Recall = \frac{TP}{TP + FP}$$

3. **F1-Score**: The harmonic mean of precision and recall:

$$F1 - Score = 2 \cdot \frac{Precision \times Recall}{Precision + Recall}$$

4. **Accuracy**: The ratio of correctly predicted instances to the total instances in the dataset:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

4.1 XGBoost Results

The XGBoost model for software defect prediction displayed favorable performance, achieving an accuracy of 89%. For class 0 (non-defective instances), the model attained a precision of 93%, signifying that 93% of the instances classified as non-defective were genuinely non-defective. The recall (also known as sensitivity) for class 0 was 87%, indicating that the model correctly identified 87% of the actual non-defective instances. Concerning class 1 (defective instances), the precision was 86%, meaning that 86% of the instances classified as defective were truly defective. The recall for class 1 was 93%, showing that the model accurately detected 93% of the actual defective instances. Both classes had an F1-score of 89%, which indicates a balanced trade-off between precision and recall. This 2 × 2 matrix (Fig. 4) summarizes the model's performance. It shows 76 true negative (TN) predictions, 73 true positive (TP) predictions, 14 false negative (FN) predictions, and 8 false positive (FP) predictions.

4.2 Random Forest Results

The confusion matrix reveals that the model accurately predicted 76 true negative (TN) instances and 73 true positive (TP) instances. However, it misclassified 14 instances as false negatives (FN) and 8 instances as false positives (FP). The precision for class 0 (non-defective instances) was 90%, indicating that 90% of the instances classified as non-defective were indeed non-defective. For class 1 (defective instances), the precision was 84%, showing that 84% of the instances classified as defective were genuinely defective.

The recall (also known as sensitivity) for class 0 was 84%, indicating that the model correctly identified 84% of the actual non-defective instances. For class 1, the recall was 90%, demonstrating that the model accurately detected 90% of the actual defective instances. The F1-score, which balances precision and recall, was 87% for both classes, indicating a harmonious balance between precision and recall (Fig. 5).

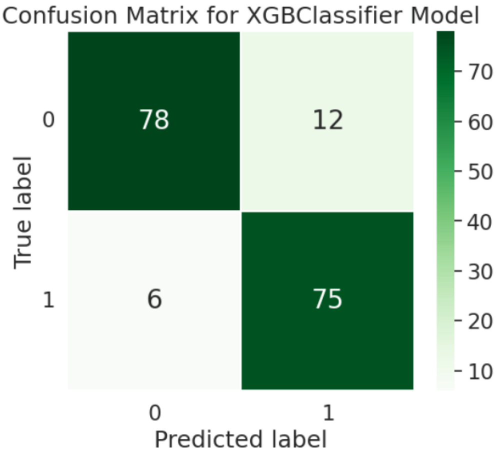


Fig. 4. Confusion matrix of XGBoot model.

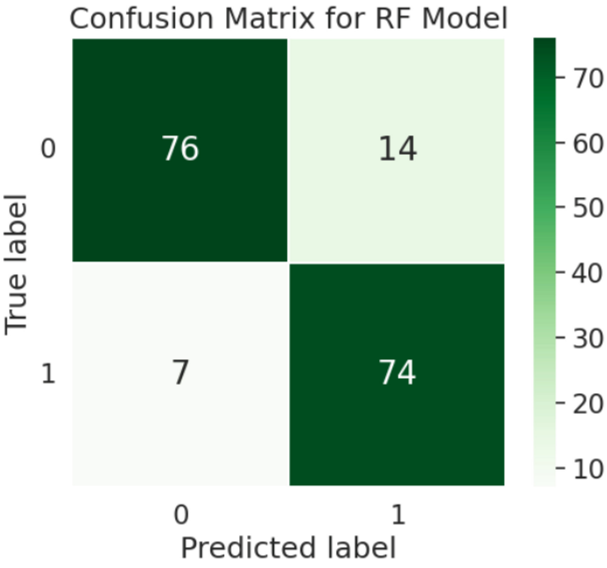


Fig. 5. Confusion matrix of RF model.

4.3 ANN Results

The Artificial Neural Network (ANN) model for software defect prediction achieved a confusion matrix (Fig. 6). This 2 × 2 matrix summarizes the model's performance, showing 79 TN predictions, 77 TP predictions, 11 FN predictions, and 4 FP predictions. The high number of TN and TP predictions indicates that the model accurately classified a significant portion of non-defective and defective instances, respectively. However, the presence of FN and FP predictions suggests some misclassifications.

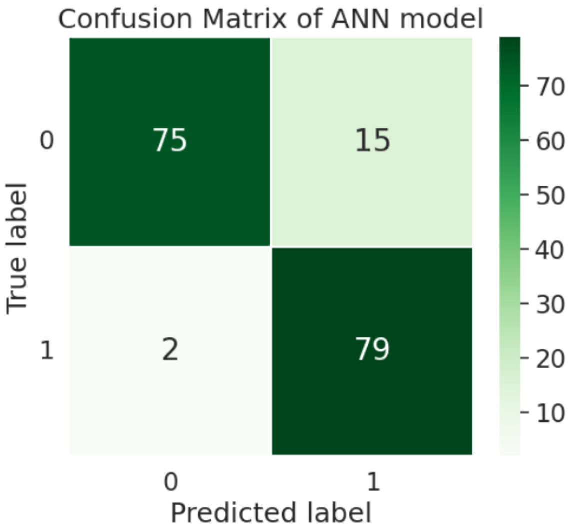


Fig. 6. Confusion matrix of the ANN model.

The ANN model for software defect prediction demonstrated strong performance, with an overall accuracy of 91%. The model achieved a precision of 95% for class 0 (non-defective instances), indicating that 95% of the instances classified as non-defective were indeed non-defective. For class 1 (defective instances), the precision was 88%, indicating that 88% of the instances classified as defective were genuinely defective. The recall (also known as sensitivity) for class 0 was 88%, indicating that the model correctly identified 88% of the actual non-defective instances. For class 1, the recall (also known as sensitivity) was 95%, showing that the model accurately detected 95% of the actual defective instances. The F1-score, which balances precision and recall, was 91% for both classes, indicating a harmonious balance between precision and recall.

4.4 Explainable AI

SHAP is a well-established, comprehensive framework for interpreting various models. It explains the predictions for a specific instance by calculating each feature's impact on the final decision, which can be either positive or negative [27] and [28]. Unlike linear methods, SHAP is applicable to any model or classifier. Instead of concentrating just on local interpretations, SHAP takes into account global interpretations by averaging each feature independently and adding up the input values of the features. The explanation for an instance using SHAP can be derived as follows [29]:

$$g(s) = v_0 + \sum_{i=1}^{N} v_i s_i \tag{1}$$

In this Eq. 1, N is the maximum size of the feature vector, and v_i is the Shapley value for feature i. The Shapley value represents the contribution of each feature to the model's prediction, with higher values indicating a larger contribution. SHAP values are computed by the rule:

To identify the most important features, the SHAP value is computed by the equation:

$$v_j = \frac{1}{N!} \sum_{S \subseteq N \backslash j} |S|! \cdot (N - |S| - 1)! \cdot \left[f\left(S \bigcup \{j\}\right) - f(S)\right] \tag{2}$$

Here in Eq. 2, $|S|$ is the number of features in the subset S, and $f(.)$ represents the output of the model. By computing v_j for all features, we can identify the most influential features in the model's predictions. Utilizing the SHAP (Shapley Additive Explanations) method, the CM1 dataset was scrutinized for binary class classification within the context of the ANN model, with the results illustrated in Fig. 6 (Fig. 7).

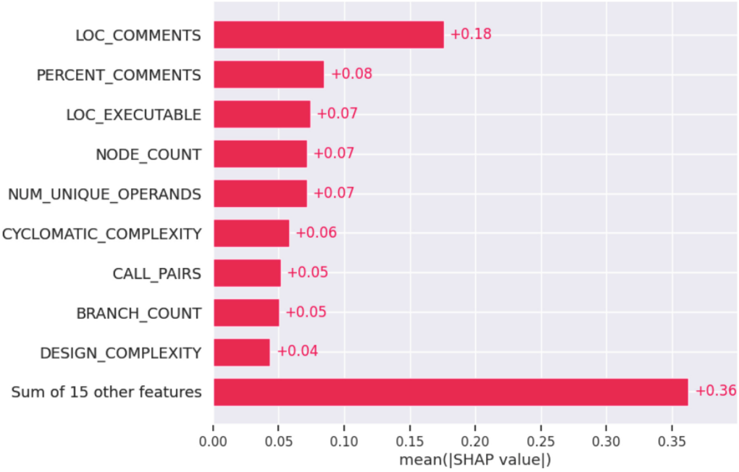


Fig. 7. The ANN Feature importance scores using SHAP technique.

Among the features identified with high importance scores are 'LOC COMMENTS', 'PERCENT COMMENTS', 'LOC EXECUTABLE', 'NODE COUNT', 'NUM UNIQUE OPERANDS', 'CYCLOMATIC COMPLEXITY', 'CALL PAIRS' among other features. These features were instrumental in the model's ability to accurately predict software defects, demonstrating the effectiveness of the SHAP method in identifying key characteristics.

Table 1 compares the SHAP importance values of algorithms Random Forest and XGBoost classifier. It is noticeable that the influential features vary, and the order of importance varies. The table allows the comparison of the sum of SHAP values of ANN, RF and XGBoost algorithms. The sum of SHAP values reflects the performance of the algorithm. The accuracy is evaluated to be 91% for ANN, 89 for XGBoost and 76.4 for Random Forest. Henceforth the sum of SHAP values of the same algorithms are 1.03, 0.93 and 0.71 correspondingly.

Table 1. SHAP Values of RF and XGBoost algorithms.

Random Forest		XGBoost Classifier	
Features	Mean SHAP Values	Features	Mean SHAP Values
LOC COMMENTS	0.09	LOC COMMENTS	0.17
DESIGN COMPLEXITY	0.05	CALL PAIRS	0.1
LOC EXECUTABLE	0.05	DESIGN COMPLEXITY	0.09
PERCENT COMMENTS	0.05	NUM UNIQUE OPERANDS	0.06
HALSTEAD COUNT	0.05	LOC EXECUTABLE	0.06
CALL PAIRS	0.04	PERCENT COMMENTS	0.05
NUM UNIQUE OPERANDS	0.04	CYCLOMATIC COMPLEXITY	0.04
NODE COUNT	0.04	HALSTEAD CONTENT	0.04
LOC TOTAL	0.03	NODE COUNT	0.04
Total	0.71	Total	0.93
Total of mean SHAP' Values relatives to ANN model			1.03

5 Conclusion

In this research, we delved into the intricate landscape of Software Defect Prediction, examining the capabilities and challenges of cutting-edge machine learning models, including Random Forest, ANN, and XGBoost. By applying these models to the Promise CM1 dataset, a recognized benchmark in software engineering, we not only highlighted their predictive accuracies but also explored the inner mechanics through the lens of Explainable AI (XAI), specifically utilizing the SHapley Additive exPlanations (SHAP) method. Our findings demonstrated that these models are not only robust in predicting defects but also transparent, fostering trust and broader adoption within the industry. The results signal a promising solution for enhancing software quality, mitigating costs, and preventing system collapses or degradation in valuation and reputation. Nevertheless, our work also illuminated the “black box” challenge, underscoring the vital need for balancing predictive efficacy with interpretability. The exploration of machine learning in software defect prediction has only scratched the mystery of their performance.

References

1. del Rio-Chanona, R.M., Mealy, P., Pichler, A., Lafond, F., Farmer, J.D.: Supply and demand shocks in the Covid-19 pandemic: an industry and occupation perspective. Oxford Rev. Econ. Policy **36**(1), S94–S137 (2020)
2. Nawaz, A., Rehman, A.U., Abbas, M.: A novel multiple ensemble learning models based on different datasets for software defect prediction. arXiv preprint arXiv:2008.13114 (2020)
3. Sharma, R.: Quantitative analysis of automation and manual testing. Int. J. Eng. Innov. Technol. **4**(1), 252–257 (2014)
4. Gewaltig, M.-O., Cannon, R.: Current practice in software development for computational neuroscience and how to improve it. PLoS Comput. Biol. **10**(1), e1003376 (2014)
5. Jamil, M.A., Arif, M., Abubakar, N.S.A., Ahmad, A.: Software testing techniques: a literature review. In: 2016 6th International Conference on Information and Communication Technology for the Muslim World (ICT4M), pp. 177–182. IEEE (2016)
6. Singh, S.K., Singh, A.: Software Testing. Vandana Publications (2012)
7. Tian, Z., Xiang, J., Zhenxiao, S., Yi, Z., Yunqiang, Y.: Software defect prediction based on machine learning algorithms. In: 2019 IEEE 5th International Conference on Computer and Communications (ICCC), pp. 520–525. IEEE (2019)
8. Kassab, M., DeFranco, J.F., Laplante, P.A.: Software testing: the state of the practice. IEEE Softw. **34**(5), 46–52 (2017)
9. Khan, A.H., Siddqui, J., Sohail, S.S.: A survey of recommender systems based on semi-supervised learning. In: Khanna, A., Gupta, D., Bhattacharyya, S., Hassanien, A.E., Anand, S., Jaiswal, A. (eds.) International Conference on Innovative Computing and Communications. AISC, vol. 1394, pp. 319–327. Springer, Singapore (2022). https://doi.org/10.1007/978-981-16-3071-2_27
10. Park, M., Hong, E.: Software fault prediction model using clustering algorithms determining the number of clusters automatically. Int. J. Softw. Eng. Appl. **8**(7), 199–204 (2014)
11. Nasteski, V.: An overview of the supervised machine learning methods. Horizons B **4**, 51–62 (2017)
12. Akimova, E.N., et al.: A survey on software defect prediction using deep learning. Mathematics **9**(11), 1180 (2021)
13. Catal, C., Diri, B.: Investigating the effect of dataset size, metrics sets, and feature selection techniques on software fault prediction problem. Inf. Sci. **179**(8), 1040–1058 (2009)
14. Zahavy, T., Ben-Zrihem, N., Mannor, S.: Graying the black box: understanding DQNs. In: International Conference on Machine Learning, pp. 1899–1908. PMLR (2016)
15. Alicioglu, G., Sun, B.: A survey of visual analytics for explainable artificial intelligence methods. Comput. Graph. **102**, 502–520 (2022)
16. Antoniadi, M., et al.: Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review. Appl. Sci. **11**(11), 5088 (2021)
17. Goyal, S.: Predicting the defects using stacked ensemble learner with filtered dataset. Autom. Softw. Eng. **28**(2), 14 (2021)
18. Jin, S.-W., Ye, J.-M.: Artificial neural network-based metric selection for software fault-prone prediction model. IET Softw. **6**(6), 479–487 (2012)
19. Mustaqeem, M., Saqib, M.: Principal component-based support vector machine (PC-SVM): a hybrid technique for software defect detection. Clust. Comput. **24**(3), 2581–2595 (2021)
20. Alazba, A., Aljamaan, H.: Software defect prediction using stacking generalization of optimized tree-based ensembles. Appl. Sci. **12**(9), 4577 (2022)
21. Yang, H., Li, M.: Software defect prediction based on SMOTE-Tomek and XGBoost. In: Pan, L., Cui, Z., Cai, J., Li, L. (eds.) BIC-TA 2021. CCIS, vol. 1566, pp. 12–31. Springer, Singapore (2022). https://doi.org/10.1007/978-981-19-1253-5_2

22. Gupta, A., Sharma, S., Goyal, S., Rashid, M.: Novel XGBoost tuned machine learning model for software bug prediction. In: 2020 International Conference on Intelligent Engineering and Management (ICIEM), pp. 376–380. IEEE (2020)
23. García, V., Sánchez, J., Martín-Félez, R., Mollineda, R.A.: Surrounding neighborhood-based smote for learning from imbalanced data sets. Prog. Artif. Intell. **1**, 347–362 (2012). https://doi.org/10.1007/s13748-012-0027-5
24. Shepperd, M., Song, Q., Sun, Z., Mair, C.: Data quality: some comments on the NASA software defect datasets. IEEE Trans. Software Eng. **39**(9), 1208–1215 (2013)
25. Quinlan, J.R.: Induction of decision trees. Mach. Learn. **1**(1), 81–106 (1986). https://doi.org/10.1007/BF00116251
26. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794 (2016)
27. Dazeley, R., Vamplew, P., Foale, C., Young, C., Aryal, S., Cruz, F.: Levels of explainable artificial intelligence for human-aligned conversational explanations. Artif. Intell. **299**, 103525 (2021)
28. Yeung, C., et al.: Elucidating the behavior of nanophotonic structures through explainable machine learning algorithms. ACS Photon. **7**(8), 2309–2318 (2020)
29. Rodríguez-Pérez, R., Bajorath, J.: Interpretation of machine learning models using Shapley values: application to compound potency and multi-target activity predictions. J. Comput. Aided Mol. Des. **34**, 1013–1026 (2020). https://doi.org/10.1007/s10822-020-00314-0

Explainable AI for Predicting User Behavior in Digital Advertising

Ashraf Al-Khafaji(✉) and Oguz Karan

Altinbas University, 34217 Istanbul, Turkey
ashraffarhan226@gmail.com, oguz.karan@altinbas.edu.tr

Abstract. Online advertising has ushered in a new era of digital communication and business transformation. However, the inundation of digital content necessitates a deeper understanding of user behavior to ensure meaningful engagement. This paper investigates the potential of machine learning in predicting and analyzing user behavior in the realm of online advertising. Utilizing a dataset encompassing user interactions with advertisements, we deployed three machine learning models: Random Forest, Logistic Regression, and Gradient Boosting. Our findings highlight that the Random Forest model outperformed with an accuracy of 97.67%, followed closely by Logistic Regression and Gradient Boosting. Furthermore, recognizing the opaque nature of machine learning models, our research leverages SHAP and LIME, tools of explainable AI, ensuring that our models' decisions remain interpretable. This study under-scores the power of a data-driven approach in online advertising, emphasizing the necessity for both precision and transparency in this digital age.

Keywords: Online Advertising · User Behavior Analysis · Machine Learning · SHAP · LIME

1 Introduction

Online advertising has fundamentally transformed the realm of business, fostering a digital space where individuals, organizations, and governments interact and exchange information across myriad domains such as commerce, sociality, politics, and education [1]. This digital evolution has been epitomized by the rising popularity of social media platforms like Facebook and the exponential increase in advertising expenditures, indicating the profound interest in the digital advertising landscape [2, 3]. However, this burgeoning realm presents both opportunities and challenges [4]. On one hand, the vast scope of digital content offers businesses unprecedented reach to their target audiences; on the other, the sheer volume of advertising content threatens to overwhelm users, leading to potential disengagement [2, 3]. The heart of effective digital advertising lies in understanding the behavior of the website visitor [5]. As users navigate through websites, engage with content, or even merely hover over a particular advertisement, they generate a wealth of data that, if analyzed correctly, can yield insights into their preferences, habits, and motivations [6]. Traditionally, understanding user behavior relied on methods

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 520–531, 2024.
https://doi.org/10.1007/978-3-031-56728-5_43

such as surveys or focus groups. Still, with the advent of machine learning, there has been a paradigm shift towards data-driven advertising. Machine learning, with its capability to process vast datasets and identify intricate patterns, offers a robust framework to tailor advertisements in real-time, optimizing both content and placement to ensure maximum user engagement. However, harnessing the power of machine learning for advertising is not without its challenges. The rapidly evolving digital landscape requires continuous adaptation and a proactive approach to remain relevant and effective. Moreover, issues such as ad-blockers and click fraud present tangible obstacles in fully realizing the potential of machine learning-driven advertising. An equally pressing concern lies in the interpretability of these machine learning models [7]. The inherent complexity of such models often renders them as 'black boxes,' making it difficult for stakeholders to understand and trust their decisions [8]. Recognizing this challenge, our research incorporates the principles of explainable AI using tools like SHAP and LIME [9]. These methodologies provide transparency into our model's decision-making processes, fostering trust, and ensuring that our advertising recommendations are both effective and interpretable [10, 11]. This paper delves into the complex interplay between user behavior, machine learning, and the interpretability of models in the context of online advertising. We present a comprehensive machine learning-based approach to analyze and predict user behavior while emphasizing the importance of explainability. By leveraging advanced machine learning algorithms, ensemble learning techniques, and explainability tools, we aim to enhance user experience, increase engagement, and maximize return on investment for businesses in the digital advertising sector. This paper is structured as follows: Sect. 2 provides an overview of related work in the field of fraud detection and BC technology. In Sect. 3 we illustrate the Explainable Artificial Intelligence (XAI). In Sect. 4, we present the proposed system model and define the problem statement. Section 5 discusses the simulation results and the performance of the implemented ML models and the application of XAI methods (SHAP and LIME). Finally, Sect. 6 concludes the paper and outlines potential directions for future research in this area.

2 Related Work

The sensitivity of the K-means method to the k-value in data clustering has been addressed by integrating the kernel density selection technique. This adjustment capitalizes on the unique density distribution characteristics [12]. This enhanced method is executed on the big data Flink platform, ensuring efficient parallel processing. Practical application in mobile e-commerce was tested using both the enhanced and the traditional K-means methods in concurrent and sequential modes. The results affirm the enhanced technique's superiority in clustering quality. Beyond its precision and efficiency, the revised algorithm showcases notable advantages when implemented in real-world scenarios. In their research, [13] delved into predicting consumer behavior on social media by leveraging big data methodologies. Utilizing mathematical models paired with Machine Learning algorithms, they assessed customer behavior rooted in their social media inter-actions and various attributes. Among the different models tested, the Decision Tree (DT) stood out, boasting an impressive accuracy in predicting consumer behavior, with deviations oscillating between 12.22% and 99.51%. The overall accuracy of this model spanned a range from 0.22% to 98%. In their study, [14] introduced a technique aimed at identifying click fraud within advertising networks. The system they presented effectively pinpointed various click fraud categories, offering a means to safeguard advertisers against potential financial repercussions. The research under-scored the importance of fine-tuning criteria weights for accurate fraud classification in practical scenarios. On the other hand, [15] unveiled an innovative algorithm tailored for refining targeting on corporate Facebook pages. This method harnessed DTs to fine-tune marketing strategies and emulate user interactions. The resultant simulation data indicated that the algorithm pinpointed target demographics that surpassed conventional industry profitability metrics, underscoring its efficacy. Emotion analysis in social media posts has risen as a pivotal research area, primarily due to its potential in identifying conditions like depression. The sentiments expressed by individuals, especially in product reviews and technology interactions, offer insights into their emotional states [16, 17]. Modern studies in this domain have pinpointed specific linguistic traits, including pre-dominant self-referential language and negative emotional expressions, as indicators of depression [18, 19]. As an illustration, [20] analyzed a Twitter dataset from individuals diagnosed with depression. Their findings revealed a distinct trend towards negatively charged words in the tweets. Moreover, their research underscored a compelling observation: expectant mothers predisposed to postpartum depression began showcasing altered emotional articulation, linguistic nuances, and social interactions on Twitter even before their child's birth [16]. In pursuit of refining the online methods for detecting depression, contemporary research is channeling efforts into the extraction and categorization of features indicative of user behaviors. These categorizations encompass metrics like posting intervals, temporal distribution of posts, and follower dynamics. A significant study by [21] categorized these features into user profiles, user behavior, and user content. They subsequently harnessed a multi-kernel SVM for classifying these features.

3 Explainable Artificial Intelligence (XAI)

XAI aims to elucidate the decision-making processes of AI models, making them transparent and comprehensible to human users [22]. By emphasizing transparency and intelligibility, XAI seeks to bolster trust, accountability, and reliability in AI systems [23]. This becomes particularly indispensable in the realm of cybersecurity. While AI models have the prowess to sift through vast datasets, identifying potential cyber threats, their opaque nature often leaves cybersecurity professionals in the dark about the underlying rationale for such detections [24]. XAI bridges this gap. It empowers these professionals with insights into the AI's decision-making, facilitating well-informed responses to the evolving landscape of cyber threats [11].

3.1 Local Interpretable Model-Agnostic Explanations (LIME)

LIME strives to elucidate complex models by generating a simplified, easily interpretable representation that maintains fidelity to the original model's local decisions. Considering an instance characterized by its original representation, x R^d, and an explanatory model,

g G (with G denoting a collection of transparent, visually interpretable models like linear models), the explanation proposed by LIME can be represented as:

$$\varphi(x) = \arg\min_{s\in G}\left[C(f, g, \omega_x) + \Omega(g)\right] \quad (1)$$

In the above Eq. 1, f stands for the primary classification model, while ω_x defines a similarity measure between the instance's original and transformed representations (where a higher value indicates a stronger similarity). The loss function L assesses the agreement in predictions between the original classifier and its interpretation. $\Omega(g)$, on the other hand, quantifies the complexity of the interpretative model g. LIME's objective is to craft a model with a dual focus: locality and comprehensibility. This is achieved by minimizing the combined term $L(f, g, \omega_x) + \Omega(g)$. Here, f signifies the primary model, g the local interpretative model, and ω_x acts as a weight vector for the instance x. The regularization term, $\Omega(g)$, ensures the interpretative model remains concise, avoiding over-complication.

Upon optimizing the objective function, LIME then formulates an instance-specific explanation through the locally derived model, $\phi(x)$. Designed for transparency, $\phi(x)$ aids humans in grasping the rationale behind specific predictions. Through its emphasis on localized, comprehensible models, LIME illuminates the intricate decision-making of more convoluted models.

3.2 Shapley Additive Explanations (SHAP)

SHAP provides a robust method for elucidating model predictions, offering clarity on how individual features influence a specific prediction outcome, either positively or negatively. Unlike approaches that might be restricted to linear models, SHAP is versatile, designed to interpret a wide range of classifiers. It not only offers insight into localized explanations but also incorporates a global perspective. It does this by averaging the effect of each feature across the dataset and summing the input feature values. The explanatory function for an instance via SHAP is given as:

$$g(s) = v_0 + \sum\nolimits_{i=1}^{N} v_i s_i \quad (2)$$

In the above Eq. 2, N signifies the total number of features, while v_i denotes the Shapley value associated with the i^{th} feature. This Shapley value quantifies the relative importance of the respective feature in contributing to the final model prediction, where larger values suggest a more prominent role. To rank and discern the significance of features, we can utilize the following importance metric:

$$IF_j = \sum\nolimits_{i=1}^{n} |v_j(x_i)| \quad (3)$$

Within Eq. 3, n denotes the overall number of data instances, and IF_j indicates the average magnitude of the Shapley value associated with the j^{th} feature. By evaluating IF_j across all features, we can gain insights into the hierarchies of feature significance in influencing model outcomes.

4 Methodology

4.1 Dataset

In our research, we utilize a carefully assembled dataset containing 1,000 records, each with 8 distinct features. These records represent user interactions with online advertisements. Key features in this dataset include 'Daily Time Spent on Site', 'Age', 'Area Income', 'Daily Internet Usage', and 'Male'. Central to our analysis is the binary target variable "Clicked on Ad". A value of '1' signifies a user clicked on the ad, whereas '0' indicates no interaction. Our primary objective, using this dataset, is to delve into user behavior patterns in the realm of online advertising, aiming to bolster understanding in this domain. The age distribution within the dataset ranges from 19 to 61 years, showcasing a diverse age group. Of the total, 51.9% (519 users) are females, and 48.1% (481 users) are males, indicating a balanced gender representation, which helps to avert notable gender biases. It's pivotal to note that our dataset reflects an even distribution concerning user-ad interactions. Roughly 50% of the users engaged with the ads, while the other half abstained (as visualized in Fig. 1). This parity is beneficial for our research, as it curtails potential skews towards any particular class, laying a robust foundation for ensuing predictive analysis.

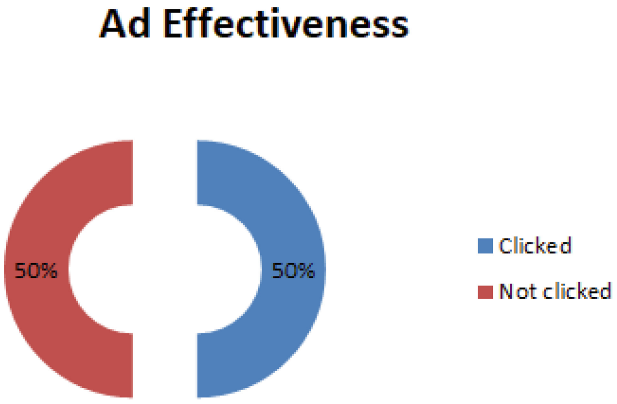


Fig. 1. Ad effectiveness label distribution.

4.2 Preprocessing

To uphold the credibility and authenticity of our research, we executed meticulous quality assessments on our dataset. This involved ensuring there were no missing values or duplicate records, thereby preserving the data's integrity. The dataset's completeness and precision offer a robust base for our subsequent analyses, ensuring that our findings and outcomes are derived from consistent and bias-free information. In our analytical framework, we identify the feature matrix (X) and the target variable (y) as critical elements. The matrix, X, incorporates attributes such as 'Daily Time Spent on Site', 'Age', 'Area Income', 'Daily Internet Usage', and 'Male'. These factors are instrumental in investigating the correlation between user demographics and their ad interactions. Conversely, the target variable, y, signifies if a user engaged with an ad. This binary outcome is central to our ensuing modeling, facilitating the creation of models that segment users based on their propensity to interact with ads. For an unbiased evaluation of our models' performance and adaptability, we divided the dataset into training and test subsets. By randomizing the dataset, we allocated 70% to training and the residual 30% to testing. This ensures that while a significant part of the data is utilized for training, a distinct segment is retained for independent evaluation, providing a genuine gauge of our model's prediction proficiency.

4.3 Modeling

4.3.1 Random Forest

In our investigation, we utilized the Random Forest classifier [25], a versatile ensemble learning method, and optimized its hyperparameters through GridSearchCV from the sklearn library. This optimization involved a systematic search over various parameter combinations, such as the number of trees, tree depth, minimum samples required for splitting, minimum samples at leaf nodes, and feature selection criteria, with a 3-fold cross-validation to ensure robustness. Upon completion of the grid search, the optimal parameters for our dataset were identified as: max_depth of 20, max_features set to 'log2', min_samples_leaf of 4, min_samples_split of 2, and n_estimators of 50.

4.3.2 Gradient Boosting

For our study, we employed the Gradient Boosting classifier [26], an ensemble learning algorithm that builds on weak learners to improve accuracy. To optimize its hyperparameters, we made use of GridSearchCV, exploring different combinations of parameters such as the number of boosting stages, learning rate, depth of the trees, minimum samples required for both splitting and leaf nodes, and the fraction of samples used for fitting each tree, all within a 3-fold cross-validation framework. Post grid search, the best parameters identified for our dataset were: a learning_rate of 1, max_depth set at 3, min_samples_leaf of 1, min_samples_split of 4, n_estimators of 100, and a sub-sample rate of 0.8.

4.3.3 Logistic Regression

Upon executing hyperparameter tuning for the Logistic Regression model using GridSearchCV, we sought to refine this well-established algorithm best suited for binary classification problems. Logistic Regression [27] operates by estimating probabilities using a logistic function, thereby ensuring results are within the range of 0 and 1. The optimal settings tailored to our dataset were determined to be: a regularization strength of 1, inclusion of an intercept term, l1 penalty, a maximum of 100 iterations, and the use of the liblinear solver. This optimization ensured robust performance tailored specifically to the underlying data structure.

5 Results

5.1 Random Forest Results

For the optimized Random Forest model, the evaluation metrics presented an impressive performance. The model boasted an accuracy of approximately 97.67%, which points to a substantial rate of correct predictions among all the predictions made. Additionally, the sensitivity (or true positive rate) stood at approximately 98.05%, which means the model correctly identified 98.05% of the positive instances. On the other hand, the specificity (or true negative rate) was approximately 96.58%, indicating the model's proficiency in correctly identifying the negative cases. The confusion matrix (Fig. 2), which breaks down the model's performance in a detailed manner, reported 141 true positives, 152 true negatives, 5 false positives, and 2 false negatives. As for precision and recall:

- For class 0, the precision is 0.99 and recall is 0.97, leading to an F1-score of 0.98. This suggests that the model is highly accurate in predicting this class and has a minimal rate of false negatives.
- For class 1, the model has a precision of 0.97 and a recall of 0.99, resulting in an F1-score of 0.98. This indicates the model's strong ability to correctly identify instances of this class while keeping the false positive rate low.

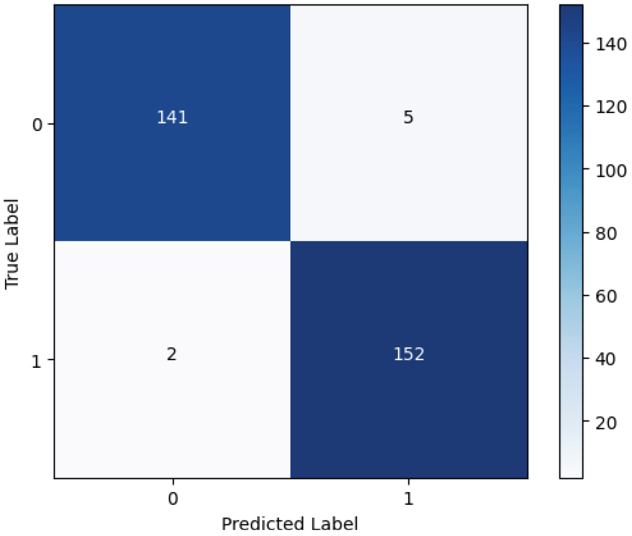


Fig. 2. Confusion matrix of Random Forest.

5.2 Gradient Boosting Results

The Gradient Boosting model, another ensemble technique, displayed a robust predictive capability. The model achieved an accuracy of approximately 96.67% highlighting its capacity to produce a high rate of correct predictions. In terms of sensitivity (or true positive rate), the model showcased an impressive 97.40%, signifying its strength in correctly identifying positive instances. The specificity (or true negative rate) was approximately 95.89%, indicating a high ability of the model to accurately identify negative cases without many false alarms. From the confusion matrix We observed 140 true positives, 150 true negatives, 6 false positives, and 4 false negatives. In terms of precision and recall (Fig. 3):

- For class 0, the model achieved a precision of 0.97 and a recall of 0.96, leading to an F1-score of 0.97. This highlights the model's proficient accuracy in predicting this class while maintaining a relatively low rate of false negatives.
- For class 1, the precision was 0.96 and the recall stood at 0.97, culminating in an F1-score of 0.97. This underscores the model's ability to make accurate predictions for this class with a minimal false positive rate.

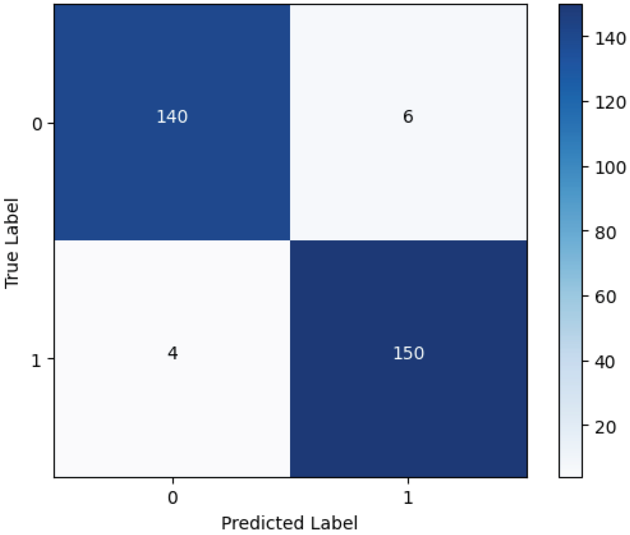


Fig. 3. Confusion matrix of Gradient Boosting.

5.3 Logistic Regression Results

The Logistic Regression model, a fundamental yet robust linear classification technique, showcased impressive results on the dataset. The model clinched an accuracy of roughly 97.33%, indicative of its high proficiency in predictions. From the provided confusion matrix, there were 143 true positives and 149 true negatives, while the instances of false positives and false negatives were minimal, with counts of 3 and 5 respectively. When we break down the precision and recall (Fig. 4):

- For class 0, the model demonstrated a precision of 0.97 and a recall of 0.98, leading to a harmonized F1-score of 0.97. This represents the model's efficiency in predicting this class with high accuracy and minimal false negatives.
- Conversely, for class 1, both precision and recall were outstanding, with values of 0.98 and 0.97 respectively. The resulting F1-score of 0.97 echoes the model's adeptness in predicting this class with few false positives.

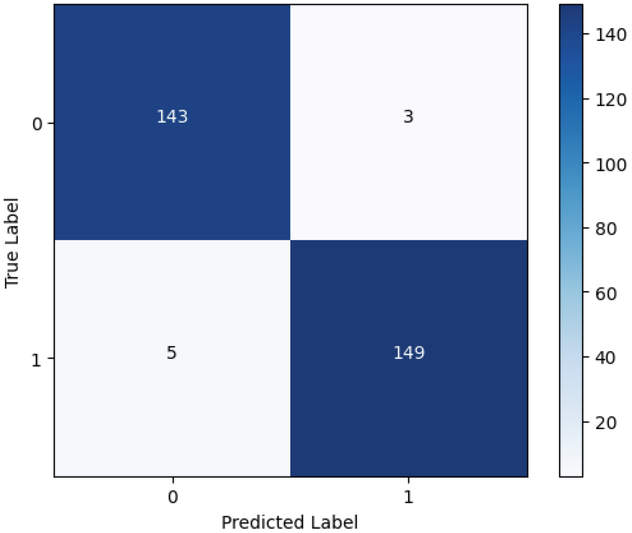


Fig. 4. Confusion matrix of Logistic Regression.

5.4 Comparison

Across all three models (Table **??**), the performance was exemplary, with the Random Forest achieving the highest accuracy at 97.67%, followed closely by Logistic Regression at 97.33%, and Gradient Boosting at 96.67%. The differences in accuracy are minimal, signifying that all three models were effective in generalizing the patterns in the data. While all three models demonstrated impressive performance metrics, the Random Forest slightly edged out in most categories, especially in sensitivity. The Logistic Regression also put forth commendable specificity and precision for Class 1. The Gradient Boosting model, while slightly lagging behind in sensitivity, still maintained competitive metrics across other categories. Making a choice among the three would depend on the specific use case and the importance of each metric for the application at hand.

5.5 Interpretative Analysis of Random Forest Predictions

In light of the superior performance metrics exhibited by the Random Forest model, we have chosen it for a deeper interpretative analysis. To demystify the intricate decision-making processes within this ensemble model and to gain more insights into the feature influences, we will utilize two prominent model interpretation tools: SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations). Both SHAP and LIME are renowned for their capability to provide transparent and understandable explanations, even for complex models like Random Forest. This interpretative analysis aims not only to validate our model's decisions but also to shed light on the primary drivers behind its predictions, enabling more informed decision-making in practical applications.

In the visual interpretations derived from both LIME (Fig. 5) and SHAP (Fig. 6) methodologies, the features "Daily Internet Usage", "Daily Time Spent on Site", "Area Income", "Age", and "Male" prominently emerged. These features play a pivotal role in the model's decision-making process, highlighting their significance in influencing the model's predictions.

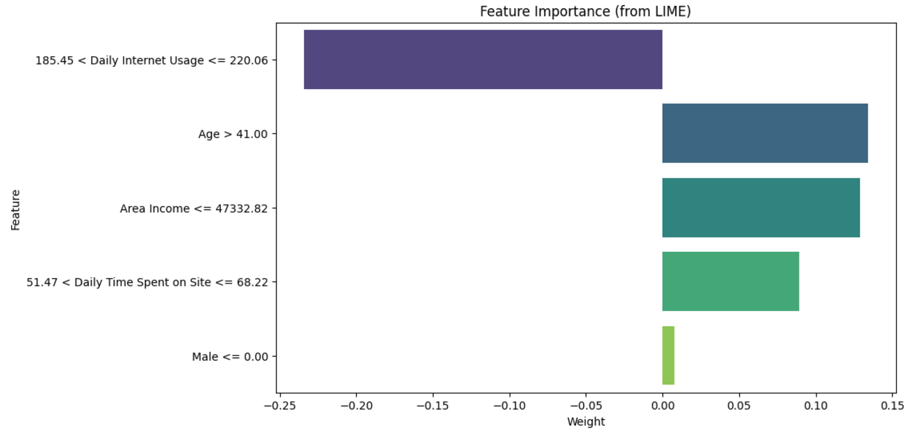


Fig. 5. Visual Interpretation using LIME technique.

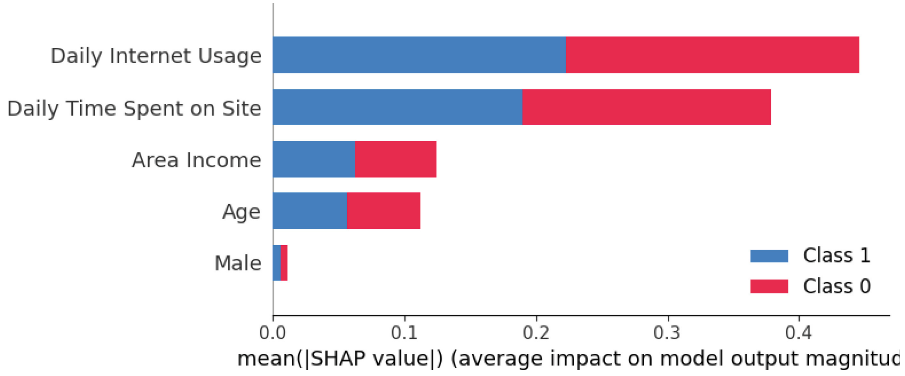


Fig. 6. Visual Interpretation using SHAP technique.

6 Conclusion and Future Work

The inexorable evolution of online advertising demands an equally progressive approach to understanding and predicting user behavior. This study successfully demonstrated the application of machine learning techniques in analyzing and predicting user interactions with online advertisements. Our comprehensive approach, encompassing models like Random Forest, Gradient Boosting, and Logistic Regression, provided insights with exceptional accuracy, further accentuated by the integration of explainability tools like SHAP and LIME. Our results, punctuated by an accuracy of 97.67% for the Random Forest model, reaffirm the potential of machine learning in harnessing the vast and intricate datasets generated in the realm of online advertising. Yet, beyond sheer accuracy, this study underscores the paramount importance of model transparency and interpretability. In a domain where trust can significantly influence user behavior, making machine learning models comprehensible to stakeholders ensures that advertising recommendations are not only effective but also trustworthy.

As we reflect on the findings of this research, it is essential to recognize that the digital realm is an ever-evolving landscape, and there remains a multitude of avenues to explore in future studies.

- As the digital landscape continually evolves, future iterations of our work should incorporate models that adapt in real-time, learning from newer data to remain relevant and effective.
- In the wake of multi-device internet access, understanding user behavior across various devices like smartphones, tablets, and desktops can offer richer insights, necessitating models that can handle such diversified data.

In conclusion, while this study provides significant strides in understanding online user behavior through machine learning, the ever-evolving nature of the digital realm means that this is but a stepping stone. The future beckons with newer challenges and opportunities, and harnessing them effectively will require a blend of technological prowess, ethical considerations, and a deep understanding of the human element.

References

1. Chibudike, C., Abdu, H., Chibudike, H., Ngige, O., Adeyoju, O., Obi, N.: Machine learning - a new trend in web user behavior analysis. Int. J. Comput. Appl. **183**, 19–25 (2021)
2. Dixon, S.: Facebook: Global daily active users 2022. https://www.statista.com/statistics/346167/facebook-global. Statista (2023)
3. Statista: Social media advertising expenditure as share of digital advertising spending worldwide from 2013 to 2017 (2017). https://www.statista.com/statistics/271408/share-of-social-media-in-online-advertising-spending-worldwide/
4. Valdez, D., Ten Thij, M., Bathina, K., Rutter, L.A., Bollen, J.: Social media insights into us mental health during the covid-19 pandemic: longitudinal analysis of twitter data. J. Med. Internet Res. **22**(12), e21418 (2020)
5. Lee, H., Cho, C.-H.: Digital advertising: present and future prospects. Int. J. Advert. **39**(3), 332–341 (2020)
6. Statista. Global advertising spending from 2010 to 2017 (in billion u.s. dollars) (2017). https://www.statista.com/statistics/236943/global-advertising-spending/

7. Bayer, S., Gimpel, H., Markgraf, M.: The role of domain expertise in trusting and following explainable AI decision support systems. J. Decis. Syst. **32**, 110–138 (2021)
8. Gramegna, A., Giudici, P.: Why to buy insurance? An explainable artificial intelligence approach. Risks **8**(4), 137 (2020)
9. Gramegna, A., Giudici, P.: Shap and lime: an evaluation of discriminative power in credit risk. Front. Artif. Intell. **4**, 752558 (2021)
10. Lundberg, S.M., Lee, S.-I.: A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
11. Ribeiro, M.T., Singh, S., Guestrin, C.: Why should I trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144 (2016)
12. He, X., et al.: Practical lessons from predicting clicks on ads at Facebook. In: Proceedings of the Eighth International Workshop on Data Mining for Online Advertising, pp. 1–9 (2014)
13. Chaudhary, K., Alam, M., Al-Rakhami, M.S., Gumaei, A.: Machine learning-based mathematical modelling for prediction of social media consumer behavior using big data analytics. J. Big Data **8**(1), 1–20 (2021)
14. Almeida, P.S., Gondim, J.J.: Click fraud detection and prevention system for ad networks. J. Inf. Secur. Crypt. (Enigma) **5**(1), 27–39 (2018)
15. Lipyanina, H., Sachenko, A., Lendyuk, T., Nadvynychny, S., Grodskyi, S.: Decision tree based targeting model of customer interaction with business page. In: CMIS, pp. 1001–1012 (2020)
16. Choudhury, M., Counts, S., Horvitz, E.: Predicting postpartum changes in emotion and behavior via social media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 3267–3276 (2013)
17. Tago, K., Takagi, K., Kasuya, S., Jin, Q.: Analyzing influence of emotional tweets on user relationships using naive bayes and dependency parsing. World Wide Web **22**(3), 1263–1278 (2019)
18. Aragon, M., López-Monroy, A., González-Gurrola, L., Montes, M.: Detecting depression in social media using fine-grained emotions. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, vol. 1 (Long and Short Papers), pp. 1481–1486 (2019)
19. Vedula, N., Parthasarathy, S.: Emotional and linguistic cues of depression from social media. In: Proceedings of the 2017 International Conference on Digital Health, pp. 127–136 (2017)
20. Choudhury, M., Counts, S., Horvitz, E.: Social media as a measurement tool of depression in populations. In: Proceedings of the 5th Annual ACM Web Science Conference, pp. 47–56 (2013)
21. Peng, Z., Hu, Q., Dang, J.: Multi-kernel svm based depression recognition using social media data. Int. J. Mach. Learn. Cybern. **10**(1), 43–57 (2019)
22. Alicioglu, G., Sun, B.: A survey of visual analytics for explainable artificial intelligence methods. Comput. Graph. **102**, 502–520 (2022)
23. Arrieta, A.B., et al.: Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion **58**, 82–115 (2020)
24. Zahavy, T., Ben-Zrihem, N., Mannor, S.: Graying the black box: understanding DQNs. In: International Conference on Machine Learning, pp. 1899–1908. PMLR (2016)
25. Ho, T.K.: Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition, vol. 1, pp. 278–282. IEEE (1995)
26. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232 (2001)
27. Cox, D.R.: The regression analysis of binary sequences. J. Roy. Stat. Soc.: Ser. B (Methodol.) **20**(2), 215–232 (1958)

CryptStego: Powerful Blend of Cryptography and Steganography for Securing Communications

Shraiyash Pandey[1](✉), Pashupati Baniya[1], Parma Nand[1], Alaa Ali Hameed[2], Bharat Bhushan[1], and Akhtar Jamil[3]

[1] Department of Computer Science and Engineering School of Engineering and Technology, Sharda University, Greater Noida, India
shraiyash.pandey@gmail.com, parma.nand@sharda.ac.in

[2] Department of Computer Engineering, Istinye University, Istanbul, Turkey
alaa.hameed@istinye.edu.tr

[3] Department of Computer Science, National University of Computer and Emerging Sciences, Islamabad, Pakistan
akhtar.jamil@nu.edu.pk

Abstract. In today's era, security is one of the most critical issues in the development of electronic communications applications, especially when sending private data. The data may be encrypted with several algorithms; however, an extra layer of security can improve protection by a significant amount. Therefore, in this paper, we have developed an application, CryptStego, to secure data using two techniques, cryptography and steganography, to transmit data securely. The encryption of original data is executed using Blowfish algorithm, a cryptographic technique. Additionally, the encrypted data is hidden through using Least Significant Bit (LSB), a steganography technique. The implementation of both techniques offers an extensive level of security, since an intruder must firstly identify the encrypted text within the image to attain the encrypted text, then secondly to decrypt using the algorithm to obtain the original message. Therefore, any intruder must encounter multiple levels of security to obtain the original message from the cipher image.

Keywords: Web-Application · Cryptography · Blowfish Algorithm · Steganography · Least Significant Bit (LSB)

1 Introduction

Nowadays, an internet connection is required for an increasing number of human activities. Humans are able to obtain data and information thanks to the internet. Large amounts of data and information must be stored as time passes. To preserve data integrity, an authentication system is required. This increases the need for a robust information safety system to protect data from potential attacks. As a result, study on information safety systems and their execution remains to be developed in order to secure data carried over a network of communications [1]. Privacy is the most pressing issue that individuals confront. Therefore, we introduce CryptStego, a strong combination of cryptography

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 532–543, 2024.
https://doi.org/10.1007/978-3-031-56728-5_44

and steganography, to render data unreadable by other parties and disguise the fact that data has been received from third parties.

Cryptography is a foundation of privacy and security, including a wide range of approaches for safeguarding data transmission and storage. The basic goal of cryptography is to encrypt information in such a way that only approved receivers may decode and comprehend it. In symmetric key encryption, the same keys are used to encrypt and decrypt information. Only one key is utilized to encrypt and decode data in this symmetric system. That is, the same unique key is utilized for both encryption and decryption, and because this key is known to both the sender and the receiver, it remains hidden [2]. Because the key employed here is shorter in length, data processing is simple. Therefore, we implement the F-function of Blowfish algorithm in our encryption standard for data. Asymmetric encryption, commonly referred to as public key cryptography, employs a pair of keys, one for encryption and the other for decryption. The secret key and the public key are the names given to these two keys. Data is encrypted using the receiver's public key, which is available to everyone, and decrypted using the receiver's private key, which is known only to the receiver [3]. As a result, security is increasingly important in these types of systems. However, cryptography does not conceal the data; rather, it concentrates on developing a robust encryption system to protect it.

Steganography, on the other hand, supports encryption by allowing covert communication by concealing data within harmless carriers such as photos, audio files, or text [4]. Unlike cryptography, which focuses on rendering signals unreadable. Steganography promotes transmission secrecy, allowing data to remain concealed in plain sight. Least Significant Bit (LSB) replacement technique is implemented in order to hide the encrypted data in an image that generates a cipher image. It is a popular technique that involves substituting the least significant bits of a pixel's value of color in a picture or audio file with concealed data [5]. This change is unnoticeable to the human sight or ear, yet it may transmit significant quantities of hidden information. Spread spectrum steganography spreads concealed data across numerous carrier components. The ensuing modifications are modest and difficult to detect, ensuring that information is sent covertly. The major contributions of the work are enumerated below.

- This work presents the methodology of a web application built with Python, and powered by Flask with the goal of securing digital landscape and data transmission.
- This work presents an overview of the various technologies implemented in the development process of CryptStego such as Python, Flask, etc.
- This work provides an in-depth review of the cryptography and steganography methods implemented in the web application.

The remainder of the paper is organized as follows. Section 2 presents the literature review of all existing technologies with similar end goal but different approaches and implementations. Section 4 explains the methodology of the developed project through flow charts and theoretical concepts. Section 5 presents the final results obtained through the development of CryptStego. Finally, Sect. 6 presents the conclusions followed by future research directions.

2 Literature Review

Islam et al. [6] presented a new data hiding method using LSB image steganography. In this method, confidential information is embedded using only selected image pixels. The image pixel information is employed to filter the complete image, determining the candidate pixel for embedding. Additionally, a user-defined password enhances the security of the LSB steganography process. Tabassum et al. [7] introduced a method that integrates Steganography and Cryptography. This method utilizes the Blowfish algorithm and the Advanced Encryption Standard algorithm, in addition to leveraging features of the residue numbering system, the Least Significant Bit algorithm, and the operators of the Genetic Algorithm.

Rashmi et al. [8] introduced the idea of securing digital data in real-time applications by ensuring authenticity, preventing modification, and restricting access using techniques like cryptography and steganography. They specifically focused on audio steganography and aimed to combine both methods for enhanced data protection. Gladwin et al. [9] introduced a robust and effective algorithm based on elliptic curve cryptography combined with Hill cipher to mitigate such threats and enhance information security. Adithya et al. [10] investigated recently published DNA steganography algorithms, which utilize DNA to encrypt confidential data transmitted through an insecure communication channel.

Zeena et al. [11] introduced a novel method for encrypting vital data using circular shape information extracted from the cover image, employing updated traditional techniques to enhance security and confidentiality. Denis et al. [12] presented an efficient Visually Imperceptible Hybrid Crypto-Steganography (VIHCS) model developed using Hybrid Cryptosystems followed by an Adaptive Genetic Algorithm assisted Least Significant Bit (LSB) embedding process. Hammad et al. [13] investigated enhancing information security by combining Vigenère and Caesar ciphers, converting them to chemical element names, and using steganography for robust encryption.

Antonio et al. [14] introduced a method to enhance information security through combined bit matching steganography and AES cryptography, focusing on pixel location, key generation, AES encryption, speed, high payload capacity, and countering statistical steganalysis. Pabbi et al. [15] introduced a method aiming to enhance message security via steganography, concealing plain text within images using LSB steganography and AES encryption for improved confidentiality during transmission. Malak et al. [16] investigated the amalgamation of lightweight cryptography and enhanced Arabic text steganography for heightened data security, highlighting the concealment of encrypted secrets using diacritics and providing simulations for comparison.

Patil et al. [17] investigate data security concerns and discuss using Cryptography and Steganography as techniques for encrypted and concealed data protection. Rajesh et al. [18] investigated the issue of data breaching during information sharing by utilizing cryptography and steganography. They discussed a novel approach that involves incorporating Huffman coding within Image Steganography to enhance security and optimize data transfer. Krishna et al. [19] presented a study that investigates the enhancement of data security for internet communication through the amalgamation of Cryptography and

Image Steganography. The authors analyzed the utilization of XOR encryption and user-chosen key-based pseudo-random embedding to achieve this goal. The previous contributions to the relevant field of applications based on cryptography and steganography are presented in Table 1 below.

Table 1. Literature Review

References	Proposed Work
Islam et al. [6]	Introduced novel LSB image steganography for data hiding
Tabassum et al. [7]	Introduced integrated Steganography and Cryptography technique
Rashmi et al. [8]	Enhanced real-time data security with cryptography and steganography for authenticity and restriction
Gladwin et al. [9]	Introduced robust elliptic curve-Hill cipher algorithm for enhanced information security and threat mitigation
Adithya et al. [10]	Explored DNA steganography for encrypting confidential data in insecure channels
Zeena et al. [11]	Innovated circular shape encryption with traditional methods for enhanced data security and confidentiality
Denis et al. [12]	Introduced efficient VIHCS model combining Hybrid Cryptosystems, Genetic Algorithm, and LSB embedding
Hammad et al. [13]	Studied security enhancement using Vigenère, Caesar ciphers, and steganography with chemical element names
Antonio et al. [14]	Proposed security enhancement: combined bit matching steganography, AES cryptography, emphasizing pixel location, key generation, speed, and payload capacity
Pabbi et al. [15]	Introduced method for enhanced message security: concealing text in images with LSB steganography and AES encryption
Malak et al. [16]	Studied lightweight cryptography, enhanced Arabic steganography for heightened data security, emphasizing diacritic-concealed encryption, and comparative simulations
Patil et al. [17]	Explored data security concerns, discussed Cryptography and Steganography for encrypted data protection
Rajesh et al. [18]	Investigated the issue of data breaching during information sharing by utilizing cryptography and steganography
Krishna et al. [19]	Presented a study that investigates the enhancement of data security for internet communication through the amalgamation of Cryptography and Image Steganography

3 Technologies Implemented

There's a wide range of technologies used in the development of the project. Python is used to implement the computations of the algorithm for cryptography technique, and accessing RGB values of the image to extract least significant bits then replace with cipher text bits for steganography technique. Flask is used to utilize its various API's that support flexible web development. HTML is used to build dynamic structure of different web pages throughout the website. CSS is used to emphasize user-friendly attributes in our website, and give an overall better look. JavaScript is used on the project to program the behavior of the website. Each technology is discussed in the subsections below.

3.1 Python

Python is a versatile and widely used programming language with applications spanning web and software development, mathematical computations, system scripting, and more. It excels in server-side web development, integrating with frameworks like Django and Flask for robust applications. Python's strength extends to software development, scientific research, and data analysis, facilitated by libraries such as NumPy and pandas. Its automation abilities are evident in system scripting, including interactions with databases, while its support for complex mathematics aids fields like engineering and machine learning. Python's cross-platform compatibility, user-friendly syntax, and support for varied programming paradigms contribute to its popularity, enabled by a vast library ecosystem [20].

3.2 Flask

Flask, a lightweight yet powerful Python web framework, streamlines and supports flexible web app development. It offers structured tools and conventions through an API, prioritizing minimalism and explicit control over components. Unlike Django, it's ideal for smaller projects due to its lean foundation, and its small codebase aids newcomers while maintaining capability. As a web app framework, Flask abstracts low-level tasks with pre-built modules, simplifying HTTP requests, sessions, and databases for unique app development. Built on the WSGI toolkit, Flask ensures server compatibility, while its Jinja2 template engine cleanly separates logic and presentation for dynamic web pages. In conclusion, Flask's explicit, minimalistic approach fosters dynamic web apps with essential tools, a gentle learning curve, and streamlined elegance, supported by WSGI and Jinja2 integration [21].

3.3 HTML

HTML, which stands for Hypertext Markup Language, serves as a foundational language for crafting the structure and content of web pages. It constitutes the backbone of any web document by offering a standardized approach to defining elements and their interrelationships within a webpage. Essentially, HTML is the essential building block that underpins website construction, playing a pivotal role in the presentation of text, images, links, forms, and diverse media formats across the internet. This language delineates a web page's structure through a series of elements, each conveying specific instructions to the browser on how to render content. These elements label distinct content components, such as headings, paragraphs, and links, thus facilitating cohesive and structured web content presentation [22].

3.4 CSS

CSS, a cornerstone of web development, empowers designers and developers to control web page aesthetics by instructing how HTML content should be presented. Collaborating with HTML, it defines styles encompassing fonts, colors, spacing, and positioning, maintaining a separation of content and presentation. While HTML structures content, CSS customizes visuals, catering to headings, paragraphs, links, and more with features like color, font, alignment, and borders. Employing a cascading approach, CSS enforces rules in order of specificity, with inheritance dictating child elements' styles inherited from parents. External.css files streamline consistency across pages, while CSS also plays a pivotal role in shaping web layout, facilitating adaptable grids, intricate designs, and responsiveness through media queries for diverse devices. Although widely compatible, browser discrepancies require testing for uniformity. Essentially, CSS empowers developers to finesse HTML element appearance, elevating code clarity, maintenance, and overall uniformity through stylesheets and cascading principles [23].

3.5 JavaScript

JavaScript is a widely used programming language that is primarily known for its role in building dynamic and interactive web applications. It is a versatile language that can be executed in web browsers, making it a crucial tool for creating client-side functionality on websites. JavaScript can also be used on the server-side, thanks to technologies like Node.js, to build scalable and efficient server applications. It is the programming language of the Web and is easy to learn. JavaScript is used to program the behavior of web pages. It is one of the 3 languages that all web developers must learn [24].

4 Methodology

Researchers have worked on numerous cryptography and steganography techniques. However, we have attempted to bring the combination of both approaches, namely cryptography and steganography, in a novel way. Our proposed methodology includes of implementing the F-function of the blowfish algorithm to encrypt the original data, then through the help of LSB algorithm, hide the data within an image. Then, that image can be securely sent through any communication channel. The proposed methodology is shown in Fig. 1 below.

Here, the initial message is encrypted into ciphertext using the suggested symmetric cryptography technique, the F-function of the blowfish algorithm, which divides the 128 bits into four equal parts of 32 bits each. Each 32-bit block is then subjected to various

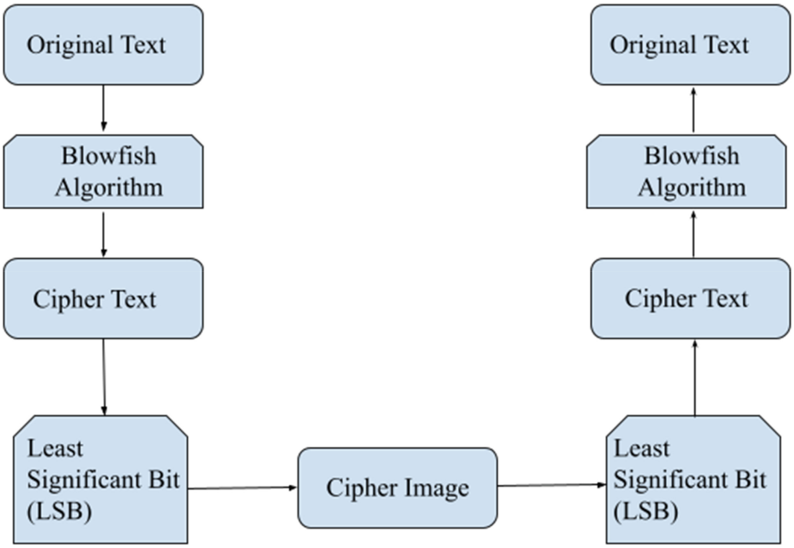


Fig. 1. Proposed Methodology

circular shift and XOR operations using secret keys. The algorithm for encryption and decryption using Blowfish algorithm is shown below.

Algorithm 1 Blowfish Algorithm Encryption

```
Input: M is the original message
Output: C, Ciphertext
1. Split the plain text 64-bit blocks: L0, R0
2. For each round i = 1 to 16:
3. $L(i) = R(i-1)$
4. $R(i) = L(i-1)\ XOR\ F(R(i-1), Subkey[i])$
5. Swap the final halves: L16, R16.
6. XOR L16, R16 with final P-Array value.
7. $C = (L16 \ll 32)\ OR\ R16$
```

Algorithm 2 Blowfish Algorithm Decryption

```
Input: C, Ciphertext
Output: M, the original message
Split ciphertext into 64-bit blocks: L16, R16.
XOR L16, R16 with initial P-Array value.
For each round i = 16 to 1:
    $R(i) = L(i - 1)$
    $L(i) = R(i)\ XOR\ F(L(i), Subkey[i])$
Swap the final halves: L0, R0.
XOR L0, R0 with initial P-Array value.
$M = (L0 \ll 32)\ OR\ R0$
```

After the encryption process for the initial data is completed and the ciphertext is generated, the process of steganography is initiated. The ciphertext is hidden in a Cover Image using the suggested steganography approach, which employs least significant bit (LSB) where a mixture of LSB-1, LSB-2, and LSB-3 is used alternatively. This produces our cipher image that is delivered over the communication channel, and the identical process occurs at the receiver side, but in the reverse order, beginning with extracting encoded material from cipher image and then decrypting using the suggested decryption method. The algorithm for encryption and decryption using LSB is shown below.

Algorithm 3 LSB Encryption

```
Input: C, the original image, and D, the binary data
Output: S, Cipher Image
For each pixel P (x, y) in C:
    Retrieve the RGB values: R, G, and B
    For each bit b in D:
        Retrieve the least significant bit of the color channel (R, G, or B) to be used
        Replace the least significant bit with the current bit b from D
        Move to the next bit in D
    If all bits in D are not embedded, move to the next pixel
    If all bits in D are embedded, break
S is obtained
```

Algorithm 4 LSB Decryption

```
Input: S, Cipher image
Output: D, the encrypted text
1. Initialize an empty binary data container D
2. For each pixel P (x, y) in S:
3. Retrieve the significant bit of each color channel (R, G, B)
4. Append the retrieved bits to D
5. If the length of D matches the length of the embedded data, break
6. D is obtained
```

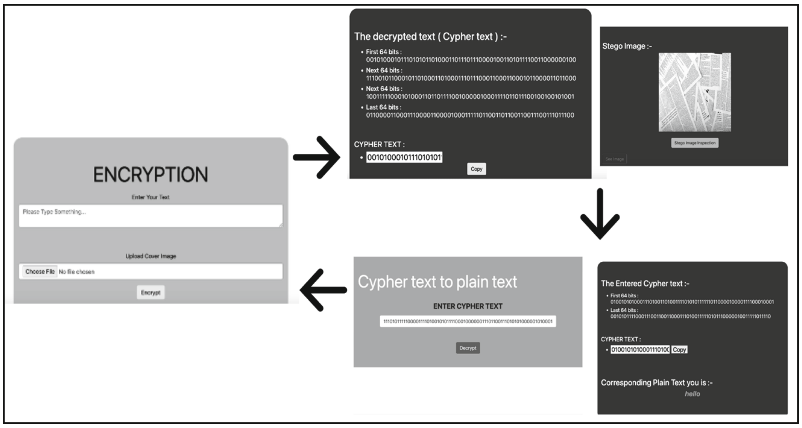
5 Results and Discussion

The initial home page of the web application, CryptStego, encompasses a user-friendly interface, presented by the welcoming text displayed at the very top of the home page. The main page also highlights the main purpose of the website through the help of a grey-colored rectangle displaying the text, "Want to Encrypt Text?". The navigation bar incorporates of different navigation routes to different pages such as sign-up, login, developer's info, search bar, and home.

The entire project consists of two main functions, encryption and decryption. The encryption of data is done via the help of encrypter that can be accessed after clicking on the "Encrypter" button situated at the home page. The encrypter takes in two parameters, the original text that needs to encrypted, and the image that needs to be encrypted through LSB to perform steganography.

Further, after the encryption is completed, the website is automatically navigated to the page that outputs the cipher text, and the image with hidden cipher text. Nonetheless, at last, the data is decrypted using the "Decrypter" which can be accessed below the page that outputs the cipher text and image. The entire process can be described in just three steps, shown in Fig. 2.

Firstly, the encrypter encrypts the original message using F-function of the blowfish algorithm, and outputs a ciphertext. Secondly, the encrypter also takes a cover image as an input, and hides the encrypted text generated within the image using LSB. Lastly, the entire process can be reversed in order to obtain the original message. The least significant bits of the cipher image are identified until the length of the encrypted text is obtained, then using the decryption algorithm, the original message will be accessed. Any intruder may only obtain the original message from the cipher image if he or she decrypts by identifying the least significant bits, and is given the length of the encrypted text. Otherwise, the total number of least significant bits can or cannot be equal. Therefore, it requires internal information in order to obtain the original message from the cipher image.



ENCRYPTION

Enter Your Text

Please Type Something...

Upload Cover Image

The decrypted text (Cypher text) :-

- First 64 bits : 0010100010111010101101000110111011100001001101011110011000000100
- Next 64 bits : 1110010110001011010001101000111011100011000110001011000011011000
- Next 64 bits : 1001111100010100011011011110010000010001111011011100100100101001
- Last 64 bits : 0110000110001110000110000100011111011001101100110011100111011100

CYPHER TEXT :

- 0010100010111010101

Stego Image :-

Cypher text to plain text

ENTER CYPHER TEXT

1110101111100001111010010101111000100000011101100111010101000001010001

The Entered Cypher text :-

- First 64 bits : 0100101010001110100110100111101010111111011000010000111100010001
- Last 64 bits : 0010101111000111001100110001110100111101011100000100111110111110

CYPHER TEXT :

- 010010101000111010(

Corresponding Plain Text you is :-

hello

Fig. 2. Process of applying cryptography and steganography on the data

6 Conclusion and Future Research Directions

The crucial importance of data security has expanded enormously in today's rapidly changing environment. The necessity to secure sensitive information has grown critical as the digital transition shapes our communication landscape. The transmission of private information through electronic mediums has increased the importance of ensuring information's secrecy, integrity, and validity. As a result, adopting effective security strategies is critical to maintaining trust, privacy, and the integrity of digital interactions. In this light, this paper introduces CryptStego, an integration of cryptography and steganography that provides a strong security solution for data transfer. The F-function of the Blowfish algorithm is used for the original data's encryption. The Blowfish algorithm protects the data throughout transmission from unauthorized access. CryptStego, combined with cryptography, employs the Least Significant Bit (LSB) steganography approach to further strengthen the security. LSB allows for the seamless hiding of secret information within pictures, a covert process that adds another layer of deception to the sent data. The program displays its ability to prevent unauthorized access, data breaches that strike unprotected networks by combining the strength of cryptography with steganography.

Further research can be done towards incorporating a video in the proposed methodology to enhance the security. The video may require extra set of resources for the device that implements CryptStego, however, it's may be possible to reduce the strength of rounds for encryption to fulfill the resource constraints. A more complex algorithm such as RSA, and ECC to significantly increase the level of security, but again with only respect to modifying other areas to not significantly increase the resource requirements unless not a constraint.

References

1. Bhushan, B., Sahoo, G.: Recent advances in attacks, technical challenges, vulnerabilities and their countermeasures in wireless sensor networks. Wirel. Pers. Commun. **98**(2), 2037–2077 (2017). https://doi.org/10.1007/s11277-017-4962-0
2. Sinha, P., Rai, A.K., Bhushan, B.: Information Security threats and attacks with conceivable counteraction. In: 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT) (2019). https://doi.org/10.1109/icicict46008.2019.8993384
3. Setiadi, D.R.I.M., Jumanto, J.: An enhanced LSB-image steganography using the hybrid canny-sobel edge detection. Cybern. Inf. Technol. **18**(2), 74–88 (2018)
4. Almalki, K.A., Mohammed, R.: A novel steganography approach to embed secret information into a legitimate URL. In: 2022 2nd International Conference on Computing and Information Technology (ICCIT), Tabuk, Saudi Arabia, pp. 180–185. (2022). https://doi.org/10.1109/ICCIT52419.2022.9711647
5. Hasan Talukder, M.S., Hasan, M.N., Sultan, R.I., Rahman, M., Sarkar, A.K., Akter, S.: An enhanced method for encrypting image and text data simultaneously using AES algorithm and LSB-based steganography. In: 2022 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE), Gazipur, Bangladesh, pp. 1–5 (2022). https://doi.org/10.1109/ICAEEE54957.2022.9836589
6. Islam, M.R., Tanni, T.R., Parvin, S., Sultana, M.J., Siddiqa, A.: A modified LSB image steganography method using filtering algorithm and stream of password. Inf. Secur. J. Global Perspect. **30**(6), 359–370 (2020). https://doi.org/10.1080/19393555.2020.1854902
7. Tabassum, T., Mahmood, M.A.: A multi-layer data encryption and decryption mechanism employing cryptography and steganography. In: 2020 Emerging Technology in Computing, Communication and Electronics (ETCCE), Bangladesh, pp. 1–6 (2020). https://doi.org/10.1109/ETCCE51779.2020.9350908
8. Rashmi, N.: Analysis of audio steganography combined with cryptography for RC4 and 3DES encryption. In: 2020 Fourth International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, pp. 96–100 (2020). https://doi.org/10.1109/ICISC47916.2020.9171215
9. Gladwin, S.J., Lakshmi Gowthami, P.: Combined cryptography and steganography for enhanced security in suboptimal images. In: 2020 International Conference on Artificial Intelligence and Signal Processing (AISP), Amaravati, India, pp. 1–5 (2020). https://doi.org/10.1109/AISP48273.2020.9073306
10. Adithya, B., Santhi, G.: DNA computing using cryptographic and steganographic strategies. Data Integrity Qual. (2021a). https://doi.org/10.5772/intechopen.97620
11. Al-Kateeb, Z.N., Al-Shamdeen, M.J., Al-Mukhtar, F.S.: Encryption and steganography a secret data using circle shapes in colored images. J. Phys: Conf. Ser. **1591**(1), 012019 (2020). https://doi.org/10.1088/1742-6596/1591/1/012019
12. Denis, R., Madhubala, P.: Evolutionary computing assisted visually-imperceptible hybrid cryptography and steganography model for secure data communication over cloud environment. Int. J. Comput. Netw. Appl. **7**(6), 208 (2020). https://doi.org/10.22247/ijcna/2020/205321
13. Hammad, R., et al.: Implementation of combined steganography and cryptography Vigenere cipher, Caesar cipher and converting periodic tables for securing secret message. J. Phys. Conf. Ser. **2279**(1), 012006 (2022). https://doi.org/10.1088/1742-6596/2279/1/012006
14. Antonio, H., Prasad, P.W., Alsadoon, A.: Implementation of cryptography in steganography for enhanced security. Multimedia Tools Appl. **78**(23), 32721–32734 (2019). https://doi.org/10.1007/s11042-019-7559-7

15. Pabbi, A., Malhotra, R., Manikandan, K.: Implementation of least significant bit image steganography with advanced encryption standard. In: 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, pp. 363–366 (2021). https://doi.org/10.1109/ESCI50559.2021.9396884
16. Alkhudaydi, M.G., Adnan, A.A.: Integrating light-weight cryptography with diacritics Arabic text steganography improved for practical security applications. J. Inf. Secur. Cybercrimes Res. **3**, 13–30 (2020). https://doi.org/10.26735/FMIT1649
17. Patil, A., Mulik, S., Pathak, P., Raut, K.: Review paper on data security using cryptography and steganography. VIVA-IJRI **1**(4), Article 59, 1–6 (2021)
18. Rajesh, P., Alam, M., Tahernezhadi, M., Ravi, T., Phaneendra, V.: Secure communication across the internet by encrypting the data using cryptography and image steganography. Int. J. Adv. Comput. Sci. Appl. **11**(10) (2020). https://doi.org/10.14569/ijacsa.2020.0111057
19. Nunna, K.C., Marapareddy, R.: Secure data transfer through internet using cryptography and image steganography. In: 2020 SoutheastCon, Raleigh, NC, USA, pp. 1–5 (2020). https://doi.org/10.1109/SoutheastCon44009.2020.9368301
20. Adee, R., Mouratidis, H.: A dynamic four-step data security model for data in cloud computing based on cryptography and steganography. Sensors **22**(3), 1109 (2022). https://doi.org/10.3390/s22031109
21. Jogar, S., Handral, D.S.: Secure file storage on cloud using hybrid cryptography. Int. J. Adv. Res. Sci. Commun. Technol. 540–551 (2022). https://doi.org/10.48175/ijarsct-5861
22. Dhanani, C., Panchal, K.: HTML steganography using relative links & multi web-page embedment. Int. J. Eng. Dev. Res. (IJEDR) **2**(2), 1960–1965 (2014). ISSN: 2321-9939. http://www.ijedr.org/papers/IJEDR1402108.pdf
23. Kabetta, H., Dwiandiyanta, B., Suyoto, S.: Information hiding in CSS: a secure scheme text-steganography using public key cryptosystem. SSRN Electron. J. (2012). https://doi.org/10.2139/ssrn.3635340
24. Astuti, N.R., Aribowo, E., Saputra, E.: Data security improvements on cloud computing using cryptography and steganography. In: IOP Conference Series: Materials Science and Engineering, vol. 821, no. 1, p. 012041 (2020). https://doi.org/10.1088/1757-899x/821/1/012041

Applications and Associated Challenges in Deployment of Software Defined Networking (SDN)

Pashupati Baniya[1], Atul Agrawal[1], Parma Nand[1], Bharat Bhushan[1], Alaa Ali Hameed[2], and Akhtar Jamil[3](✉)

[1] Department of Computer Science and Engineering, School of Engineering and Technology, Sharda University, Greater Noida, India
parma.nand@sharda.ac.in

[2] Department of Computer Engineering, Istinye University, Istanbul, Turkey
alaa.hameed@istinye.edu.tr

[3] Department of Computer Science, National University of Computer and Emerging Sciences, Islamabad, Pakistan
akhtar.jamil@nu.edu.pk

Abstract. SDN, a rising technology within the realm of Internet of Things (IoT), has been increasingly well-received in recent times. This article presents a summary of SDN along with its different elements, advantages, and difficulties. The paper aims to provide practical solutions for introducing OpenFlow into commercial routers without hardware modifications and extending the integration of OpenFlow with legacy control protocols and control planes. In addition, the paper presents a refactoring process for migrating traditional network applications to OpenFlow-based ones, focusing on the security challenges and techniques of open technologies like SDN, OpenROADM, and SDN-based Mobile Networks (SDMN). The document also examines the advantages and possible uses of SDMN in enhancing network adaptability, streamlining network administration, and bolstering network security. The article also discusses O-RAN network innovations and difficulties, such as AI and ML workflows that are made possible by the architecture and interfaces, security concerns, and, most importantly, standardization issues.

Keywords: Software-Defined Networking (SDN) · Network flexibility · Network security · Hybrid SDN Architecture · Artificial intelligence (AI)

1 Introduction

Software-Defined Networking (SDN) and its various components, advantages, and drawbacks are discussed in this paper. It makes sense how SDN isolates the control plane from the information plane, empowering brought together and programmable command over network framework. The introduction also emphasizes the advantages of SDN, which include opportunities for innovation, simplified network management, increased network flexibility, and enhanced network security.

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 544–557, 2024.
https://doi.org/10.1007/978-3-031-56728-5_45

This paper discusses the challenges of integrating Software-Defined Networking (SDN) and OpenFlow with legacy systems. It proposes a hybrid SDN architecture that combines the distributed functions of legacy routers with the centralized control and management capabilities of OpenFlow. The paper aims to provide practical solutions for introducing OpenFlow into commercial routers without hardware modifications and extending the integration of OpenFlow with legacy control protocols and control planes. The authors also emphasize the need to accelerate the evolution from legacy networks to SDN while leveraging existing investments in infrastructure. Overall, the paper aims to address the challenges and provide solutions for integrating SDN with legacy systems, highlighting the benefits and potential applications of SDN in improving network flexibility, simplifying network management, and enhancing network security.

This paper is to address the challenges involved in integrating Software-Defined Networking (SDN) and OpenFlow with legacy systems. These challenges include the need for a hybrid architecture that combines the distributed functions of legacy routers with the centralized control and management capabilities of OpenFlow, practical solutions for introducing OpenFlow into commercial routers without hardware modifications, and extending OpenFlow to integrate with legacy control protocols and control planes. The paper aims to overcome these challenges and provide a promising approach for accelerating the evolution from legacy networks to SDN while leveraging existing investments in infrastructure. The major contributions of the paper are to addresses challenges in merging SDN and OpenFlow with legacy systems. It suggests a hybrid SDN architecture that blends legacy router functions and OpenFlow's control capabilities to balance compatibility and advantages. Emphasizing firmware updates for seamless integration into commercial routers, it minimizes hardware costs. The hybrid architecture enables a smoother SDN transition without major disruptions. In short, the paper offers integration insights, guiding organizations toward SDN adoption while safeguarding existing investments.

The paper starts with the topic of SDN and its architecture. It then discusses the challenges of integrating SDN and OpenFlow with legacy systems, followed by a proposed hybrid SDN architecture for integrating SDN and legacy systems. The paper then provides practical solutions for introducing OpenFlow into commercial routers and extending OpenFlow to integrate with legacy control protocols and control planes. The advantages and potential applications of SDN are also discussed. Finally, the paper concludes with future work.

2 Software Defined Network

SDN is a creative organization engineering that isolates the control plane from the information plane, empowering brought together and programmable command over network foundation. In SDN, a centralized controller manages the network and instructs the data plane devices on how to handle traffic flows. By decoupling control from individual network devices, SDN offers benefits such as simplified network management, dynamic provisioning of resources, and rapid deployment of new services. It enhances network flexibility, agility, and scalability, allowing administrators to adapt quickly to changing requirements [1]. SDN also promotes automation, improves network security, and fosters innovation by providing open interfaces for developers to create and experiment with network applications and services. Overall, SDN revolutionizes network architecture by enabling centralized control, programmability, and adaptability, leading to more efficient and agile networks.

2.1 SDN Architecture and Component of SDN

SDN represents a structural method for designing and managing networks, with the goal of enhancing the flexibility, scalability, and programmability of computer networks. It enables network administrators to centrally oversee and govern the network using software applications, achieved by separating the network's control plane from the underlying hardware infrastructure [2].

In conventional systems administration, network gadgets like switches and switches pursue sending choices freely founded on their singular designs. In SDN, nonetheless, the control plane is decoupled from the information plane, and a brought together regulator is liable for overseeing and coordinating the traffic stream in the organization. This controller communicates with the network devices using a standardized protocol, such as OpenFlow, to instruct them on how to handle network traffic. The key components of an SDN architecture include [3]:

1. Infrastructure: This alludes to the actual organization gadgets, for example, switches and switches, which are answerable for sending bundles in light of the guidelines, got from the regulator.
2. Controller: The controller serves as the core of the SDN architecture. It offers a centralized perspective of the network, administers network policies, and determines how to manage traffic considering network conditions, application needs, and established policies. The controller communicates with the infrastructure devices to configure their behavior.
3. Southbound Interfaces: These interfaces are employed by the controller to establish communication with the network devices within the data plane. They empower the controller to program and manage the functioning of switches and routers.
4. Northbound Interfaces: These interfaces allow higher-level applications and services to communicate with the SDN controller. They provide APIs and protocols that enable external applications to make requests and receive information about the network state, allowing for programmability and automation.

SDN architecture offers improved network agility, simplified management, scalability, increased security, and easy integration of new services. Centralized control allows quick network adaptation, resource optimization, and simplified troubleshooting. It's used in data centers, WANs, campus networks, providing flexibility and control. Figure 1 depicts SDN's elements: controller, devices, control/data plane separation, interfaces, and benefits like agility, management ease, and security.

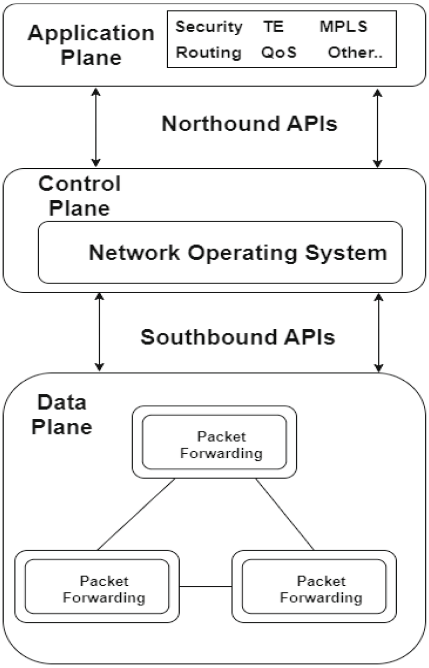


Fig. 1. SDN Architecture and Components

2.2 SDN Controller

An SDN controller is a critical component of an SDN architecture that serves as the central point of control and management for the network. It functions as the network's "intelligence," allowing administrators to programmatically define and manage the actions of network devices.

The SDN controller communicates with network devices using standardized protocols like OpenFlow or vendor-specific ones, instructing switches, routers, and other devices on traffic forwarding and processing [4]. Its main functions include:

- Network Visibility: The controller gathers real-time network data, including topology, device status, and traffic. This provides administrators a centralized view to monitor performance, detect issues, and make informed decisions.
- Policy Management: The controller establishes and enforces network policies, covering QoS, security, and traffic prioritization, maintaining consistent application throughout the network.
- Traffic Engineering: The controller intelligently routes network traffic using defined policies and adapts to optimize performance, reduce congestion, and balance loads.
- Service Orchestration: The controller facilitates network service deployment and management through northbound interfaces, enabling external applications to interact programmatically for resource requests, network status updates, and behavior control.

- Network Automation: The controller automates network tasks, reducing manual configurations and errors. It enables defining and deploying services through software, simplifying network operations.

SDN controllers vary, including open-source options (e.g., OpenDaylight, ONOS) and proprietary ones from networking vendors. They offer flexibility for network management, adapting to changes, enhancing efficiency, and fostering innovation. Table 1 outlines the controller's roles: maintaining network visibility, enforcing policies, optimizing traffic, enabling service orchestration, and facilitating automation.

Table 1. Key Functions of Network Controller.

Function	Description
Network Visibility	Controller maintains real-time network view, aiding monitoring, analysis, and decision-making
Policy Management	Controller enforces policies, including QoS, security, and traffic handling. Consistent network-wide application
Traffic Engineering	Intelligently routes traffic, optimizes paths, minimizes congestion, balances network load
Service Orchestration	Enables service deployment, APIs for external control, resource requests, status updates
Network Automation	Automates provisioning, reduces errors, deploys services through software, simplifying network

3 Benefits of SDN

SDN revolutionizes network management and infrastructure by separating control and data planes, providing flexibility, streamlined management, rapid resource allocation, and new service introduction. It saves costs, enhances scalability, improves security through centralized policy enforcement, and encourages innovation. In short, SDN offers unmatched flexibility, simplified management, cost-effectiveness, scalability, security enhancement, and innovation opportunities, transforming modern networking.

3.1 Enhanced Network Flexibility

The integration of SDN and network slicing in 5G networks enhances network flexibility by enabling dynamic resource allocation. SDN, aided by machine learning algorithms, optimizes network slice allocation for various applications, ensuring efficient resource utilization, high-speed data transfer, and low latency. This synergy empowers network operators to deliver tailored services, optimize user experiences, and adapt to changing network demands in the evolving digital landscape [5].

3.2 Simplified Network Management

SDN decouples control and data planes, centralizing management through software-defined controllers. It facilitates quick app deployment, precise traffic shaping, efficient resource use, and streamlined QoS and security management. Programmability enables dynamic adjustments, exploring new mechanisms, and cost-effective virtualization [6], enhancing network efficiency and agility.

3.3 Improved Network Security

SDN enhances network security through centralized control, improved visibility, and real-time threat detection. Adaptive security measures are possible with dynamic policy enforcement, while segmentation and isolation contain breaches. Rapid response capabilities enable swift action during security incidents, ultimately fortifying defenses, improving threat response, and safeguarding critical assets [7].

3.4 Increased Scalability

SDN offers scalability demonstrated in the study "Cloudy with a Chance of Pain." Using cell phones, Ramantas et al. [8] collected data from 2,658 patients over 15 months to examine pain and weather connections. SDN's centralized control enables gathering and analyzing vast datasets from various sources, enhancing scalability in data collection. In this study, SDN facilitated processing of smartphone data, overcoming challenges in obtaining patient-reported pain and weather data. This scalability enables comprehensive research, addressing health questions, and improving pain management through forecasts. SDN's scalability empowers efficient data handling, driving research, decision-making, and service improvements.

3.5 Cost Savings

SDN offers cost savings and optimized network solutions in the rapidly evolving technology landscape, including the transition to 5G and beyond. It provides a cost-effective approach to network management and security. Bendale et al. [9] address security challenges related to open technologies like SDN and SDMN. SDN centralizes control, enhancing resource allocation and utilization, enabling dynamic provisioning and scaling of services without costly physical upgrades. Intelligent security mechanisms in SDN ensure robust protection against various attacks, benefiting future technologies like 5G and 6G. Leveraging SDN's cost-saving and security features allows organizations to significantly reduce expenses while maintaining network performance and security.

Table 2 outlines SDN benefits for network management: 5G integration for flexibility, centralized control for streamlined management, enhanced security through dynamic policies, improved scalability, and cost savings via optimized infrastructure and performance.

Table 2. Benefits of Software-Defined Networking (SDN) in Network Management.

Benefit of SDN	Description
Enhanced Network Flexibility	5G SDN integration enhances flexibility, optimizing resources for diverse services
Simplified Network Management	SDN streamlines management, centralizing control, enabling rapid deployment, and enhancing agility
Improved Network Security	SDN strengthens security with centralized view, threat detection, dynamic policies
Increased Scalability	SDN enhances research, pain management with scalable data collection and analysis
Cost Savings	SDN cuts costs, optimizes infrastructure, deploys security, maintaining performance effectively

4 Application of SDN

Software-Defined Networking (SDN) revolutionizes network operations and architectures across domains. In data centers and cloud computing, it enhances scalability, agility, and resource allocation through dynamic provisioning of virtual networks and centralized management. It's interconnected with Network Function Virtualization (NFV), providing a flexible infrastructure for virtual network function deployment. For IoT networks, SDN centralizes control, optimizing traffic for expansive deployments. Moreover, SDN fosters innovation in protocols, algorithms, and architectures as a research platform. Its diverse applications transform network management, security, scalability, and innovation in varied environments.

4.1 Software-Defined Data Centers (SDDC)

SDN finds application in software-defined data centers, exemplified by NetFlash research. NetFlash co-designs SDN and storage stack to enhance end-to-end performance. Traditional data centers manage SDN and software-defined flash (SDF) separately, leading to suboptimal results. Integrating SDN into software-defined data centers improves resource management for compute, memory, and storage. NetFlash showcases network/storage co-design benefits, enabling cross-stack coordination and state sharing. Storage management functions merge with SDN infrastructure for global resource management, while key functions stay on storage servers. NetFlash optimizes I/O requests using network packet formats and tracking mechanisms, reducing delays. The synergy of SDN and SDF in software-defined data centers boosts performance, flexibility, and scalability [10], paving the way for co-design opportunities in rack-scale hardware resource management.

4.2 Internet of Things (IoT)

The integration of SDN in the IoT offers numerous benefits, including improved network management, customization, flexibility, and reduced maintenance costs. IoT networks are more effective and responsive thanks to SDN's centralised control and separation of the control and data planes. In terms of security, SDN can address vulnerabilities in IoT networks by offering a more secure and scalable architecture [11]. Ferrão et al. [11] explored machine learning algorithms to enhance intrusion detection systems and prevent false alarms. As IoT adoption increases, SDN is expected to play a crucial role in developing efficient and secure IoT networks.

4.3 Cloud Computing

SDN innovation streamlines network control in cloud computing. Cloud resources are shared online, necessitating effective management and data security. SDN enhances resource management and data security in cloud computing. It optimizes network traffic and bandwidth allocation, aiding efficient resource utilization. Moreover, SDN strengthens data security by enabling precise control and stricter security policies, mitigating unauthorized access and breaches [12].

4.4 Network Function Virtualization (NFV)

SDN in NFV, combined with blockchain, benefits IoT. SDN centrally controls network resources for IoT deployment, while NFV virtualizes functions for scalability. Blockchain adds trust, security, and scalability [13]. Integration ensures IoT data security, reliable transfer, interoperability, and smart contracts, addressing cloud-driven IoT challenges. This convergence enables secure, scalable, and efficient IoT deployments.

4.5 Traffic Engineering and Load Balancing

SDN offers valuable applications in traffic engineering and load balancing through its centralized control plane and programmability. By isolating the control plane from the information plane, SDN enhances adaptable traffic management. The central controller dynamically optimizes network performance, allocates resources, and balances traffic loads based on real-time data, preventing congestion and improving overall efficiency [14]. This flexibility leads to effective resource utilization, enhanced user experience, and the fulfillment of Service Level Agreements (SLAs). SDN empowers administrators to dynamically manage traffic, allocate resources wisely, and improve network performance while avoiding congestion.

Table 3 illustrates diverse SDN applications: enhancing data centers, optimizing networks (campus and WAN), IoT management, cloud resource control, integrating with NFV for IoT security, and centralized traffic engineering and load balancing.

Table 3. Applications and Benefits of Software-Defined Networking (SDN).

Application	Summary
Software-Defined Data Centers (SDDC)	SDN enhances software-defined data centers through network/storage co-design, improving compute, memory, and storage resource management
Internet of Things (IOT)	SDN benefits IoT by enhancing network management, customization, flexibility, and security through centralized control
Cloud Computing	SDN aids in efficient resource management, data security, and traffic control in cloud computing, improving user experience
Network Function Virtualization (NFV)	SDN-NFV integration with blockchain enhances IoT security, scalability, interoperability, and trust in cloud-driven ecosystems
Traffic Engineering and Load Balancing	SDN's centralized control optimizes traffic engineering and load balancing, dynamically managing network resources and flows

5 Challenges of SDN

SDN implementation faces challenges including legacy system integration, security concerns, scalability, performance issues, vendor lock-in, and standardization problems. Integrating with legacy systems requires a hybrid architecture, combining central control with distributed legacy functions. Security challenges arise from centralized control, encompassing controller security, communication security, API openness, virtualization security, and more. Scalability and performance are crucial for managing growing networks efficiently. Vendor lock-in is a concern, particularly in disaggregated Optical Transport Networks; addressing this involves standardized data models. Standardization problems also emerge, especially in SDN adoption within Open Radio Access Networks, highlighting the need for industry-wide standards. Overcoming these challenges can lead to improved network management and flexibility.

5.1 Integration with Legacy Systems

Integrating SDN and OpenFlow with legacy systems presents challenges. A hybrid SDN architecture is proposed to address these, combining distributed legacy functions with centralized control. This preserves existing infrastructure while enhancing network management. Feng et al. [15] suggest a refactoring process for migrating traditional network apps to OpenFlow-based ones, introducing OpenFlow to routers via firmware updates. The architecture includes an OpenRouter module for dynamic adjustments through an external OpenFlow controller, facilitating the transition to SDN while reducing overhead. This hybrid approach leverages existing investments and enhances network management without discarding legacy systems.

5.2 Security Concerns

Integrating SDN with legacy systems brings security challenges that demand a thorough security approach. These challenges involve centralized control, controller and communication security, API openness, virtualization security, flow-based security, insider threats, and complex security policies. Securing the SDN controller, communication channels, and access controls is crucial. Addressing vulnerabilities from open APIs and securing virtualized components is essential. Strengthening flow-based security is necessary against spoofing and manipulation. Insider threats need monitoring and access controls. Handling complex security policies and standardizing security mechanisms are also important. Implementing these measures ensures strong security for SDN deployments, safeguarding network integrity and mitigating integration risks with legacy systems [16].

5.3 Scalability and Performance

SDN adoption revolutionizes networking with flexibility and control. Challenges arise with network growth, requiring scalable solutions. Issues include managing wireless networks, optimizing controller placement, ensuring resilience, enhancing Wi-Fi support, and monitoring metrics. Bandwidth demands can overwhelm networks, mitigated by standardized hardware. Controller placement impacts performance and resilience; distributed architectures reduce risks. Limited Wi-Fi support hampers SDN; integration with techniques like Mininet-Wi-Fi and Ryu Controller improves scalability [17]. Monitoring metrics like bandwidth is crucial. Addressing challenges optimizes SDN in Software-Defined Wireless Networking (SDWN), ensuring better scalability and performance.

5.4 Vendor Lock-in

In SDN, addressing vendor lock-in is crucial, especially for disaggregated Optical Transport Networks (OTN). Standardized data models like OpenConfig or OpenROADM are used to represent configuration and operational parameters of optical network elements (NEs) in disaggregated OTN. A unified data model approach reduces vendor lock-in, costs, and boosts flexibility in OTN deployments. However, it adds complexities in supporting end-to-end network services. A solution is partial disaggregation, separating the network into Optical Terminal (OT) and Optical Line System (OLS), including multiple autonomous OLS domains. Karunakaran et al. [18] evaluated OpenROADM's challenges from data and control plane perspectives, comparing it with OpenConfig and vendor models. They explored open-source SDN controllers supporting OpenROADM to tackle vendor lock-in, aiming to enhance SDN-based OTN's interoperability and flexibility.

5.5 Standardization Issues

The adoption of SDN in Open Radio Access Network (O-RAN) poses challenges in standardization. O-RAN, guided by O-RAN Alliance specs, transforms telecom with virtualized RANs, interconnecting disaggregated parts via open connection points and intelligent controllers. This approach envisions multi-vendor, upgradable O-RAN networks managed through a centralized layer. Standardization hurdles hinder O-RAN's potential. Polese et al. [19] offer a comprehensive O-RAN tutorial covering concepts, architecture, and interfaces. They stress O-RAN RAN Intelligent Controllers' role and highlight AI/ML workflows, security, and standardization. The authors review experimental phases for O-RAN network design, concluding with directions for tackling standardization challenges for widescale O-RAN implementation.

Table 4 outlines challenges in SDN implementation: integration with legacy systems, security, scalability, performance, vendor lock-in, and standardization.

Table 4. Challenges and Considerations in SDN Implementation.

Challenge	Summary
Integration with Legacy Systems	Hybrid SDN architecture integrates SDN with legacy systems, improving end-to-end performance and enabling gradual migration
Security Concerns	Legacy SDN integration demands robust security: control, communication, access
Scalability and Performance	SDN tackles growth issues through optimized placement, standardized hardware, metrics
Vendor Lock-in	Disaggregated OTN uses OpenConfig for flexibility, cost reduction, vendor lock-in
Cloud Computing	SDN aids in efficient resource management, data security, and traffic control in cloud computing, improving user experience

6 Future Trends

SDN's applications span diverse domains. Network slicing, foremost, forms virtual networks in a physical one, serving varied needs. Merging SDN with slicing demands resource distribution and end-to-end coordination study. Edge computing melded with SDN offers promise in healthcare and IoT, needing power and security remedy. Intent-Based Networking (IBN) automates virtual network control through abstract goals, urging intent lifecycle management and optimization growth. Lastly, fusing AI with SDN for load balancing boosts network efficiency, mandating AI-based balancing enhancement. These trends amplify SDN, revealing novel management chances.

6.1 Network Slicing

Network slicing creates virtual networks within a physical network for tailored services and efficient resource use. Combined with SDN, it enables effective network control.

SDN's centralized control suits slicing, allowing dynamic resource allocation and rapid service setup. Components like network slice controllers, P4 switches, and INT data are key in implementing SDN-based slicing. Hwang et al. [20] find SDN with P4 switches improves performance for various slice types. Integrating slicing, SDN, and related tech optimizes 5G resource use and network performance.

6.2 Edge Computing and SDN

The integration of SDN and edge computing has significant potential for transforming IoT-based healthcare. Edge processing reduces latency and enables real-time monitoring by bringing cloud capabilities closer to the data source. SDN provides centralized control and management of IoT devices, improving device management, interoperability, and resource allocation. Challenges such as power limitations, security, and interoperability require attention through frameworks, architectures, and solutions [21]. Technologies like MEC and machine learning enhance resource allocation and data processing, while security frameworks, AI algorithms, and blockchain technology bolster security measures and protect healthcare data.

6.3 Intent-Based Networking (IBN)

Intent-Based Networking (IBN) automates SDN-based virtual network (VN) configuration and management using tenant intents. It allows multiple VNs on one infrastructure for separate tenant control, with layers for protocol adaptation, abstraction, virtualization, and intent management. IBN employs open-source tools like ONOS SDN controller and OpenVirteX hypervisor, enhancing multi-VN support. Its goals include simplified network management, secure VN isolation, but it needs further work for native virtualization and optimal VN embedding [22]. IBN offers promising network management but requires ongoing research and development for real-world assessment.

Table 5 presents various innovative applications and considerations in Software-Defined Networking (SDN), including network slicing, edge computing transformation, intent-based networking automation, and the integration of artificial intelligence for load balancing and enhanced performance.

Table 5. Innovation Application and Consideration in SDN.

Application	Summary
Networking Slicing	SDN-integrated network slicing optimizes resource usage, enables rapid service deployment. Challenges include security, privacy, and coordination
Edge Computing and SDN	SDN in edge computing transforms IoT-based healthcare, addressing power, security, and interoperability challenges
Intent-Based Networking (IBN)	IBN automates virtual network provisioning, enhancing management, isolation, and scalability

7 Conclusion and Future Research Directions

SDN is a transformative technology that offers centralized control, programmability, improved network performance, and efficient resource utilization. Its flexible architecture addresses the limitations of traditional approaches, enabling effective management and optimization of network resources. SDN finds applications in traffic engineering, load balancing, IoT, and various domains, showcasing its versatility and potential for enhancing network performance. SDN's dynamic resource allocation, traffic rerouting, and adaptability ensure efficient bandwidth utilization and improved user experiences. The future prospects of SDN are promising, with ongoing research exploring its integration with emerging technologies like AI, ML, and 5G networks. These collaborations open avenues for intelligent network automation, real-time analytics, and predictive network management. As more organizations recognize the benefits of SDN, its adoption is expected to expand, driving standardization, interoperability enhancements, and advanced SDN-based applications.

References

1. Hussain, M., Shah, N., Amin, R., Alshamrani, S.S., Alotaibi, A., Raza, S.M.: Software-defined networking: categories, analysis, and future directions. Sensors **22**(15), 5551 (2022). https://doi.org/10.3390/s22155551
2. Benzekki, K., El Fergougui, A., Elbelrhiti Elalaoui, A.: Software-defined networking (SDN): a survey. Secur. Commun. Netw. **9**(18), 5803–5833 (2016). https://doi.org/10.1002/sec.1737
3. Montazerolghaem, A.: Software-defined load-balanced data center: design, implementation and performance analysis. Cluster Comput. **24**(2), 591–610 (2021). https://doi.org/10.1007/s10586-020-03134-x
4. Paliwal, M., Shrimankar, D., Tembhurne, O.: Controllers in SDN: a review report. IEEE Access **6**, 36256–36270 (2018). https://doi.org/10.1109/ACCESS.2018.2846236
5. Arif, M.A.I., Kabir, S., Khan, M.F.H., Dey, S.K., Rahman, M.M.: Machine learning and deep learning based network slicing models for 5G network. In: 2022 25th International Conference on Computer and Information Technology (ICCIT), Cox's Bazar, Bangladesh, pp. 96–101 (2022). https://doi.org/10.1109/ICCIT57492.2022.10054696
6. Ergen et al.: Implementation of a SDN (software defined network). Int. Inst. Inform. Syst. **IIIS** (2014). ISDN 9781941763049, https://hdl.handle.net/20.500.12469/1368
7. Aditya, T., Donald, A.D., Thippanna, G., Kousar, M.M., Murali, T.: NFV and SDN: a new era of network agility and flexibility. Int. J. Adv. Res. Sci. Commun. Technol. 482–493 (2023). https://doi.org/10.48175/IJARSCT-8526
8. Ramantas, K., Antonopoulos, A., Kartsakli, E., Mekikis, P.V., Vardakas, J., Verikoukis, C.: A C-RAN based 5G platform with a fully virtualized, SDN controlled optical/wireless fronthaul. In: 2018 20th International Conference on Transparent Optical Networks (ICTON), pp. 1–4. IEEE (2018). https://doi.org/10.1109/ICTON.2018.8473489
9. Bendale et al.: State of the art challenges and technique for 5G and 6G using software defined network. In: 2023 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, pp. 1–6 (2023). https://doi.org/10.1109/ESCI56872.2023.10099588
10. Reidys, B., Huang, J.: Building next-generation software-defined data centers with network-storage co-design
11. Ferrão, T., Manene, F., Ajibesin, A.A.: Multi-attack intrusion detection system for software-defined internet of things network. Comput. Mater. Contin. **75**, 4987–5007 (2023). https://doi.org/10.32604/cmc.2023.038276

12. Rahman, A., et al.: Towards a blockchain-SDN-based secure architecture for cloud computing in smart industrial IoT. Digit. Commun. Netw. **9**(2), 411–421 (2023). https://doi.org/10.1016/j.dcan.2022.11.003
13. Rahman, A., et al.: Impacts of blockchain in software-defined Internet of Things ecosystem with network function virtualization for smart applications: present perspectives and future directions. Int. J. Commun. Syst. e5429 (2023). https://doi.org/10.1002/dac.5429
14. Fawaz, H., et al.: Graph convolutional reinforcement learning for load balancing and smart queuing. In: 2023 IFIP Networking Conference (IFIP Networking), pp. 1–9. IEEE (2023). https://doi.org/10.23919/IFIPNetworking57963.2023.10186430
15. Feng, T., et al.: Hybrid SDN architecture to integrate with legacy control and management plane: an experiences-based study. In: 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM), pp. 754–757. IEEE (2015). https://doi.org/10.1109/INM.2015.7140368
16. Kazmi, S.H.A., et al.: Survey on joint paradigm of 5G and SDN emerging mobile technologies: architecture, security, challenges and research directions. Wirel. Pers. Commun. **130**, 2753–2800 (2023). https://doi.org/10.1007/s11277-023-10402-7
17. Ali, M., et al.: Performance and scalability analysis of SDN-based large-scale Wi-Fi networks. Appl. Sci. **13**(7), 4170 (2023). https://doi.org/10.3390/app13074170
18. Karunakaran, V., Patri, S.K., Zimmermann, S., Autenrieth, A., Bauschert, T.: OpenROADM for disaggregated optical networks: challenges, requirements and evaluation. In: Photonic Networks; 24th ITG-Symposium, pp. 1–5. VDE (2023)
19. Polese, M., et al.: Understanding O-RAN: architecture, interfaces, algorithms, security, and research challenges. IEEE Commun. Surv. Tutor. **25**(2), 1376–1411 (2023). https://doi.org/10.1109/COMST.2023.3239220
20. Wu, Y.-J., Hwang, W.-S., Shen, C.-Y., Chen, Y.-Y.: Network slicing for mMTC and URLLC using software-defined networking with P4 switches. Electronics **11**(14), 2111 (2022). https://doi.org/10.3390/electronics11142111
21. Kumhar, M., Bhatia, J.B.: Edge computing in SDN-enabled IoT-based healthcare frameworks: challenges and future research directions. Int. J. Reliab. Qual. E-Healthc. **11**(4), 1–15 (2022). https://doi.org/10.4018/IJRQEH.308804
22. Han, Y., et al.: An intent-based network virtualization platform for SDN. In: 2016 12th International Conference on Network and Service Management (CNSM), pp. 353–358. IEEE (2016). https://doi.org/10.1109/CNSM.2016.7818446

Combining Text Information and Sentiment Dictionary for Sentiment Analysis on Twitter During COVID

Vidushi[1], Anshika Jain[1], Ajay Kumar Shrivastava[1], Bharat Bhushan[2], Alaa Ali Hameed[3], and Akhtar Jamil[4](✉)

[1] KIET Group of Institutions, Delhi-NCR, Ghaziabad, India
{anshika.1821mca1080,ajay}@kiet.edu

[2] Department of Computer Science and Engineering, School of Engineering and Technology, Sharda University, Greater Noida, India

[3] Department of Computer Engineering, Istinye University, Istanbul, Turkey
alaa.hameed@istinye.edu.tr

[4] Department of Computer Science, National University of Computer and Emerging Sciences, Islamabad, Pakistan
akhtar.jamil@nu.edu.pk

Abstract. Presence of heterogenous huge data leads towards the 'big data' era. Recently, tweeter usage increased with unprecedented rate. Presence of social media like tweeter has broken the boundaries and touches the mountain in generating the unstructured data. It opened research gate with great opportunities for analyzing data and mining 'valuable information'. Sentiment analysis is the most demanding, versatile research to know user viewpoint. Society current trend can be easily observed through social network websites. These opportunities bring challenges that leads to proliferation of tools. This research works to analyze sentiments using tweeter data using Hadoop technology. It explores the big data arduous tool called Hadoop. Further, it explains the need of Hadoop in present scenario and role of Hadoop in storing ample of data and analyzing it. Hadoop cluster, Hadoop Distributed File System (HDFS), and HIVE are also discussed in detail. The Dataset used in performing the experiment is presented. Moreover, this research explains thoroughly the implementation work and provide workflow. Next session provides the experimental results and analyzes of result. Finally, last session concludes the paper, its purpose, and how it can be used in upcoming research.

Keywords: Hadoop · Analysis · HIVE · Covid · Twitter · Sentiment

1 Introduction

In present era, twitter is a social media platform that offers a microblogging service. It is a web-based entertainment stage for PC intervened web-based correspondence, which influences the social construction that emerges. This correspondence stage has

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 558–569, 2024.
https://doi.org/10.1007/978-3-031-56728-5_46

1.3 billion records and 336 million dynamic clients posting 500 million tweets each day [1]. Aladwani et al. [2] presented massive amounts of rich Twitter data. Worldwide people or various organizations are gaining and transforming information using this media. It provides a public medium through which users interaction enhanced by exchanging emotion, thought, pushing status. Since from 2006, twitter gains the success among all present platforms of social networking [3]. Analysis of twitter data can be valuable in number of uses such like political elections [4]. People's emotions can be monitored effectively and regularly. In this competitive market, cooperate world is getting benefit by analyzing user information [5]. Person sentiment analysis could be a useful tool in getting hotel reviews [6]. Currently, the research on the analysis of sentiments becomes a hot topic for various applications like judgement, sentiment analysis [7].

Since it was initially used as a task roughly 20 years ago, sentiment analysis as a field has advanced significantly. It has several commercial uses in a variety of industries, including marketing, risk management, market analysis, and politics, to mention a few. There is a general view that this subject has matured because to its saturation in some subtasks, such as emotion polarity classification, and datasets [8].

With the development of web 2.0, a huge amount of unstructured data, including comments, opinions, and other types of data, is produced in real-time. Building an accurate predictive model for sentiment analysis is difficult due to the data's unstructured nature. Additionally, modern techniques do not effectively use semantic and sentiment knowledge to extract significant important contextual sentiment characteristics [9]. More than five million people have already died as a result of the COVID-19 epidemic, prompting a rush of research from a range of fields to offer remedies [10]. Since the COVID-19 epidemic, social media has been essential for keeping in touch with friends, family, and coworkers as well as for staying updated and discussing new policy updates and regulations [10]. The COVID-19 is a hot topic on Twitter. During the pandemic, people all over the world have used Twitter to communicate their thoughts and experiences [10].

In this research, Hadoop framework is used for storing and analyzing covid tweets dataset. The objective is to categorize users' tweets into one of three categories by analyzing the sentiment of their tweets – positive, negative, neutral. The tweets are classified into these categories using a dictionary text file which contains English words with its score. At last, a graph is plotted showing total number of tweets that are positive (good), negative and neutral. Visualization of the sentiment counts will give better understanding.

Presently, twitter analysis becomes the good choice to get insight analysis of person emotions. Analysis through twitter messages removes man-machine interaction barrier. Recently, world is facing a significant problem called COVID. In this time, twitter becomes a platform through which people shared their emotions. So, it becomes important to analyze these tweets. Therefore, this research focuses on covid tweets dataset analysis using dictionary text file. To accomplish this task, various literature study has been done. After that tweet dataset storage and analysis work has been done using Hadoop Distributed File System (HDFS) and HIVE. Furthermore, visualization is done using BI power tool.

In this paper, sentiment analysis is done using covid tweets dataset. Along with the covid tweets dataset, dictionary text file is also used. The rest of this article comprises of five sections. Firstly, the related literature work is expounded. Secondly, the dataset with dictionary file is described. Thirdly, the methodology is discussed. Fourthly, experiments are completed, and the exploratory outcomes are examined and talked about. Finally, the work is concluded, and future work is also discussed here.

2 Literature Review

Sentimental examination is a course of making a decision about feeling articulation through language. It incorporates getting emotional data handling, examining information, and characterizing the feeling of chosen text [6]. The fundamental undertakings of feeling investigation are feeling acknowledgment, extreme recognition and emotional registration [11–13]. Analysis of text sentiments is an important technique in NLP field. It is used for mining sentiments. It is broadly utilized in popular assessment observing, AI and business insight [14].

On the basis of the various literature studied in this research, it can be inferred that presently various machine learning algorithms are drastically used. In AI based approaches, a classifier based on sentiment is prepared utilizing a pre-marked corpus. Ye et al. [15] analyzed the opinion grouping impacts of NB, SVM and N-gram model on text remarks and figured out that the exactness of SVM and NB calculation is essentially higher than that of N-gram model. Considering the semantic connections between words, Zhang et al. [16] proposed a technique in light of Word2vec and SVMperf to arrange Chinese remark texts. The exploratory outcomes exhibit the hybrid algorithm adequacy, which can be more than 90% exactness in classification of sentiments. Yang et al. [17] proposed a STSM is intended to catch subject opinion relationship and gauge fine-grained feelings. To achieve the best results in sentiment classification, different arrangement models have been recently developed. These include a sentiment model [18, 19], a hybrid ensemble pruning model [20] a bag of meta-words model [21], deep learning methods based on neural networks [19–21], and an LSTM model [21].

This article employs Twitter text information and a sentiment lexicon to analyze the sentiments during COVID based on previously explored related work. In contrast with the previous work, this study uses the Hadoop technology to store and analyze the twitter data. Motive of this study is to analyze sentiments during covid. To accomplish this task, Hadoop components, HDFS, and HIVE are used. This technology is used due to its capability to handle voluminous data.

3 Dataset Description

In this research, Covid tweets dataset, and dictionary text file is used. The objective is to categorize users' tweets into one of three categories by analyzing the sentiment of their tweets – positive, negative, neutral. The tweets are classified into these categories using a dictionary text file which contains English words with its score. We found twitter dataset on Kaggle Website. This data set contains 13 columns. All the entries and Tweet's text in data set are related to Covid19 situation. Table 1 displays dataset details.

Second file used in the research is dictionary.txt. The file has scores associated with each word. The scores are given in such way, the more the word is positively expressed the higher will be its value and for negative words, more the word is negative more

Table 1. Covtweets Dataset

No. of features	No. of Records	Dataset taken from	Dataset Link	Data Format
13	179108	Kaggle	https://www.kaggle.com/gpreda/covid19-tweets	Textual

negative value is specified in the score. If the score of any word is 0 then it means that word is Neutral in nature.

4 Methodology

The Apache HIVE technology has been implemented over the dataset to query different types of data in an organized manner to run the project. The process of HIVE management and file management along with the methodology that incorporates an algorithm used in the proposed work are presented in the subsections below.

4.1 HIVE Working over Dataset

Large dataset management and querying are made easier by the Apache HIVE bigdata software. HIVE is a data warehouse and SQL-like query language system based on the HDFS. Hive is intended to make querying and handling massive datasets in a distributed setting easier, making it appropriate for analytical and corporate intelligence activities. HIVE offers a way to query the data using HiveQL and project structure onto the data. Figure 1 shows all the steps required to set hive. HiveQL, a high-level query language similar to SQL (Structured Query Language) used in relational database systems, is provided by HIVE. Individuals are likely to be comfortable with HIVE, who are accustomed with SQL. Hive connects with the system via command-line and web-based interfaces, as well as connectivity with major data visualization and BI tools.

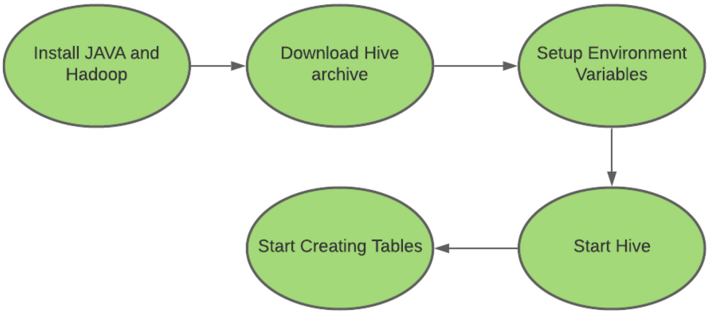


Fig. 1. Steps to setup HIVE

Figure 2 depicts twitter sentiment analysis system. This system uses tweets dataset file and dictionary file. The tweets dataset was downloaded from the previous mentioned source, and the dictionary file includes all of the data necessary to input in the twitter sentiment analysis system. Both of the files, tweets dataset, and dictionary file are sent as inputs to the twitter sentiment analysis system from which the analysis result is generated.

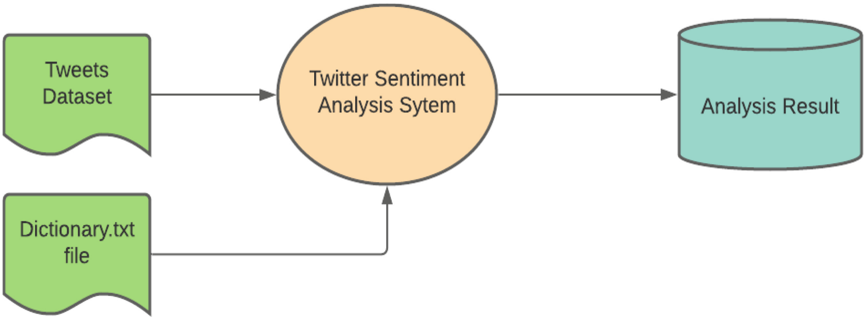


Fig. 2. Twitter sentiment analysis stream

The collecting and storing of Twitter data in the HDFS is the first step in the process. The HDFS architecture stores raw tweets sourced from the Twitter site in predefined folders. Concurrently, the development and maintenance of a dictionary file is an essential part of the process. This dictionary, which contains a vocabulary of preset sentiment-bearing terms and their associated sentiment scores, is critical in the sentiment analysis. The method then moves on to integrate the saved Twitter data with the HIVE data warehouse system. HIVE tables, which are organized repositories that allow for both the storing and accessing of data in a tabular format, are used for this. The previously created vocabulary file is also ingested into a HIVE table, providing a centralized home for sentiment analysis-related resources. The computation of sentiment scores for each tweet is at the heart of the study. This complex phase makes use of both the dictionary and the tweets table in HIVE. A systematic and complete process of how the data is managed is shown in Fig. 3.

4.2 Proposed Algorithm

Within a Hadoop ecosystem, the provided technique defines a systematic approach for sentiment analysis of Twitter data. The first step in the procedure is collecting and setting up the Twitter dataset for further processing and analysis. When you start Hadoop and its components, a directory is created in the Hadoop HDFS. The data from Twitter is then fed into this directory, allowing for dispersed storage. Simultaneously, a dictionary file containing sentiments is saved on the local drive. With Hadoop's services operating, HIVE is enabled, allowing the generation of tables inside HIVE to handle the Twitter data set and the dictionary. The HDFS dataset is copied to the HIVE database, and the dictionary data is copied to a second HIVE table. Tweets are converted into individual words using HQL, laying the groundwork for sentiment assessment. The sentiment-scoring algorithm uses a lexicon to give scores to words, which is then used to calculate

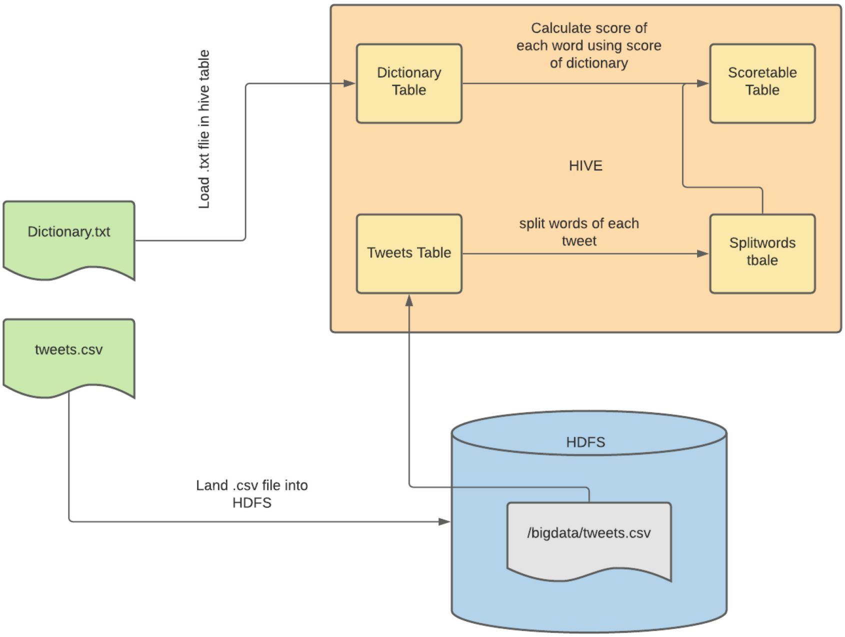


Fig. 3. Systematic work to analyze twitter data

a total rating for each tweet. This score-based technique categorizes feelings as positive (score $>$ 0), negative (score 0), or neutral (otherwise), with sentiment counts shown as a result. The technique provides a thorough approach for sentiment analysis. The algorithm consists of a total of 12 steps that need to be executed in a sequential manner are presented below.

Algorithm 1 Proposed Algorithm

```
Find and prepare the tweeter dataset to load and analyze.
Start Hadoop and all its daemons.
Create a directory in HDFS.
Feed the tweeter data into HDFS inside directory.
Store the dictionary file in local disk.
Start HIVE
Create table in HIVE and store dataset from HDFS to HIVE table.
Now, create a table to load dictionary data.
Using HQL, split tweets into individual words.
Calculate the score using dictionary.
{Score > 0:  positive
    Score < 0: negative
              otherwise, neutral}
Display sentiment counts
```

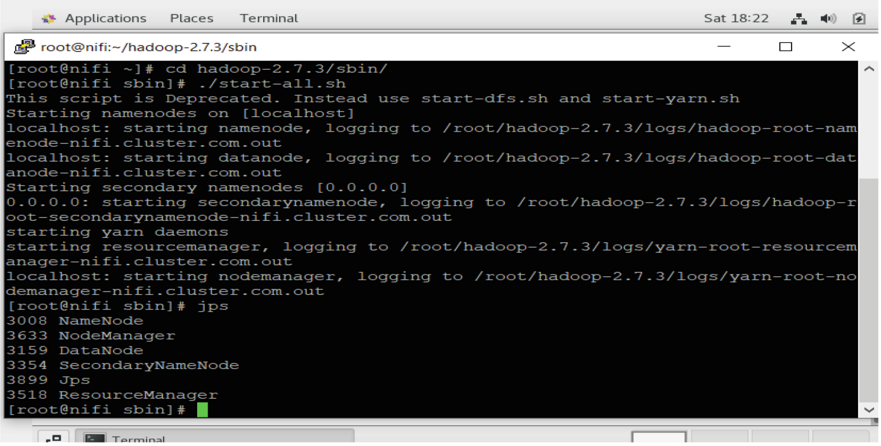
4.3 Project Execution Steps

The project execution refers to making sure the setup is ready to be executed, and that there are no errors will performing the needed steps. The project execution steps are presented below. Pre-requisite to execute the steps is an up-to-date installed Java on the system.

1. Check whether java is properly installed or not.
2. Check java version.
3. Set up Apache Hadoop and HIVE.
4. Check Hadoop version.
5. Check whether all the Hadoop daemons are running or not.
6. If yes, then go ahead
7. Else, set Hadoop properties again.
8. Run HIVE.
9. Using HQL, analyze the result.

5 Results and Discussions

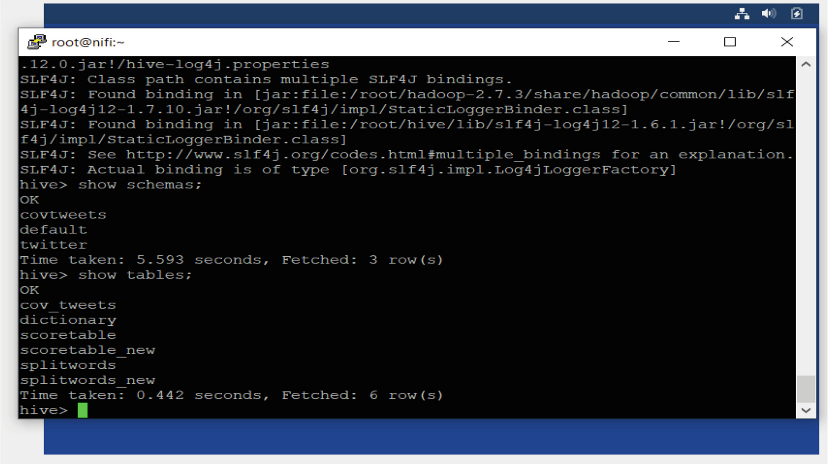
The acquired results include a variety of features such as Hadoop daemon orchestration, HIVE table configuration, tweet classification, sentiment frequency enumeration, and subsequent analytical insights. Initially, an inspection is performed to ensure that all Hadoop daemons are running well. This critical phase ensures the distributed processing framework's core operation. Figure 4 is shows of how these daemons work, with each piece executing its assigned duty in the data processing and analysis.



```
[root@nifi ~]# cd hadoop-2.7.3/sbin/
[root@nifi sbin]# ./start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
Starting namenodes on [localhost]
localhost: starting namenode, logging to /root/hadoop-2.7.3/logs/hadoop-root-nam
enode-nifi.cluster.com.out
localhost: starting datanode, logging to /root/hadoop-2.7.3/logs/hadoop-root-dat
anode-nifi.cluster.com.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /root/hadoop-2.7.3/logs/hadoop-r
oot-secondarynamenode-nifi.cluster.com.out
starting yarn daemons
starting resourcemanager, logging to /root/hadoop-2.7.3/logs/yarn-root-resourcem
anager-nifi.cluster.com.out
localhost: starting nodemanager, logging to /root/hadoop-2.7.3/logs/yarn-root-no
demanager-nifi.cluster.com.out
[root@nifi sbin]# jps
3008 NameNode
3633 NodeManager
3159 DataNode
3354 SecondaryNameNode
3899 Jps
3518 ResourceManager
[root@nifi sbin]#
```

Fig. 4. Running Hadoop Daemons

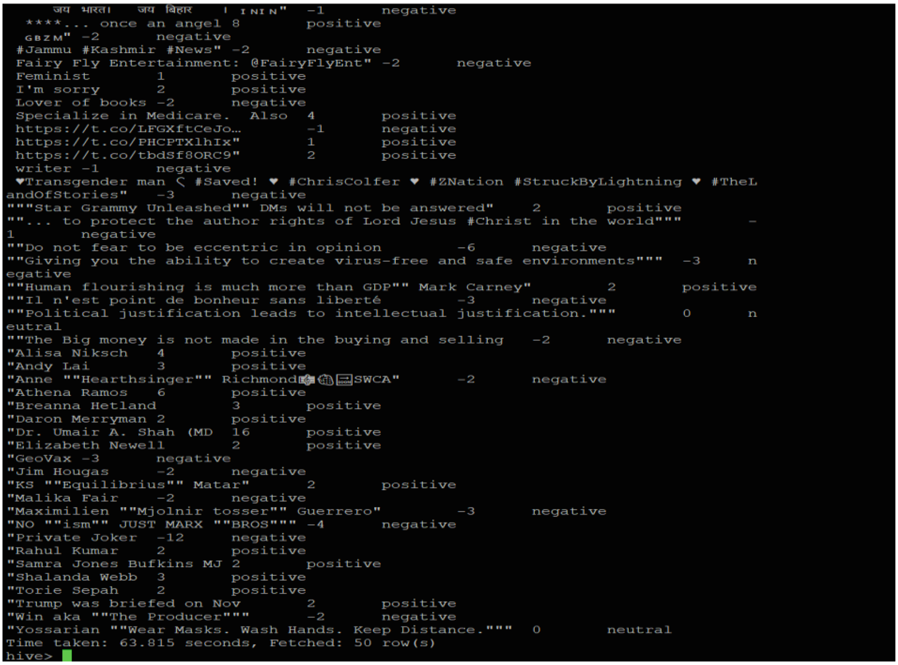
Currently, the software begins HIVE execution, methodically confirming the correct setup of all tables and schemas. It is critical to ensure that these components are created seamlessly. The next stage is to migrate data from the HDFS into HIVE while keeping its structure and organization. The planned execution of this move ensures that the data, now in HIVE, preserves its intrinsic integrity and is ready for organized analysis. Figure 5 illustrates that all the schemas present in HIVE.



```
.12.0.jar!/hive-log4j.properties
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/root/hadoop-2.7.3/share/hadoop/common/lib/slf
4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/root/hive/lib/slf4j-log4j12-1.6.1.jar!/org/sl
f4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
hive> show schemas;
OK
covtweets
default
twitter
Time taken: 5.593 seconds, Fetched: 3 row(s)
hive> show tables;
OK
cov_tweets
dictionary
scoretable
scoretable_new
splitwords
splitwords_new
Time taken: 0.442 seconds, Fetched: 6 row(s)
hive>
```

Fig. 5. HIVE Tables

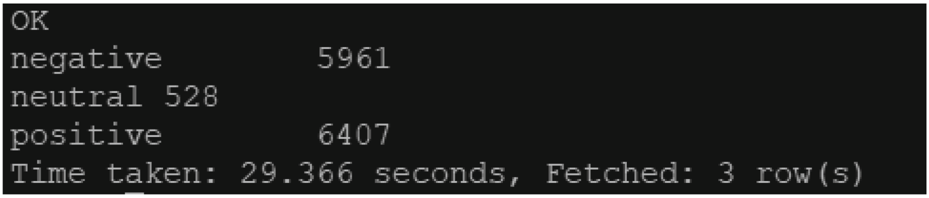
Using dictionary and Covtweets tables, analyze the tweets and sort them into tweets that are positive (excellent), negative (bad), and neutral. During the covid time, use of social media has increased tremendously. Twitter becomes the platform to share their views, especially when people are unable to talk with each other. Figure 6 shows the result after analyzing each tweet. Split tweets into words and then calculate score of each word using score of dictionaries.



```
     जय भारत।   जय बिहार   । I NI N"   -1      negative
  ****... once an angel 8       positive
  GBZM" -2      negative
 #Jammu #Kashmir #News" -2      negative
 Fairy Fly Entertainment: @FairyFlyEnt" -2      negative
 Feminist       1       positive
 I'm sorry      2       positive
 Lover of books -2      negative
 Specialize in Medicare.  Also  4       positive
 https://t.co/LFGXftCeJo…       -1      negative
 https://t.co/PHCPTX1hIx"       1       positive
 https://t.co/tbdSf8ORC9"       2       positive
 writer -1      negative
 ♥Transgender man ☇ #Saved! ♥ #ChrisColfer ♥ #ZNation #StruckByLightning ♥ #TheL
andOfStories"   -3      negative
"""Star Grammy Unleashed"" DMs will not be answered"     2       positive
""... to protect the author rights of Lord Jesus #Christ in the world"""       -
1       negative
""Do not fear to be eccentric in opinion        -6      negative
""Giving you the ability to create virus-free and safe environments"""  -3     n
egative
""Human flourishing is much more than GDP"" Mark Carney"        2       positive
""Il n'est point de bonheur sans liberté        -3      negative
""Political justification leads to intellectual justification."""      0      n
eutral
""The Big money is not made in the buying and selling   -2      negative
"Alisa Niksch   4       positive
"Andy Lai       3       positive
"Anne ""Hearthsinger"" Richmond🏕🌐🔜SWCA"       -2      negative
"Athena Ramos   6       positive
"Breanna Hetland        3       positive
"Daron Merryman 2       positive
"Dr. Umair A. Shah (MD  16      positive
"Elizabeth Newell       2       positive
"GeoVax -3      negative
"Jim Hougas     -2      negative
"KS ""Equilibrius"" Matar"      2       positive
"Malika Fair    -2      negative
"Maximilien ""Mjolnir tosser"" Guerrero"        -3      negative
"NO ""ism"" JUST MARX ""BROS""" -4      negative
"Private Joker  -12     negative
"Rahul Kumar    2       positive
"Samra Jones Bufkins MJ 2       positive
"Shalanda Webb  3       positive
"Torie Sepah    2       positive
"Trump was briefed on Nov       2       positive
"Win aka ""The Producer"""      -2      negative
"Yossarian ""Wear Masks. Wash Hands. Keep Distance."""  0       neutral
Time taken: 63.815 seconds, Fetched: 50 row(s)
hive>
```

Fig. 6. Tweets Classification

The categorization of tweet sentiment into negative, neutral, and positive categories is critical, as is the accurate measurement of the program's runtime for data output. These key indicators produced from program execution provide invaluable insights. The terminal interface, as seen in Fig. 7, is the completion of this critical procedure. This terminal display highlights the distribution of feelings throughout the tweet corpus and sheds illumination on the application's efficiency in terms of chronological execution.



```
OK
negative        5961
neutral 528
positive        6407
Time taken: 29.366 seconds, Fetched: 3 row(s)
```

Fig. 7. Sentiments Count

Use the Microsoft Power BI tool to create a visual representation of the results shown in Fig. 7. The detailed facts included within the output are transformed into an interesting visual story by utilizing the tools available of the Power BI. Figure 8 shows the examination of sentiment count analysis, which assists to comprehension by providing a clear portrayal of sentiment distribution patterns. This combination of visual insights improves understanding of the data's underlying patterns and qualities.

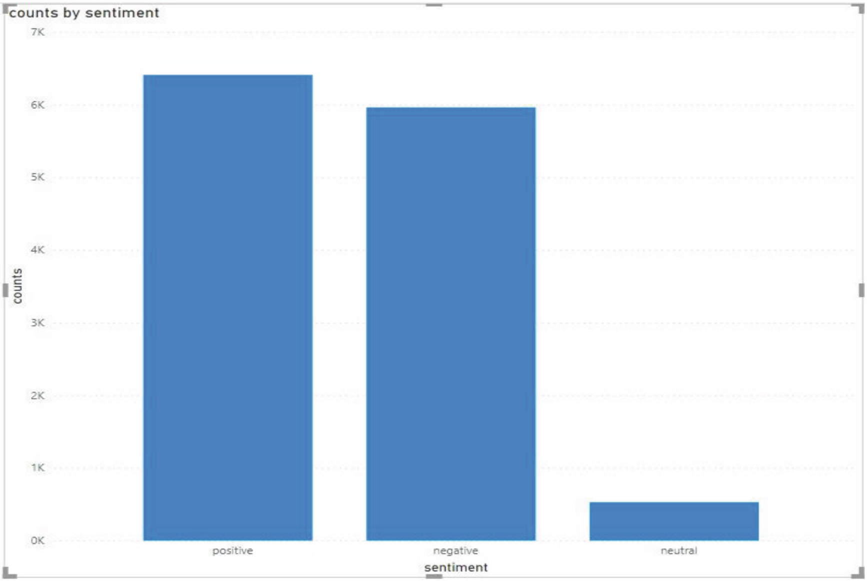


Fig. 8. Tweet sentiments count analysis

6 Conclusion and Future Research Directions

The Internet's rapid rise has led an increasing number of individuals voiced their opinions online. As a result, text-related big data is produced online. Natural language processing technology has emerged as a key tool for controlling public opinion in the age of big data by extracting sentiment patterns from vast amounts of online material. More than merely a social analytic tool, sentiment analysis has other uses as well. It's been a fascinating area of study. However, because to the complexity of this research, it is an area that is currently being investigated, albeit not in great detail. That is, the functions in this subject are too complex for machines to comprehend. For machines without feelings, it has been challenging to comprehend sarcasm, hyperbole, positive or negative emotions. The work analyzes, predicts, and gauge someone's opinion based on the users' tweets. The analysis is performed on covid tweets dataset. Using Dictionary text file, the sentiment analysis is done on tweets dataset and classified tweets into positive, negative, and neutral tweets. For better understanding of the result, visualization is performed on the result using BI tool. The emotions shown by people have not been predicted by algorithms with an accuracy of greater than 60%. However, despite various restrictions, this profession is expanding quickly across many industries. By analyzing and making predictions for massive datasets, sentiment analysis utilizing big data analytics can be very beneficial in business development and many other sectors.

In this research, tweet analysis work has been done on dataset that is downloaded from Kaggle. Based on result gained in this research, furthermore enhanced work can be done like by getting the data directly from twitter API. More upgraded work can be accomplished in future on basis of this study.

References

1. Karami, A., et al.: Twitter and research: a systematic literature review through text mining. IEEE Access **8**, 67698–67717 (2020)
2. Aladwani, A.M.: Facilitators, characteristics, and impacts of Twitter use: theoretical analysis and empirical illustration. Int. J. Inf. Manage. **35**(1), 15–25 (2015)
3. Li, H., Dombrowski, L., Brady, E.: Working toward empowering a community: how immigrant-focused nonprofit organizations use twitter during political conflicts. In: Proceedings of the 2018 ACM Conference on Supporting Groupwork (ACM 2018), pp. 335–346 (2018)
4. Wang, H., Can, D., Kazemzadeh, A., Bar, F., Narayanan, S.: A system for real-time twitter sentiment analysis of 2012 US presidential election cycle. In: Proceedings of the ACL 2012 System Demonstrations. Association for Computational Linguistics, pp. 115–120 (2012)
5. Bravo-Marquez, F., Frank, E., Pfahringer, B.: Annotate-sample-average (ASA): a new distant supervision approach for Twitter sentiment analysis. In: 22nd European Conference on Artificial Intelligence (ECAI), vol. 285, pp. 498–506. IOS Press (2016)
6. Li, M., Ma, Y., Cao, P.: Revealing customer satisfaction with hotels through multi-site online reviews: a method based on the evidence theory. IEEE Access **8**, 225226–225239 (2020)
7. Neviarouskaya, A., Aono, M.: Sentiment word relations with affect, judgment, and appreciation. IEEE Trans. Affect. Comput. **4**(4), 425–438 (2013)
8. Poria, S., Hazarika, D., Majumder, N., Mihalcea, R.: Beneath the tip of the iceberg: current challenges and new directions in sentiment analysis research. IEEE Trans. Affect. Comput. **14**(1), 108–132 (2020)
9. Khan, J., Ahmad, N., Khalid, S., Ali, F., Lee, Y.: Sentiment and context-aware hybrid DNN with attention for text sentiment classification. IEEE Access **11**, 28162–28179 (2023)
10. Braig, N., Benz, A., Voth, S., Breitenbach, J., Buettner, R.: Machine learning techniques for sentiment analysis of COVID-19-related twitter data. IEEE Access **11**, 14778–14803 (2023)
11. Mu, Y., Fan, Y., Mao, L., Han, S.: Event-related theta and alpha oscillations mediate empathy for pain. Brain Res. **1234**, 128–136 (2008)
12. Thompson, J.J., Leung, B.H., Blair, M.R., Taboada, M.: Sentiment analysis of player chat messaging in the video game StarCraft 2: extending a lexicon-based model. Knowl. Based Syst. **137**, 149–162 (2017)
13. Bloom, K., Garg, N., Argamon, S.: Extracting appraisal expressions. In: Proceedings of HLT/NAACL, Rochester, NY, pp. 308–315 (2007)
14. Cambria, E.: Affective-computing-and-sentiment-analysis. IEEE Intell. Syst. **31**(2), 102–107 (2016)
15. Ye, Q., Zhang, Z., Law, R.: Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. Expert Syst. Appl. **36**(3), 6527–6535 (2009)
16. Zhang, D., Xu, H., Su, Z., Xu, Y.: Chinese comments sentiment classification based on word2vec and SVMperf. Expert Syst. Appl. **42**(4), 1857–1863 (2015)

17. Yang, Q., Rao, Y., Xie, H., Wang, J., Wang, F.L., Chan, W.H.: Segment-level joint topic-sentiment model for online review analysis. IEEE Intell. Syst. **34**(1), 43–50 (2019)
18. Lee, S., Kim, W.: Sentiment labeling for extending initial labeled data to improve semi-supervised sentiment classification. Electron. Commer. Res. Appl. **26**, 35–49 (2017)
19. Onan, A., Korukoğlu, S., Bulut, H.: A hybrid ensemble pruning approach based on consensus clustering and multi-objective evolutionary algorithm for sentiment classification. Inf. Process. Manage. **53**(4), 814–833 (2017)
20. Shuang, K., Ren, X., Yang, Q., Li, R., Loo, J.: AELA-DLSTMs: attention-enabled and location-aware double LSTMs for aspect-level sentiment classification. Neurocomputing **334**, 25–34 (2019)
21. Abdi, A., Shamsuddin, S.M., Hasan, S., Piran, J.: Deep learning-based sentiment classification of evaluative text based on multi-feature fusion. Inf. Process. Manage. **56**(4), 1245–1259 (2019)

Cyber Threat Analysis and Mitigation in Emerging Information Technology (IT) Trends

Mohsin Imam[1], Mohd Anas Wajid[2], Bharat Bhushan[3], Alaa Ali Hameed[4], and Akhtar Jamil[5](✉)

[1] Department of Computer Science, ARSDC, University of Delhi, New Delhi, India
[2] School of Computing Science and Engineering, Galgotias University, Greater Noida, UP, India
[3] Department of Computer Science and Engineering (CSE) at School of Engineering and Technology, Sharda University, Greater Noida, India
[4] Department of Computer Engineering, Istinye University, Istanbul, Turkey
alaa.hameed@istinye.edu.tr
[5] Department of Computer Science, National University of Computer and Emerging Sciences, Islamabad, Pakistan
akhtar.jamil@nu.edu.pk

Abstract. For the information technology sector, cybersecurity is essential. One of the main issues in the modern world is sending information from one system to another without letting the information out. Online crimes, which are on the rise daily, are the first thing that comes to mind when we think about cyber security. Various governments and businesses are adopting a number of actions to stop these cybercrimes. A lot of individuals are still quite worried about cyber security after taking many safeguards. This study's primary goal is to examine the difficulties that modern technology-based cyber security faces, especially in light of the rising acceptance of cutting-edge innovations like server less computing, blockchain, and artificial intelligence (AI). The aim of this paper is to give readers a good overview of the most recent cyber security trends, ethics, and strategies. This study focuses on the present state of cyber security and the steps that may be taken to address the rising dangers posed by modern technology through a thorough investigation of the existing literature and actual case studies.

Keywords: Cyber-attack · Cyber-crime · Information Security · Social Media · Blockchain · Cyber Security

1 Introduction

With a simple click, we can now transmit various forms of information through email, audio, or video. But have we pondered the security of our data during these transfers? The solution lies in the realm of cybersecurity. The internet is rapidly becoming an integral part of contemporary life. Transformative technological advancements are reshaping human existence. These progressions, however, are also fostering an uptick

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 570–588, 2024.
https://doi.org/10.1007/978-3-031-56728-5_47

in cybercrime, as they complicate the protection of personal data. Given that over 60% of business transactions occur online, maintaining a robust security infrastructure is imperative for optimal and transparent operations. The significance of cybersecurity is increasingly evident [1]. Beyond merely securing IT data, the domain of cybersecurity encompasses a wide array of other areas, such as cyberspace.

Modern advancements like cloud computing, mobile tech, and online banking require robust security due to the inclusion of sensitive personal data. Safeguarding these technologies is crucial, as they play a vital role in a nation's security and economic stability. Enhancing cybersecurity and ensuring the protection of critical information systems are vital for a country's wellbeing. The evolution of services and government strategies hinges on creating a safer online environment for Internet users [2].

Information technology professionals frequently approach cybersecurity challenges from a technical, IT perspective. Current research advocates for an all-inclusive strategy that takes organizational psychology, commercial objectives, governance, risk management, and other elements like those mentioned in the Clinger-Cohen Act into account [3]. According to the Systems and Nonlinear Theory, issues should be addressed at every level of the organization and from every perspective. Due to the sheer amount of data, expiration of the data, technology shifts, and the numerous parties and information involved, cybersecurity is largely a knowledge management issue [4]. Some approaches and tools to deal with these difficulties are provided by business intelligence and analytics. From these aspects, cybersecurity is examined in this research.

The paper's structure is outlined as follows: Sect. 2 provides insights into Cyber Security, covering aspects of cybercrime and details about phishing. Section 3 delves into the Challenges and Evolution within the realm of Emerging Technologies, discussing Artificial Intelligence, Blockchain, Internet of Things, and Cloud Computing. Moving to Sect. 4, the focus is on Challenges and Evolution in Cybersecurity amidst the emergence of new technologies, including Web Servers, Advanced Persistent Threats (APTs), Targeted Attacks, Mobile Networks, Novel Internet Protocols, Encryption Techniques, Social Engineering, Phishing Scams, and the rise of state-sponsored cyber attacks and cyber espionage. Lastly, Sect. 5 expounds on the ascent of Advanced Cyber Threats.

2 Cyber Security

Data security and privacy are paramount concerns for businesses. In today's digital era, information is predominantly stored in electronic formats. Platforms such as Facebook and Instagram provide users with a familiar space to engage with family and friends. Cyber attackers frequently focus on social networking platforms to access personal details, particularly from home users. Therefore, adopting rigorous security measures is crucial when using social media and conducting online banking transactions.

Most of the time, knowledge management and business intelligence are not thought to be relevant to cybersecurity. However, it should have a vital role in cybersecurity. There is more to cybersecurity than just the tools. The attacker's aim are the company processes and information, especially intellectual property. It's also possible that the attack is motivated by greed or intellectual humiliation. Attackers use people as well as gaps in information technology to their advantage. As a result, every aspect of an organization is revealed. As a result, everyone gets involved in cybersecurity. With knowledge management, enterprise intelligence, and analytics, an organization can collect and analyze its mission-related information, intellectual property, and link the stakeholders to the information that it holds.

For numerous companies, handling cybersecurity in an ever-changing threat environment poses difficulties. The old-fashioned responsive methods that allocated resources to shield systems from the biggest established dangers, while leaving smaller risks unguarded, are no longer effective. To stay well-informed of evolving security threats, it is imperative to adopt a forward-thinking and flexible strategy. Prominent advisory bodies in the field of cybersecurity provide counsel. As an illustration, the National Institute of Standards and Technology (NIST) advocates the utilization of perpetual surveillance and instantaneous evaluations to protect against acknowledged and undisclosed hazards within a risk evaluation framework (Fig. 1).

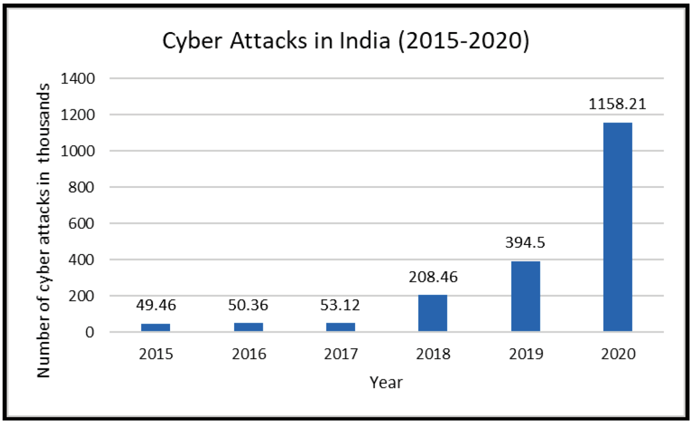


Fig. 1. Cyber Attacks in India (2015–2020)

The above graph illustrates the trend of cyberattacks in India over the past five years, and it is obvious that the rate of cyberattacks is rising exponentially, which clearly exhibits the cyber security threats.

A rise in security breaches and hacking incidents can be attributed to factors like inadequate data protection, the global epidemic's impacts, and an upsurge in exploit complexity. These issues predominantly arise from everyday workplace elements such as mobile devices and IoT gadgets. The prevalence of remote work due to COVID-19 has further exposed vulnerabilities, making the landscape susceptible to cyberattacks.

Based on recent findings in the field of security research, most companies possess inadequate cybersecurity measures, putting them at risk of data breaches. In order to effectively counteract harmful motives, organizations need to engrain cybersecurity consciousness, proactive measures, and established norms within their ethos. Presented below are several cybersecurity figures impacting numerous individuals and their private data.

- Each year, cybercrime alone affects the American economy $100 billion: Both ordinary persons and major companies, like the U.S. Navy, which experiences over

100,000 intrusions every hour, are targeted by cybercriminals. Over 100 million Americans' personal medical records that belonged to them were stolen in 2016.

- In 2016, only three industries—government, retail, and technology—accounted for 95% of records that were compromised. It's not necessarily the case that such industries are less vigilant about protecting client information. They are just very sought-after targets because to the abundance of personally identifying information are present in their databases.
- Small businesses are the target of 43% of cyberattacks.64% of firms have experienced web-based attacks. 62% of respondents had encountered social engineering and phishing scams. Botnets, malicious code, and denial-of-service attacks were all encountered by 59% and 51% of the organizations, respectively. Small businesses, defined as those with fewer than 500 employees, spend $7.68 million on average every incident.
- Cybercriminals' preferred method with respective rising trend is ransomware, which is malicious software that holds a victim's data captive until a ransom is paid. Using ransomware, a hacker can directly extort money from the victim rather than selling the victim's personal information on the underground market. The ransomware threat is predicated on either publishing the victim's personal information online (doxing) or denying them access to their online accounts.

2.1 Cyber Crimes

Cybercrime is the term for illegal activity carried out through digital tools like the Internet and electronic communication systems. This form of illicit conduct can result in broad effects on people, groups, and the entirety of society [5]. Cyberattacks, stealing personal information, deceptive activities, and the distribution of harmful software stand as some of the most common manifestations of online criminal acts. The impact of cybercrime can be significant, as it can result in financial losses, theft of intellectual property, and disruption of critical infrastructure [6].

Despite the growing awareness of cybercrime, many individuals and organizations still lack effective measures to prevent or mitigate these types of attacks [7, 8]. The reason for this is often a lack of knowledge about the best practices for cyber security, as well as the rapid pace of technological change, which can make it difficult for individuals and organizations to keep up [9].

According to the report by McAfee, the cost of cybercrime worldwide as of 2018 was above $1 trillion. The estimated cost of lost revenue due to cybercrime is $945 billion, and by 2020, it is anticipated that the global budget for cybersecurity would surpass $145 billion. The impact of this on the world economy is $1 trillion.

This McAfee report on the price of cybercrime is the fourth in the series. In earlier publications, data from non-attribution conversations with cybersecurity officials and publicly accessible information on national losses were examined. According to McAfee's 2018 research, cybercrime cost the world economy more than $600 billion. Figure 2 illustrates the newest forecast from 2018 to 2027, which predicts a more than 2700% growth in the span of 10 years.

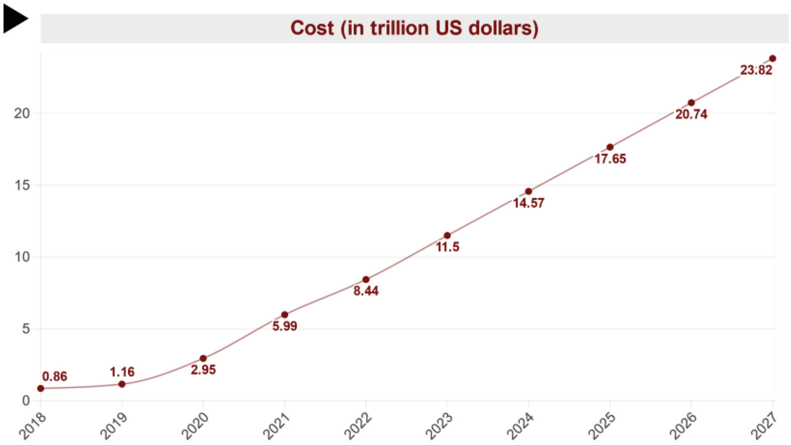


Fig. 2. Estimated Average Cost of Cybercrime

It is essential that people and businesses take proactive safety steps to combat the increasing risk of cybercrime. This includes staying up-to-date with the latest cybersecurity technologies, implementing strong password protocols, and educating employees about safe online practices [10]. Organizations should also spend money on comprehensive cybersecurity solutions that can stop cyberattacks before they have a chance to do real damage. Firewalls, intrusion detection systems, and encryption technologies can all fall under this [11].

To ensure that cybercrime remains a priority for governments, organizations, and individuals, it is also important to continue research in the field of cybersecurity. This includes exploring new technologies for detecting and preventing cyber-attacks, as well as developing educational programs and outreach initiatives to raise awareness about the importance of cyber security [12].

2.2 Malwares

Malicious software, often called malware, encompasses any software or document designed to cause damage to a computer, network, or server. Various forms of malware exist, including computer viruses, worms, Trojan horses, ransomware, and spyware, all capable of disrupting or commandeering vital computer functions. These harmful applications can modify or seize essential computing operations, purloin important information, encode it, and erase it. Furthermore, they have the capacity to observe the actions of users within a system. Due to advancements in computer software, new generation malware may now carry out more devastating operations, including previously unseen focused and persistent attacks as well as attacks that mix multiple malware types. The Table 1 below shows both conventional malware and the most recent generation of malware.

Table 1. Traditional vs New Generation Malwares

Comparison Criteria	Classical malware	New Generation Malwares
Implementation Level	Simple Coded	Hard Coded
Proliferation	Each copy is identical	Every version is dynamic
Extension for spreading	.exe extension is used	Employs various extension
Permanency	Temporal	Persistent
Interaction with System	Few Processes	Multiple Processes
Attack Type	All-purpose	For Targeted Attacks
Defensive Challenge	Easy	Hard
Specified Devices	General Computers	For Wide range of Devices

2.3 Phishing

Phishing is a sort of cyber-attack that preys on people or organizations by impersonating a reliable source. The attacker asks the receiver for personal information, like passwords or credit card details, through a message or email that looks to be from a credible source, like a bank, a government agency, or a well-known firm. The victim's personal details or money are then stolen using the information.

As per research conducted by the Anti-Phishing Working Group (APWG), phishing stands as the prevailing type of cyber assault, constituting more than 80% of documented security occurrences [13].The APWG estimated an average of about 500,000 phishing sites per month in 2020, with the COVID-19 pandemic seeing a notable spike in attacks. The study also discovered that phishing attempts are evolving, with attackers utilizing strategies such as personalized email addresses, realistic logos, and official-looking URLs to make the attack seem legitimate.

Phishing is a serious and growing threat to cyber security, and proactive behavior is crucial for both individuals and organizations in protecting against these attacks. By implementing effective cybersecurity measures and being vigilant about potential phishing attempts, people and organizations can reduce their vulnerability to these threats.

3 Challenges and Evolutions in the Era of Emerging Technologies

The development of developing technologies has created many difficulties for the cybersecurity industry. The rise of advanced innovations such as Artificial Intelligence, the Internet of Things, and cloud computing has given rise to a constantly shifting environment of risks, growing in intricacy. As a result, it is crucial for both businesses and people to take a proactive and watchful approach to cybersecurity. This requires regular assessment of their security posture and the implementation of the most current technologies and best practices to mitigate risks [14]. The rapid pace of technological advancement is expected to continue to present new challenges and opportunities in the field of cybersecurity, making it a continuously evolving discipline that requires ongoing study and practical application. This section will delve into the impact of the latest technological advancements on cybersecurity and the necessary transformations that are required to ensure its continued efficacy.

3.1 Artificial Intelligence

Artificial Intelligence (AI) has revolutionized the way in which many industries operate, providing new opportunities for automation, optimization, and decision making.

AI has captured substantial attention in scholarly investigations, encompassing various forms of artificial intelligence such as machine learning, computer vision, and deep learning. These methodologies have displayed remarkable progress, effectively addressing real-world challenges spanning from visual analysis to processing human language.

However, the quick uptake of AI has also brought up significant moral and cultural concerns about how technology may affect privacy, responsibility, and transparency. Concerns arise over AI's impact on the labor market and workforce, as well as its propensity to reinforce preexisting biases.

Lately, there's been a growing fascination with incorporating artificial intelligence (AI) through machine learning techniques into the field of cybersecurity.AI offers a wide range of potential benefits to the field of cybersecurity, including improved threat detection, faster and more accurate incident response, and better risk assessment and management [15]. However, AI also presents a number of potential risks, including issues related to bias, transparency, and ethics which is summarized in Table 2. These limitations and challenges must be carefully considered by organizations and individuals when making decisions about the use of AI in cybersecurity. To ensure the appropriate and successful application of AI in the sphere of cybersecurity, it is crucial to fully comprehend and solve these difficulties as AI technology develops.

The advent of deepfake technology has introduced a new dimension to the realm of cyber attacks. Deepfakes are artificial intelligence-generated images, audio, or video that mimic real individuals or events. In the context of cybersecurity, deepfakes can be used to create malicious and deceptive content, such as fake news or impersonating a trusted individual to steal sensitive information. This type of attack can be particularly dangerous because deepfakes are often highly convincing, making it difficult for individuals to differentiate between real and fake content [16]. As the technology behind deepfakes continues to advance, the potential for deepfake-based cyberattacks becomes more concerning, highlighting the companies and people must be on guard and proactively put security measures in place to reduce these risks.

Adversarial machine learning-based cyberattacks refer to a type of attack that exploits weaknesses in artificial intelligence (AI) systems to achieve malicious goals. Adversarial machine learning operates by introducing small perturbations, such as an imperceptible change in an image or a minor modification in the input data, which can significantly affect the AI system's output. These attacks can bypass conventional security measures that are typically effective against other types of cyberattacks [17]. The objective of adversarial machine learning-based attacks is to manipulate the AI system's behavior, making it behave in a manner that is unexpected, potentially harmful, and difficult to detect.For example, in the case of image recognition systems, adversarial attacks can manipulate an image to the extent that it appears normal to the human eye, but the AI system misidentifies it. Similarly, in the case of speech recognition systems, adversarial attacks can manipulate an audio file to produce an output that is different from what was intended by the user. Adversarial machine learning-based attacks pose a serious risk to the safety and dependability of AI systems, as they are often highly effective and difficult to detect.

It is crucial to create new methods and strategies that can recognize and thwart adversarial machine learning-based assaults in order to reduce the hazards they bring. This can include incorporating adversarial machine learning into the AI system's training process, using machine learning-based detectors, or using formal methods to verify the robustness of AI systems against adversarial attacks. By considering these factors, researchers and practitioners can help ensure that AI systems remain secure and reliable even in the face of adversarial machine learning-based attacks.

Autonomous weapon systems, also known as lethal autonomous weapons, are a growing concern in the realm of cybersecurity and are often seen as a product of the rapidly advancing field of artificial intelligence. These systems, which can be used to conduct cyberattacks, are created to operate autonomously and without human interaction. The use of autonomous weapons might significantly increase the security risks already present and endanger both organizations and people. They are programmed to operate in a pre-defined manner and can quickly scale an attack without the need for human involvement, making them difficult to detect and defend against. Additionally, autonomous weapon systems, powered by artificial intelligence, can be programmed to evade traditional security measures, making them even more dangerous. Significant ethical and legal issues are raised by the use of autonomous weapon systems, particularly those pertaining to accountability and responsibility for their acts. In order to reduce the risk involved with the use of self-navigating weapons in the realm of cybersecurity, it is crucial that the proper precautions be taken.

They present significant challenges in the realm of cybersecurity. These systems are designed to operate independently, without human intervention, and can be used to carry out cyberattacks. The deployment of autonomous weapon systems exacerbates existing security threats and poses a significant risk to organizations and individuals. One of the major difficulties associated with autonomous weapon systems is their ability to operate quickly and efficiently. They are programmed to operate in a pre-defined manner and can rapidly scale an attack without the need for human involvement, making them difficult to detect and defend against. Additionally, autonomous weapon systems can be programmed to evade traditional security measures, making them even more dangerous.

The moral and legal ramifications of using autonomous weapon systems provide another difficulty. Significant concerns about accountability and responsibility for their acts are raised by the implementation of these systems. The question of who is accountable for the results of their activities is brought up by the deployment of autonomous weapon systems in cyberattacks. A challenging and constantly changing issue in cybersecurity is the use of autonomous weapon systems. It is essential that the right safeguards are put in place to reduce the risk posed by these technologies and guarantee their responsible use.

3.2 Blockchain

Blockchain technology offers safe and transparent transaction record-keeping through a decentralized, distributed ledger system. It utilizes cryptographic algorithms to secure the data and ensures its integrity, making it immutable and tamper-proof. Since Bitcoin's foundational technology was introduced in 2008, the term "blockchain" has been used. But in recent years, it has become well known and has been used in a variety of sectors, including banking, healthcare, and supply chain management. Blockchains' decentralized structure increases security and eliminates the need for middlemen, lowering the possibility of fraud and cyberattacks. Additionally, the automatic execution of contract terms made possible by smart contracts on blockchains lowers the possibility of human error and improves efficiency.

However, the increasing adoption of blockchains has also raised new challenges and problems in the domain of cybersecurity. The decentralized nature of blockchains can make it challenging to detect and remediate security breaches, particularly in the case of malicious actors who have gained control over a significant portion of the network. Additionally, the immutability of blockchains can make it difficult to roll back unauthorized transactions or to correct errors.

The rise of blockchain technology has presented a new set of challenges and opportunities for cybersecurity. As blockchain becomes more widely adopted across various industries, secure and effective cybersecurity measures are becoming more and more necessary. The approach to cybersecurity needs to change in order to reflect the distinctive qualities and capabilities of blockchain technology in order to meet these concerns. The decentralization of data, which makes it challenging for businesses to effectively safeguard their systems, is one of the main problems presented by blockchain. Because of its decentralized nature, blockchain is more open to assault because there is no central authority in charge of running and protecting the network. This necessitates the implementation of suitable security measures to guard against nefarious individuals and potential threats.

In recent years, the use of cryptocurrency and blockchain-based systems has grown in popularity, leading to an increase in the number of cyber-attacks targeting these systems, a statistic of amount of cryptocurrency stolen in Q1 of 2021–2022 shown in Fig. 3. These attacks can take many forms, including the theft of private keys, the exploitation of software vulnerabilities, and the manipulation of transaction records. Therefore, when adopting cryptocurrencies and blockchain-based systems, individuals and businesses must be alert and pro-active in their approach to cybersecurity. This involves regularly reassessing their security posture, implementing strong security protocols, and remaining informed about the latest trends and threats in the cyber threat landscape.

The moral and legal ramifications of using autonomous weapon systems provide another difficulty. Significant concerns about accountability and responsibility for their acts are raised by the implementation of these systems. The question of who is accountable for the results of their activities is brought up by the deployment of autonomous weapon systems in cyberattacks. A challenging and constantly changing issue in cybersecurity is the use of autonomous weapon systems. It is essential that the right safeguards are put in place to reduce the risk posed by these technologies and guarantee their responsible use.

The rise of blockchain also presents an opportunity to improve the overall security of systems through decentralized and distributed systems. Due to the fact that data is held among numerous nodes rather than in a single location, the decentralized nature of blockchain, for instance, can make it simpler to secure data and avoid data breaches. In addition, blockchain can be used to strengthen system security through the use of secure transactions and smart contracts, lowering the likelihood of cyberattacks and data breaches. Blockchain technology requires a transformation in the approach to cybersecurity to effectively address the unique challenges and opportunities posed by this technology. This requires a proactive, comprehensive, and forward-thinking approach to security, to ensure that organizations and individuals are protected against potential threats and risks in the era of blockchain.

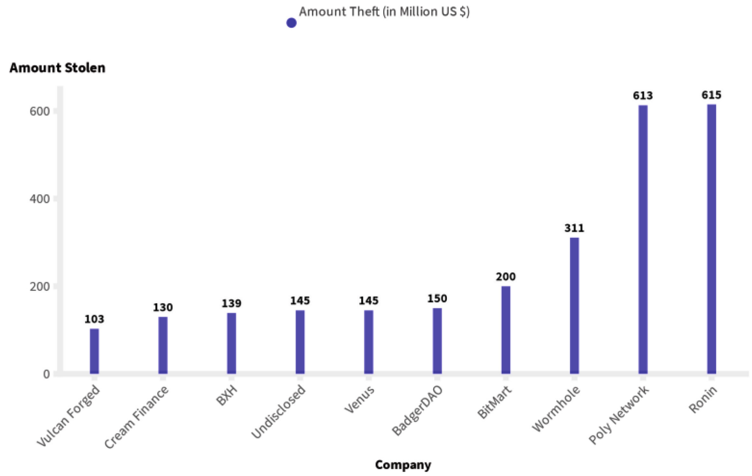


Fig. 3. Top ten cryptocurrency theft in 2021–2022

3.3 Internet of Things

The term "Internet of Things" (IoT) refers to the expanding network of physical objects, such as machinery, automobiles, home appliances, and other objects, which are connected to one another and share data. These technologies are becoming more connected to one another and more integrated into our daily lives, which has created new opportunities and efficiencies but also new security risks. It is crucial for businesses and individuals to stay knowledgeable and proactive in their approach to IoT security as more and more devices are connected to the internet, increasing the attack surface for potential cyberattacks. With an increase in the deployment of linked devices and sensors in private residences, commercial buildings, and public areas, the field of the Internet of Things (IoT) has emerged as one that is fast expanding. The security of these linked devices has become increasingly evident as the adoption of IoT technology has increased. IoT systems are particularly vulnerable to cyberattacks because to their complexity and interconnectedness, which can lead to data theft, unauthorized access to private data, and a variety of other security-related events [18]. Because of this, it is crucial that cyber security change in reaction to the growing use of IoT technology. In order to mitigate risk and guarantee the protection of sensitive data, security experts must be active in their approach, continually evaluating their security posture, and implementing the most recent technologies and best practices.

The interconnection and communication of devices has been completely transformed by the Internet of Things (IoT). But substantial security flaws have also been introduced into networks and systems due to the growing use of IoT devices. IoT devices frequently have insufficient security measures because they are made to be affordable and simple to use. This makes them a desirable target for cybercriminals who can use the devices' vulnerabilities to obtain unauthorized access to sensitive data or interfere with the proper operation of systems. It is becoming more crucial for businesses and individuals to comprehend the security threats posed by IoT devices and take the necessary precautions to reduce them as the number of connected devices increases. The risk of cyberattacks has increased due to the exponential development in the number of IoT devices, which has increased the possible attack surfaces for cybercriminals. IoT device growth has produced multiple chances for hostile actors to obtain illegal access and compromise sensitive information, necessitating the adoption of strong cybersecurity measures by both businesses and people.

The amount of data collected and saved has exponentially increased as a result of the spread of Internet of Things (IoT) devices. This glut of data, however, also poses a serious threat to cyber security. IoT devices have accumulated enormous volumes of sensitive data, making them a potential target for bad actors looking to conduct data breaches. Sensitive information theft or unauthorized access can cause a great deal of harm to people and businesses, including loss of reputation, monetary losses, and legal repercussions. In order to reduce the risk associated with the deployment of IoT devices, it is crucial that proper safeguards be put in place, especially with regard to data breaches and the safeguarding of sensitive data.

Internet of Things (IoT) technology adoption is growing, which has created new cybersecurity security issues. The lack of security solutions particularly created for IoT devices is one of the main obstacles. As a result, it may be difficult to effectively defend against cyberattacks since IoT devices have special vulnerabilities that typical security solutions may not be able to address. The lack of standards for IoT devices can also cause interoperability problems, raising the danger of cyberattacks. These problems emphasize the significance of creating thorough and efficient security solutions that can handle the particular requirements and risks of IoT technology [19].

3.4 Cloud Computing

In order to provide quicker innovation, adaptable resources, and scale economies, cloud computing refers to the supply of computer resources, including servers, storage, databases, networking, software, analytics, and intelligence, over the Internet. The use of cloud computing has grown substantially over the past few years, enabling businesses and individuals to store vast amounts of data, process it, and instantly access cutting-edge applications. The increased usage of cloud computing has brought about a number of cybersecurity issues, including data breaches, unauthorized access, and the theft of private data [20]. As a result, organizations must implement effective security measures and practices to protect their data and systems in the cloud. The increasing complexity of the threat landscape and the dynamic nature of the cloud environment make it imperative that organizations adopt a proactive and risk-based approach to cybersecurity in the cloud. Cloud computing has emerged as a transformative technology, offering organizations and individuals significant benefits in terms of cost, scalability, and accessibility. However, the rapid adoption of cloud computing has also raised a number of significant security challenges, including data breaches, unauthorized access, and security threats to cloud infrastructure. The challenge lies in maintaining the confidentiality, integrity, and availability of data while it is stored and processed in the cloud.

The shared responsibility approach, which calls on both the cloud service provider and the customer to take proactive measures to maintain security, is one of the main obstacles to protecting cloud systems. In order to effectively secure cloud computing environments, organizations need to implement comprehensive security strategies that address the specific risks inherent in the cloud. In order to identify and address security incidents, this can involve using monitoring and logging tools, multi-factor authentication, and encryption technologies. Additionally, firms using cloud computing must be careful to protect their data even after it has been transferred to the cloud. To address this, security must be proactive, with regular monitoring, security assessments, and the adoption of best practices and guidelines all being key components. The transformation of cyber security in the era of cloud computing should focus on enhancing the security of cloud infrastructures and improving the safety of cloud-based data, while also addressing the shared responsibility model and the specific security challenges associated with cloud computing. The increased utilization of cloud computing has resulted in a significant expansion of the attack surface for organizations, the types of attacks are summarized in Table 2. As organizations move sensitive data and operations to the cloud, they are exposing this information to an increased risk of potential cyber-attacks. In order to reduce the danger of cyberattacks, it is necessary for enterprises to be aware of the security problems and apply strong security measures. The enlarged attack surface can include cloud infrastructure, platforms, and apps. Additionally, in order to keep ahead of changing security risks and guarantee the protection of sensitive data, organizations must constantly monitor and review their security posture. Due to these difficulties, cyber security must change to keep up with the evolving threat environment and successfully reduce the dangers related to cloud computing.

Moreover, the increasing adoption of cloud computing has caused the attack surface to grow exponentially, making it a prime target for cyber attackers. The cloud environment, which typically involves multiple tenants and data storage centers, presents unique security challenges and increases the risk of data breaches. It is crucial to secure data privacy and compliance with laws like the General Data Protection Regulation (GDPR) in a cloud environment because enterprises are entrusted with various tenants' sensitive and private data. Ample precautions must be taken to guarantee the privacy and security of data kept in the cloud, including routine evaluations and audits to confirm adherence to laws and industry best practices for data protection. In order to discover potential vulnerabilities and adopt strong security measures to reduce the risk of cyberattacks and safeguard sensitive data, businesses must also constantly monitor and review their security posture.

As these challenges continue to evolve, it is imperative that the approach to cybersecurity must also adapt and transform to effectively address these issues.

Table 2. Types of Cloud Computing Attacks

Attack Types	Description	Impact	Mitigation
Cloud Storage Breaches	Unauthorized access to cloud storage systems	Stolen or lost sensitive data	Encryption, access control, and continuous monitoring
Denial of Service (DoS)	overloading a system with users so that it becomes unusable	Downtime and lost productivity	Network filtering, firewalls, and DDoS protection services
Injection Flaws	Exploiting vulnerabilities in application code to inject malicious content	Data theft, loss of service, or reputational damage	Input validation, sanitization, and encryption
Malicious Insider Threat	Insider abuse of privileges to steal, damage, or misuse data	Sensitive data theft or loss, system outages, and legal repercussions	Background checks, user activity monitoring, and access controls

4 Challenges and Evolutions in Cybersecurity in the Era of Emerging Technologies

4.1 Web Servers

Web servers have become a crucial aspect of modern cyber security, with an increasing number of organizations relying on them to host and manage their online presence. The rise of cloud computing has further amplified the importance of web servers, as they are now being used to host both internal and external-facing applications. The increase in the use of web servers has brought about new challenges and security risks, as they are often targeted by cyber criminals who seek to steal sensitive information or compromise the availability of critical systems. In response to these challenges, organizations must ensure that their web servers are properly secured and maintained, using a combination of network security technologies and best practices. This requires a proactive approach to cyber security, as well as a commitment to staying up-to-date with the latest security trends and threat intelligence.

4.2 APT's and Targeted Attacks

Advanced persistent threats (APTs) and targeted attacks are among the top trends changing the landscape of cybersecurity. APTs are a type of cyber-attack that involves prolonged and targeted network exploration and exploitation of a victim [21]. APTs, in contrast to conventional cyber-attacks, are made to go undetected for a protracted amount of time, enabling the attacker to gather private data and obtain access to vital systems. On the other hand, targeted attacks are made with a specific organization or person in mind. As they are frequently carried out by state-sponsored organizations or highly competent cybercriminals, both APTs and targeted assaults constitute a serious threat to both enterprises and individuals. Cybersecurity must constantly change and adapt to counter these emerging threats, making use of cutting-edge tools like artificial intelligence, machine learning, and large-scale data analytics.

4.3 Mobile Networks

The rapid growth and adoption of mobile networks have significantly altered the cybersecurity landscape. Mobile networks provide a new vector for cyberattacks, as they are often used for sensitive transactions and accessing critical information. Additionally, the limited security measures on many mobile devices make them vulnerable to attack. Furthermore, the use of mobile networks for personal and business purposes has increased the attack surface, as attackers now have access to a wider range of confidential data. Effective safety precautions for mobile devices and networks, such as encryption, secure access controls, and frequent software updates, must be put in place by enterprises to solve these issues. The increasing use of mobile networks is driving the need for more sophisticated security solutions that can protect against the growing threat of mobile cyberattacks.

4.4 New Internet Protocols

The advent of the new Internet Protocol (IPv6) is bringing significant changes to the landscape of cybersecurity. IPv6, the successor to the current Internet Protocol (IPv4), is designed to support a much larger number of devices with Internet access, which expands the attack surface. As more and more devices are added to the network, it becomes increasingly difficult to secure these devices, making them vulnerable to various types of cyber-attacks. Moreover, the implementation of IPv6 is not always straightforward and requires organizations to make significant changes to their existing security systems. This increases the risk of misconfiguration and introduces new security challenges, requiring organizations to adopt new security strategies and techniques. In order to tackle these challenges, organizations must embrace the latest security technologies, such as network segmentation, firewalls, and encryption, to maintain the security of their network and protect against possible threats.

The advent of the new Internet Protocol (IPv6) is bringing significant changes to the landscape of cybersecurity. IPv6, the successor to the current Internet Protocol (IPv4), is designed to support a much larger number of devices with Internet access, thereby increasing the attack surface. As more and more devices are added to the network, it becomes increasingly difficult to secure these devices, making them vulnerable to various types of cyber-attacks. Moreover, the implementation of IPv6 is not always straightforward and requires organizations to make significant changes to their existing security systems. This increases the risk of misconfiguration and introduces new security challenges, requiring organizations to adopt new security strategies and techniques. In order to tackle these challenges, organizations must embrace the latest security technologies, such as network segmentation, firewalls, and encryption, help keep their network secure and guard against any threats.

4.5 Encryption Techniques

The development of encryption technologies has significantly influenced the condition of cybersecurity today. Because cyberattacks are becoming more sophisticated, corporations now routinely use encryption to safeguard their critical data. Advanced encryption algorithms, such as AES and RSA, have been developed to provide stronger protection against cyber criminals who are seeking to access sensitive information. Moreover, the rise of cloud computing has further emphasized the importance of encryption, as more and more data is being stored in the cloud, making it vulnerable to cyber-attacks. As a result, organizations are constantly looking for new and improved encryption techniques to safeguard their data. The rapid evolution of encryption techniques is changing the landscape of cybersecurity by providing organizations with more robust and secure ways to protect their data, and as a result, it is becoming a critical aspect of any effective cyber security.

4.6 Social Engineering and Phishing Scams

Among the fast expanding phenomena that are reshaping the cybersecurity landscape are social engineering and phishing attacks. Social engineering assaults use psychological deception to get people to reveal private information, including passwords and financial information. On the other side, phishing scams use false emails, texts, or websites to steal personal information. These assaults pose a serious threat to both persons and companies since they are becoming more sophisticated and difficult to detect. Therefore, cybersecurity must incorporate efficient phishing prevention technologies, user awareness training, and continual monitoring for suspicious activity to respond to these changing threats. A comprehensive strategy that considers the technological, social, and psychological elements of these types of attacks is required to properly address the challenge posed by these trends.

4.7 Increase in State-Sponsored Attacks and Cyber Espionage

State-sponsored assaults and cyberespionage are on the rise, whichhas been a major trend that has transformed the landscape of cybersecurity. With the rise of geopolitical tensions and increasing use of technology in sensitive industries, state-sponsored actors have been exploiting vulnerabilities in organizations to gain access to confidential information and intellectual property. These attacks are highly sophisticated, well-funded, and can target both large corporations and small businesses alike. The need for organizations to protect against state-sponsored attacks has increased, resulting in a need for stronger cybersecurity measures and increased investment in cybersecurity technology. The challenge of detecting and defending against these types of attacks has become even greater, as these actors often use advanced techniques to evade detection and cover their tracks. Organizations need to adopt a proactive strategy to defend against these attacks, investing in cutting-edge security solutions, incident response preparation, and routine threat assessments to keep ahead of the changing threat landscape.

The increase in state-sponsored attacks and cyber espionage has been a major trend that has transformed the landscape of cybersecurity. With the rise of geopolitical tensions and increasing use of technology in sensitive industries, state-sponsored actors have been exploiting vulnerabilities in organizations to gain access to confidential information and intellectual property. These attacks are highly sophisticated, well-funded, and can target both large corporations and small businesses alike. The need for organizations to protect against state-sponsored attacks has increased, resulting in a need for stronger cybersecurity measures and increased investment in cybersecurity technology. The challenge of detecting and defending against these types of attacks has become even greater, as these actors often use advanced techniques to evade detection and cover their tracks.

5 Rise of Advanced Cyber Threats

The growth of sophisticated cyber-attacks has recently become a top concern for both businesses and individuals. Our world is becoming more interconnected, and we are relying more and more on technology, which has greatly raised the possibility of cyber-attacks. Due to this, increasingly complex cyber threats have emerged. These threats are made to bypass conventional security measures and steal important data. The necessity for sophisticated cyber security measures to combat these threats is highlighted by their growing sophistication.

Advanced cyber threats come in many forms, including Zero-day vulnerabilities, targeted assaults, and advanced persistent threats (APTs). APTs, in particular, are becoming increasingly prevalent and are often used to conduct long-term, targeted attacks on organizations. These attacks are made to hide from detection and steal private information for a long time. On the other hand, targeted attacks are made to take advantage of a particular weakness in a system or application. Zero-day exploits, on the other hand, are attacks that take advantage of unknown vulnerabilities in systems or applications.

The increasing sophistication of these threats requires organizations to take anactive and multi-layered approach to cyber security. In order to do this, sophisticated security technologies must be implemented, including firewalls, systems for detecting and preventing intrusions, and security information and event management (SIEM) programs. In addition, organizations should also implement strong password policies, regularly update software and systems, and educate employees on best security practices.

The application of artificial intelligence (AI) and machine learning (ML) technology is a key component of modern cyber security. These technologies can offer enterprises a higher level of safety by assisting in the real-time detection and reaction to cyberthreats. Large-scale data analysis and the detection of patterns suggestive of a cyber-attack can be done using AI and ML. This enables firms to respond to threats rapidly and lessen the effects of an assault.

As sophisticated cyber threats increase, businesses must adopt a proactive, multi-layered strategy to cyber protection. This entails installing cutting-edge security tools, putting in place strict security regulations, and utilizing AI and ML tools. These steps can help firms defend themselves against the increasingly sophisticated cyber threats and guarantee the protection and safety of sensitive data.

6 Future Trends in Cyber Security

To keep up with the increasingly sophisticated nature of cyber threats, the field of cyber security is continually evolving. Future emphasis is anticipated to shift toward more sophisticated security measures that take advantage of new technologies. The rising use of artificial intelligence (AI) in security solutions is one of the major developments in the sector. Organizations may more easily identify and respond to cyber threats in real time by using AI-powered systems that can automate threat detection and response. Furthermore, AI-powered solutions can deliver more sophisticated threat intelligence, enabling businesses to better comprehend the threat landscape and take preventative action to lessen risks.

Cloud security is another new development in cyber security. With the widespread use of cloud computing, it is now more crucial than ever to guarantee the security of data stored there. This is due to the fact that cloud environments are frequently accessed from anywhere in the world, which makes them open to cyberattacks. Organizations must put strong security mechanisms, such encryption and access controls, in place to secure their cloud environments in order to reduce this risk.

Another development in the area of cyber security that is picking up steam is biometric authentication. Compared to more conventional authentication techniques like passwords and security tokens, biometric authentication can offer a higher level of security. By authenticating individuals before granting them access to critical data, biometric authentication techniques like facial recognition and fingerprint scanning can add an extra layer of security. Biometric authentication is becoming more and more popular, thus it is going to be important in the future when it comes to protecting the digital environment. Emerging trends including powered by artificial intelligence security solutions, cloud security, and biometric authentication are anticipated to have an impact on the direction of cyber security in the future. In order to stay ahead of the changing threat landscape and maintain the security of their data and systems, organizations must be proactive in implementing these trends.

7 Conclusion

In conclusion, this paper thoroughly explored Cyber Threat Analysis and Mitigation in the context of Emerging Information Technology Trends, encompassing cybersecurity aspects from cybercrime to emerging technologies like AI, Blockchain, IoT, and Cloud Computing. It adeptly navigated evolving cybersecurity challenges amidst new technological frontiers, shedding light on state-sponsored cyber attacks, espionage, and Advanced Cyber Threats. The findings underscore cybersecurity's criticality in the digital landscape for both enterprises and individuals, emphasizing robust security protocols, knowledge management, and business intelligence. Moreover, it extends to the broader cybersecurity discourse, emphasizing the evolving nature of the domain and the need for proactive vigilance. Equipped with insights into trends, ethics, and strategies, this paper guides enterprises and individuals in safeguarding their digital frontiers against cyber adversaries.

References

1. Corallo, A., Lazoi, M., Lezzi, M.: Cybersecurity in the context of industry 4.0: a structured classification of critical assets and business impacts. Comput. Ind. **114**, 103165 (2020)
2. Kim, J.: Cyber-security in government: reducing the risk. Comput. Fraud Secur. **7**(2017), 8–11 (2017)
3. Kumar, A., Bhushan, B., Nand, P.: Preventing and detecting intrusion of cyberattacks in smart grid by integrating blockchain. In: Sharma, D.K., Peng, S.-L., Sharma, R., Zaitsev, D.A. (eds.) Micro-Electronics and Telecommunication Engineering. LNNS, vol. 373, pp. 119–130. Springer, Singapore (2022). https://doi.org/10.1007/978-981-16-8721-1_12
4. Madaan, G., Bhushan, B., Kumar, R.: Blockchain-based cyberthreat mitigation systems for smart vehicles and industrial automation. Stud. Big Data Multimed. Technol. Internet Things Environ. 13–32 (2020). https://doi.org/10.1007/978-981-15-7965-3_2
5. Chen, P.: The evolution of cybercrime: from hackers to organized crime. J. Financ. Crime **21**(2), 312–322 (2014)
6. McAfee & CSIS. The economic impact of cybercrime—No slowdown in sight (2016)
7. Kumar, A., Bhushan, B., Malik, A., Kumar, R.: Protocols, solutions, and testbeds for cyber-attack prevention in industrial SCADA systems. In: Pattnaik, P.K., Kumar, R., Pal, S. (eds.) Internet of Things and Analytics for Agriculture, Volume 3. SBD, vol. 99, pp. 355–380. Springer, Singapore (2022). https://doi.org/10.1007/978-981-16-6210-2_17
8. Kashyap, S., Bhushan, B., Kumar, A., Nand, P.: Quantum blockchain approach for security enhancement in cyberworld. In: Kumar, R., Sharma, R., Pattnaik, P.K. (eds.) Multimedia Technologies in the Internet of Things Environment, Volume 3. SBD, vol. 108, pp. 1–22. Springer, Singapore (2022). https://doi.org/10.1007/978-981-19-0924-5_1
9. Singh, R., Kumar, R.: Cybercrime and its effects on individuals and organizations: a review of the literature. J. Cybersecur. **4**(2), 73–87 (2018)
10. Grier, C., Song, D., Tanz, J.: A first look at computer security dividends. In: ACM Conference on Computer and Communications Security, pp. 345–354 (2010)
11. Zhou, Y., Wang, X.: A survey of intrusion detection techniques. J. Netw. Comput. Appl.Netw. Comput. Appl. **35**(1), 1–14 (2012)
12. Chakraborty, S., Biswas, P.: A review on cyber security: threats, challenges, and solutions. J. King Saud Univ.-Comput. Inf. Sci. **30**(2), 131–139 (2018)
13. Anti-Phishing Working Group. Phishing Activity Trends Report Q3 2020 (2020). https://apwg.org/reports/apwg-phishing-activity-trends-report-q3-2020/
14. Jang-Jaccard, J., Nepal, S.: A survey of emerging threats in cybersecurity. J. Comput. Syst. Sci.Comput. Syst. Sci. **80**(5), 973–993 (2014)
15. Hu, Y., et al.: Artificial intelligence security: Threats and countermeasures. ACM Comput. Surv. (CSUR) **55**(1), 1–36 (2021)
16. Li, C., Wang, L., Ji, S., Zhang, X., Xi, Z., Guo, S., Wang, T.: Seeing is living? Rethinking the security of facial liveness verification in the DeepFake era. In: 31st USENIX Security Symposium (USENIX Security 2022), pp. 2673–2690 (2022)
17. Arora, A., Kaur, A., Bhushan, B., Saini, H.: Security concerns and future trends of internet of things. In: 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT) (2019). https://doi.org/10.1109/icicict46008.2019.8993222
18. Goel, A.K., Rose, A., Gaur, J., Bhushan, B.: Attacks, countermeasures and security paradigms in IoT. In: 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT) (2019). https://doi.org/10.1109/icicict46008.2019.8993338
19. Miao, Y., Chen, C., Pan, L., Han, Q.-L., Zhang, J., Xiang, Y.: Machine learning–based cyber attacks targeting on controlled information: a survey. ACM Comput. Surv. (CSUR) **54**(7), 1–36 (2021)

20. Choo, K.-K.R.: Cloud computing: challenges and future directions. Trends Issues Crime Crim. Just. **400**, 1–6 (2010)
21. Alshamrani, A., Myneni, S., Chowdhary, A., Huang, D.: A survey on advanced persistent threats: techniques, solutions, challenges, and research opportunities. IEEE Commun. Surv. Tutor. **21**(2), 1851–1877 (2019)

Author Index

F. P. García Márquez et al. (Eds.): ICETAI 2023, LNNS 960, pp. 589–591, 2024.
https://doi.org/10.1007/978-3-031-56728-5