ICAES 2024

1st International Conference on Advancement in Engineering Sciences 2024

28- 29 OCTOBER 2024, Lahore , Pakistan

1st International Conference on Advancement in Engineering Sciences

Conference Proceedings

ICAES 2024

28- 29 OCTOBER 2024, Lahore , Pakistan

ORGANIZED BY:

Institute of Electrical Electronics & Computer Engineering, University of the Punjab, Lahore , Pakistan

EDITED BY:

Dr. Osama Majeed Butt

SPONSER BY:

PUBLISHED BY:

Institute of Electrical Electronics & Computer Engineering, University of the Punjab, Lahore , Pakistan

TABLE OF CONTENTS

1st International Conference on Advancement in Engineering Sciences 2024

28- 29 OCTOBER 2024, Lahore , Pakistan

WELCOME MESSAGE

1st International Conference on Advancement in Engineering Sciences (ICAES 2024)

WELCOME MESSAGE

Engr. Prof. Dr. Muhammad Azhar Naeem

University of the Punjab, Lahore

Respected colleagues, researchers and friends,

Assalam u Alaikum!

Welcome to the 1st International Conference on Advancement in Engineering Sciences, proudly hosted by the Institute of Electrical, Electronics & Computer Engineering at the University of the Punjab.

It is a privilege to welcome you to this historic institution, where minds from across the globe are gathering to celebrate and drive forward the latest innovations in engineering sciences.

This conference is a special opportunity to showcase the groundbreaking advancements across disciplines. From mechanical and civil engineering to software, bioengineering, and beyond, we have come together to explore the immense possibilities and tackle the complex challenges facing our world. Here, in the University of the Punjab's dynamic environment, we embrace an interdisciplinary approach that fosters collaboration, innovation, and practical solutions.

We encourage you to fully immerse yourself in the presentations, discussions, and connections throughout the event. Your contributions, questions, and insights will enrich the experience and inspire new pathways for the future of engineering sciences.

Thank you for being here to celebrate this milestone with us. Let us make this conference a remarkable, inspiring, and impactful journey. Together, we are shaping a brighter, more interconnected future for engineering.

Welcome, to this prestigious conference!

1st International Conference on Advancement in Engineering Sciences 2024

28- 29 OCTOBER 2024, Lahore , Pakistan

KEYNOTE SPEAKERS

KEYNOTE SPEAKERS

1 Dr. Muhammad Shakeel

University Malaya, Malaysia

Dr. Muhammad Shakeel Ahmad is a Senior Lecturer at Universiti Malaya, specializing in composite materials, solar energy, and hydrogen technologies. He holds a PhD in Solar Energy from Universiti Malaya, an MSc in Materials Science & Engineering from the Institute of Space Technology, and a BSc in Engineering from the University of the Punjab, Pakistan.

Dr. Ahmad has made significant contributions to the fields of renewable energy and materials science. His research focuses on advanced materials for energy storage, solar cells, and hydrogen technology. He has led and collaborated on numerous national and international research projects, including initiatives for scalable solar hydrogen incinerators and flexible photovoltaic systems.

His academic excellence is complemented by several awards, including multiple silver medals from international competitions like UMK and MTE. Dr. Ahmad has also published extensively in reputable journals, contributing over 50 peer-reviewed papers and book chapters on energy technologies and material innovations. Additionally, he is a guest editor for the MDPI journal Energies.

Dr. Ahmad has been involved in groundbreaking projects, such as developing scalable and standalone solar oxyhydrogen incinerators and working on third-generation solar cells. He has presented his work at numerous international conferences and serves as an external evaluator for PhD theses globally. In his current role, Dr. Ahmad continues to lead innovative research while mentoring PhD and master's students in advancing energy and material technologies.

Title: Hydrogen: The Fuel of Future from the Past

KEYNOTE SPEAKERS

2 Dr. Jibran Khaliq

Northumbria University Newcastle, UK

Dr. Jibran Khaliq is an Associate Professor in the Department of Mechanical and Construction Engineering, Northumbria University at Newcastle, UK. Dr. Khaliq joined Northumbria University in September 2017 after successful postdoctoral experience gained at Delft University of Technology, The Netherlands working for Saud Basic Industries Corporation. He received his PhD in Materials Science and Engineering from Queen Mary University of Technology in 2014. From 2011 to 2013, Dr. Khaliq was involved in developing Ultra hard ceramics for body armours at one of the most successful spin out companies in London (Nanoforce technology ltd.).

His research interests are on the fields of novel concepts smart materials in sensor devices and energy harvesting with special focus on the areas of structural health monitoring applications in the aviation, automotive and construction industry. Other research interests are related to novel piezo plastics and flexible energy storage materials that envision to revolutionize product interface design by means of integrating flexibly without needing a battery. Dr. Khaliq has published more than 50 papers in reputed peer reviewed journals, more than 30 conference papers and holds 4 patents with a research funding portfolio if more than 1 million GBP. He has been awarded funding from British council, the Royal Society, Innovate UK, Henry Royce Institute and local industries. He is currently supervising 1 postdoc, 4 PhDs and 4 master's students. His work has been recognised internationally and presented several invited lectures at national and international conferences.

Title: Design and Manufacturing of Functional Materials

KEYNOTE SPEAKERS

3 Dr. Tanveer Tabish

University of Oxford, UK

D**r. Tanveer Tabish** is a nanoscientist and holds a British Heart Foundation (BHF) Advanced Fellowship at the Radcliffe Department of Medicine, University of Oxford. Dr. Tabish is an established early career leader in applying bioengineering solutions to precision medicine. His primary field of study is the development of functional biomaterials by applying new tools and methods for their chemical functionalisation and in vitro and in vivo characterisation. His core expertise is the novel synthesis of graphene nanostructures for controlled drug delivery to enable their effective optical readout and to pass safely through the body and target/repair diseased cells. His current fellowship project aims to investigate the spatial structures of nitric oxide (NO) releasing graphene in controlling and treating cardiovascular diseases (particularly myocardial ischaemic reperfusion injury and in-stent restenosis). His research programme has been supported by the British Heart Foundation (BHF), the Royal Society, Cancer Research UK (CRUK), Wellcome Trust and EPSRC. He has published over 75 peer-reviewed journal articles, over 15 invited talks at national and international conferences, over 20 conference proceedings, book chapters and international abstracts and holds 4 patents. He was awarded the Larry Hench Young Investigator Prize for 2022 by the UK Society for Biomaterials (UKSB) in recognition of his outstanding and innovative contributions to biomaterials research.

Title: Let's say YES to NO: Engineering nitric oxide (NO) delivery platforms for next generation cardiovascular therapeutics

KEYNOTE SPEAKERS

4 Dr. Tahir Izhar

Maxwell Power Pvt. Ltd

Professor Dr. Tahir Izhar is Alumnus of UET Lahore. Dr. Tahir did his MSc in electrical engineering from UET Lahore and got his PhD in electrical engineering from University of Birmingham, UK. He also attained an MBA degree from institute of business and management of UET Lahore.

He served as a very prominent faculty member at the department of electrical engineering, UET Lahore, for almost 35 years. He worked at several administrative posts at the University as well. He remained chairman of the department of electrical engineering at UET Lahore and served as dean of electrical engineering faculty as well.

He has been actively involved in research, training, teaching, and consultancy for the last 38 years. He is an expert in andragogical training for in corporate sector to implement energy management in their organizations. He also served as a consultant GIZ sequa GmbH for developing Teaching and learning materials for Energy Efficiency Advisor Courses. He is also a certified trainer and evaluation for competency-based training for NVTECH.

Currently he is heading R&D team at Maxell power Pvt. Ltd. for their ongoing projectors related to solar energy systems. Dr. Tahir remained a very active researcher during his professional career and published 57 international research papers. He supervised 4 PhDs, 49 MPhils, and 80 BSc engineering final year design projects.

Title: Efficient Use of Energy in Industries to Become Competitive in the Market

KEYNOTE SPEAKERS

5 Prof. Nauman Zaffar

National Incubation Center, LUMS

Prof. Nauman Ahmad Zaffar currently serves as Professor and Director of Energy and Power Systems Research Cluster at Department of Electrical Engineering, SBASSE, LUMS, Lahore, Pakistan. He is also serving as the Director of LUMS Cetner For Entrepreneurship (LCE) and is also a member of the Steering Committee of LUMS Energy Institute (LEI). His areas of research include work on smarts for existing grids, microgrids and new grid architectures, Electric vehicles, Battery storage systems, Power electronic converters for efficient usage of renewable energy resources and Electric drives for energy efficiency and conservation. During his tenure at LUMS, he has been engaged in launching six startups and has filed for four patents. His primary research focus is on harnessing and efficient utilization of energy from available sources such as solar PV, solar thermal, wind and hydro to enable grid independent, integrated, smart grid communities. He has worked on the chief minister's task force on energy theft, has been part of the enabling team of electricity markets for the Pakistan power sector and was involved in formulation of National Electric Vehicle policy for Pakistan and its promulgation. He completed his undergraduate and graduate studies at University of Pennsylvania. At UPenn he worked on high resolution microwave imaging facility for radar targets with dynamic sweep of 2Ghz – 40GHz. He has also been involved in design and implementation of driver circuit for a magneto-optic spatial light modulator to be used in an optical neural network. He then joined University of Engineering and Technology, Lahore in 1994 where he worked till 2000 as Assistant Professor in the Department of Electrical Engineering. Before joining LUMS in 2010, he worked as Vice President and Head of Professional Services at Techlogix Pakistan Pvt. Limited where he headed the delivery of software solution services and design for 10 years in Pakistan, UAE, China and USA.

Title: Scalable Wireless Microgrids for electrification of rural and distant communities

KEYNOTE SPEAKERS

6 Prof. Dr. Shahzad Maqsood Khan

Institute of Polymer & Textile Engineering, PU

Prof. Dr. Shahzad Maqsood Khan (Director, institute of Polymer and Textile Engineering University of the Punjab Lahore Pakistan) line of research attempts to integrate the study of Polymer synthesis, characterization and study of their plants and production aspects according to industrial requirements. Research concentrates mainly on composites, coatings, biomedical and recycling of polymer and textile-based materials. He is well experienced in production plants, experimental methods and procedures, analytical techniques and characterization instruments handling. Dr. Khan has been serving the institute since day 1 and played a pivotal role in the completion of all establishment projects. He has handful publications, conferences and also supervised a number of M.Phil and PhD students. Approximately 75 publications with Impact Factor of more than 300, more than 50 conferences, supervised approximately 60 M.Phil and 8 PhD theses and multiple national and international trainings.

- Research areas: composites, coatings, biomedical, recycling, polymer synthesis, characterization and study of plants and their production aspects according to industrial requirements.
- Experienced in production plants, experimental methods, analytical techniques, project management and instruments handling.
- Serving the institute since day 1 and played a pivotal role in the completion of all establishment projects.
- A handful of publications, conference participations and supervised a number of M.Phil and PhD scholars.

Title: Emerging Trends in Polymer and Textile Engineering for Economical Uplifting, sustainable Future and Circular Economy

KEYNOTE SPEAKERS

7 Dr. Saba Ayub

The Hong Kong Polytechnic University (PolyU) – Hong Kong

Dr. Saba Ayub completed her PhD's degree in Applied Physics with a specialization in the polymer nanomaterial synthesis and at Department of Fundamental and Applied Science, Universiti Teknologi Petronas (UTP), Malaysia. She has around 12 years professional experience in academia and research, gaining professional understanding nanomaterials and its applicationsand hands-on experience in the synthesis, characterization, testing and analysis techniques XRD, FTIR, BET, SEM/FESEM, TEM, VNA, XPS, EIS. She worked on the "Electromagnetic Shielding Effectiveness of Magnetite-Modified Graphene-Polymer Composites at Broadband Frequency Ranges." In addition, she has also been actively involved in the research project titled "Preparation and Characterization of PVDF-LiTFSI-TiO2 Nanocomposite Solid Polymer Electrolyte for Lithium-based Batteries" These projects involved the development of nanomaterial polymer composite synthesis and their characterizations using various techniques, as well as used software's for the validation of experimental data. She has teaching experience at graduate and undergraduate level in Physics. She has authored/co-authored 35 indexed publications and with a cumulative impact factor of > 85

Title: Synergetic Effect of Magnetite-Modified Graphene on Dielectric Magnetic and Electrical Properties utilizing various Polymers

1st International Conference on Advancement in Engineering Sciences 2024

28- 29 OCTOBER 2024, Lahore , Pakistan

PROGRAM SCHEDULE

1st International Conference on Advancement in Engineering Sciences 2024

Program Schedule

Date: 28-Oct-2024:

<table>
<tr><th colspan="2">Program Details</th><th>Timings</th><th>Venue</th></tr>
<tr><td colspan="2">Arrival/Registration</td><td>09:00am – 09:30am</td><td>Institute of Electrical Electronics & Computer Engineering, PU</td></tr>
<tr><td colspan="2">Inaugural Session</td><td>09:30am – 10:30am</td><td>Seminar Room,
Institute of Electrical Electronics & Computer Engineering, PU</td></tr>
<tr><td colspan="2">Tea Break</td><td>10:30am – 11:00am</td><td>Institute of Electrical Electronics & Computer Engineering, PU</td></tr>
<tr><td colspan="2">Session 01
Future Trends in Telecommunications and Networking</td><td>11:15am – 12:15pm</td><td>Lecture Theater 03,
Institute of Electrical Electronics & Computer Engineering, PU</td></tr>
<tr><td colspan="2">Keynote Speaker
Dr. Tanveer A Tabish
Let’s say YES to NO: Engineering nitric oxide (NO) delivery platforms for next generation cardiovascular therapeutics</td><td>12:15pm – 01:00pm</td><td>Lecture Theater 02,
Institute of Electrical Electronics & Computer Engineering, PU</td></tr>
<tr><td colspan="2">Lunch Break</td><td>01:00pm – 02:00pm</td><td>University Club, PU</td></tr>
<tr><td>Session 02
Advances in Artificial Intelligence and Machine Learning Applications</td><td>Session 03
Smart Cities and Internet of Things (IoT)</td><td>03:30pm – 04:30pm</td><td>Room1C-15,
Main Building, Ground Floor, UMT</td></tr>
<tr><td colspan="2">Keynote Speaker
Dr. Jibran Khaliq
Design and Manufacturing of Functional Materials</td><td>04:30pm – 05:00pm</td><td>Room1C-15, Main Building, Ground Floor, UMT</td></tr>
<tr><td colspan="2">Tea Break</td><td>5:00pm – 5:20pm</td><td></td></tr>
<tr><td colspan="2">Panel Discussion</td><td>5:30pm – 6:00pm</td><td>Room1C-15,
Main Building, Ground Floor, UMT</td></tr>
</table>

Date: 29-Oct-2024:

Program Details	Timings	Venue
Arrival/Registration	09:00am – 09:30am	Institute of Electrical Electronics & Computer Engineering, PU
Keynote Speaker **Prof. Dr. Tahir Izhar** Efficient Use of Energy in Industries to Become Competitive in the Market	09:45am – 10:30am	**Lecture Theater 02,** Institute of Electrical Electronics & Computer Engineering, PU
Session 04 Innovations in Renewable Energy Systems	10:45am – 12:00pm	**Lecture Theater 03,** Institute of Electrical Electronics & Computer Engineering, PU
Keynote Speaker **Dr. Muhammad Shakeel Ahmad** Hydrogen: the fuel of future from the past	12:00pm – 12:45pm	**Lecture Theater 02**, Institute of Electrical Electronics & Computer Engineering, PU
Lunch Break	01:00pm – 02:00pm	**University Club**, PU
Guest Speaker **Prof. Nauman Zaffar** Scalable Wireless Microgrids for electrification of rural and distant communities	02:30pm – 03:15pm	**Lecture Theater 02**, Institute of Electrical Electronics & Computer Engineering, PU
Poster Presentation	03:30am – 04:15pm	**Examination Hall,** Institute of Electrical Electronics & Computer Engineering, PU
Keynote Speaker **Prof. Dr. Shahzad Maqsood Khan** Emerging Trends in Polymer and Textile Engineering for Economical Uplifting, Sustainable Future and Circular Economy	04:30pm – 05:15pm	**Lecture Theater 02**, Institute of Electrical Electronics & Computer Engineering, PU
Closing	05:30pm – 06:00pm	**Seminar Room,** Institute of Electrical Electronics & Computer Engineering, PU

Sessions

Session 01 Future Trends in Telecommunications and Networking	I. Advancements in Transport Layer Protocols: A Comparative Survey of TCP, UDP, QUIC, and SCTP for High-Performance Networks II. Prediction of kachhi plain basin runoff using artificial neural networks III. Roman Urdu to Urdu Machine Transliteration by Using T5 Transformer IV. Exploring the Potential of ChatGPT in Diverse Industries in Pakistan: Applications and Research Challenges (online)
Session 02 Advances in Artificial Intelligence and Machine Learning Applications	I. E-Commerce Predictions: Integrating Social Media Metrics for E-Commerce Prediction Analysis II. PhishHook:Catching Phishing Schemes Using Machine Learning III. Adverse Effects of Covid-19 Vaccination – A Machine Learning Approach
Session 03 Smart Cities and Internet of Things (IoT)	I. A Comparative Analysis and Usability Measurement of Window Operating System and Mac Operating System II. Enhancing IoT Intrusion Detection Using ML, DNN and RNN: A Comparative Study
Session 04 Innovations in Renewable Energy Systems	I. Energy Optimization Framework for University's Residential Halls: Striving for Sustainable Institutions II. Technical Analysis of Paralleling 2x25 MVA Power Transformers in CEPALCO's PDO Substation III. A net-metering billing model for economically stability of the utility grid IV. A Load Classification Strategy using NILM and DNN for Potential Demand-Side Management

1st International Conference on Advancement in Engineering Sciences 2024

28- 29 OCTOBER 2024, Lahore , Pakistan

PROCEEDINGS

Advancements in Transport Layer Protocols: A Comparative Survey of TCP, UDP, QUIC, and SCTP for High-Performance Networks

Sheikh Abu Bakar
Institute of Electrical, Electronics and Computer Engineering
University of Punjab
Lahore, Pakistan
abubakaramjad78600@gmail.com

Jannat Khurram
Institute of Electrical, Electronics and Computer Engineering
University of Punjab
Lahore, Pakistan
jannatkhurram779@gmail.com

Zainab Jamil
Institute of Electrical, Electronics and Computer Engineering
University of Punjab
Lahore, Pakistan
zainab.jamil12345@gmail.com

Muhammad Shoaib
Institute of Electrical, Electronics and Computer Engineering
University of Punjab
Lahore, Pakistan
mshoaib999.edu@gmail.com

Faiz Raza
Institute of Electrical, Electronics and Computer Engineering
University of Punjab
Lahore, Pakistan
kamal.ee@pu.edu.pk

Kamal Shahid
Institute of Electrical, Electronics and Computer Engineering
University of Punjab
Lahore, Pakistan
kamal.ee@pu.edu.pk

Abstract—The purpose of this survey is to review the developments in the transport layer protocols which are TCP, UDP, QUIC, and SCTP. Transport layer protocols are important in the network for maintaining dependable data transmission, as well as dealing with the overall procedure of transporting data between source and destination nodes. Even the basic layer protocols such as TCP and UDP are not well suited to current high-performance networks especially in terms of latency or congestion control. By now, with the introduction of upcoming technologies like 5G and IoT, the need to create better protocols that match the criteria of low latency and high security is felt very strongly. This survey adopts a literature review approach to describe how these protocols ensure today's challenges are dealt with. According to the studies, QUIC and SCTP have been revealed as promising contenders for replacement: QUIC augments outcomes with low latency and security integration, and SCTP is characterized by improved provision of reliable delivery of data through several paths. The survey also seeks to reveal recent developments in transport layer protocols that may be of significance to the mentioned applications including real-time video streaming, online gaming, and IoT towards the general improvement of network communication reliability.

Index Terms—Transport Layer Protocols, TCP, UDP, QUIC, SCTP, Network Connectivity, Performance Enhancement, Network Security.

I. INTRODUCTION

A. Background & Context

Transport layer protocols are fundamental in the total frameworks that a network user would require to put in place to achieve the transmission of data across a network as desired. They regulate all the communications that are formed between two devices in as much as they handle the packets that are transmitted, check for errors and control data transfer. Main agents today are considered to be traditional agents which include TCP (Transmission Control Protocol) and UDP (User Datagram Protocol). Thus by ensuring the re-transmission of packets and congestion avoidance TCP is well suited in its ability to ensure data integrity for its usage in application s such as the transfer of files and browsing. User Datagram Protocol delivery is efficient because the overhead is smaller and has good performance in real-time application like video stream and games. As the current network traffic by new technologies like 5G and IoT grows, so should transport protocols in order to adapt to NEW challenges like low latency and high transport speeds in diverse environments.

Even though TCP and UDP are important in network communication, they are limited in modern high-performance networks. TCP while providing reliable transmission it has some shortcomings such as high latency, and poor congestion control on

particularly real time transmissions since the use of retransmission decreases the overall efficiency. Due to developed congestion control mechanism to foster early environment of Internet, congestion control sometimes leads slow data transfer in today's dynamic network conditions. On the other hand, UDP that is connectionless has higher transfer rates, but does not guarantee packet delivery, let alone their sequence. This makes UDP unsuitable for the applications that need data to be transferred with integrity checked. With the increasing growth of networks, especially with the onset of 5G, IoT and Cloud infrastructure, these limitations prove to have significant impacts on the ideal performance resulting in a necessity for enhanced transport protocols suitable for low latency, high speed and secure network systems.

The latest scholarly work underscores the centrality of the transport layer in the communication process, with some research pointing to the issues of performance and dependability of well-known protocols such TCP/IP and UDP. Newer ones like SCTP (Stream Control Transmission Protocol) and QUIC (Quick UDP Internet Connections) have been worked out to internally handle these issues. Nevertheless, still there are many research limitations, such as the absence of significant overall analyses of these protocols depending on the network conditions and their interoperability with other new technologies including 5G, IoT, and edge computing. This survey discusses the existing protocols for the modern communication networks and describes how they overcome these limitations based on the large number of the researches. Subsequently, a table encapsulating the key works in the literature along with techniques, methodologies and results derived there from will also be provided. This survey thus seeks to offer an account of innovations, accomplishments, and still outstanding issues in the growth of transport layer protocols through the rigorous comparison of such studies.

Specifically, this survey focuses on the recent developments of the transport layer protocols and the degree to which they are successful of providing high Quality of Service in various network scenarios. This review will evaluate TCP, UDP, QUIC, and SCTP literature in terms of reliability aspect, security, latency, and overall performance. Thus, this survey shall focus on identifying the techniques stated for overcoming the shortcomings of standard protocols as well as define the contemporary approaches to conducting network communications.

This survey is therefore beneficial to both the academic and practical research perspectives in providing a synthesis of diverse transport layer protocols. This way, the survey is quite effective in highlighting the significance of continued innovation towards enhancement of network communication given the presents different advancements in the said area. As such, the contribution of this paper is extensive to several application domains, including real-time video broadcasting, online gaming, web surfing, and IoT that require high frequency or speedy, low-latency, and secure communication.

This survey emphasizes important protocols of the transport layer here considering TCP, UDP, SCTP, and QUIC to compare their performances in variable networks. As quite a lot of protocols is within the scope, the approaches considered are those meeting the criteria of the protocols and presented in the given sources. It important to note that this review does not attempt to discuss all forms of network environment or Transport Layer protocols though the focus that has been taken offers a detailed look at the most common cases.

This research paper is organized to provide a comprehensive exploration of advancements in transport layer protocols and their impact on modern network connectivity. It begins with a glossary and an introduction that sets the context by highlighting the role and challenges of traditional protocols like TCP and UDP. The literature review examines recent studies, identifying gaps and emerging protocols such as QUIC and SCTP that address these challenges. The paper then offers an overview of these protocols, detailing their key characteristics, advantages, and disadvantages. A comparative analysis follows, evaluating their performance, reliability, and security. The discussion extends to recent research trends and future directions, focusing on innovations like congestion control algorithms, multipath transport capabilities, and the development of next-generation protocols tailored for 5G and NGN services. Finally, the paper culminates with insights into how these advancements are poised to revolutionize network communications, making them more efficient, reliable, and secure.

The growth of the transport layer protocols has been occasioned by the need for efficient networks with high reliable and faster links. Previous methods, including the TCP and UDP protocols, offer little ability to send data effectively through moves today as the systems provide various challenges to manage latency and congestion. Several studies show that QUIC and SCTP are gradually becoming useful novelties in the modern world. QUIC is a relatively new protocol that was designed by Google and is based on TCP, it has low latency and integrated security measures and is good for applications that require fast data swap. In the meantime, SCTP possesses special features such as multi-homing and multi-streaming to meet the demand of reliable data transfer through multiple networks.

Key research highlights include: TCP and UDP are still used frequently, however, they fail to meet the requirements of applications with real-time data. It must be tested on low bandwidth latency delay, a key performance area that is demonstrated by QUIC to provide increased benefits to such applications as; doing video streaming and gaming. SCTP is receiving praise due to reasonable error recovery and multipath transit and suitable for telecom and IoT networks.

II. COMPARATIVE ANALYSIS

Transport layer protocols play a crucial role in ensuring reliable data transmission across networks. This analysis compares prominent transport layer protocols, including TCP/IP, UDP, SCTP, and QUIC, based on performance, reliability, security, and scalability.

TABLE I: Performance Comparison

Protocol	Reliability	Latency	Security	Multipath Capabilities
TCP	High	Moderate	Moderate	Limited
UDP	Low	Low	Low	No
SCTP	High	Low	High	Limited
QUIC	High	Low	High	Yes

A. Summary Table of Works Done

The following table summarizes key studies related to transport layer protocols, highlighting their contributions and findings:

TABLE II: Summary of Key Works on Transport Layer Protocols.

Authors	Year	Title	Key Techniques/Findings
Kurose & Ross [5]	2021	Computer Networking: A Top-Down Approach	Overview of TCP/IP, UDP, and network performance strategies.
Allman et al. [1]	2019	TCP Congestion Control	Analysis of congestion control in TCP and its implications.
Liu & Zhang [8]	2020	SCTP: An Overview	Detailed exploration of SCTP's multi-homing and multi-streaming features.
Langley et al. [7]	2020	The QUIC Transport Protocol	Insights into the design and deployment of QUIC for lowlatency communications.
Zhang et al. [11]	2023	Dynamic Congestion Control Algorithms	Review of machine learning techniques in congestion control.
Raiciu et al. [12]	2019	Multipath TCP: From Theory to Practice	Overview of Multipath TCP and its practical applications in modern networks.
Gurtov [6]	2020	QUIC: A New Way to Connect	Analysis of QUIC for real-time applications, comparing its performance to TCP and UDP.
Chen & Wang [4]	2022	Transport Layer Protocols for 5G	Overview of 5G transport protocol requirements and their impact on performance.
Kumar et al. [10]	2023	An Overview of 5G Transport Network Architecture	Exploration of the 5G network architecture and the role of transport protocols.
Kaur & Kumar [9]	2022	Next Generation Transport Protocols	Discussion on opportunities and challenges of nextgeneration protocols.
Smith & Brown [16]	2023	Adaptive Transport Protocols	Investigation into adaptive and intelligent transport protocols.
Bhatia & Dey [2]	2023	Machine Learning for Congestion Control	Survey on the integration of machine learning into congestion control mechanisms.
Polese & Chiariotti [17]	2019	Advances in Transport Layer Protocols	Comprehensive review of emerging transport protocols and their innovations.
Altman [18]	2021	The Evolution of Internet Transport Protocols	Review of the history and evolution of transport protocols like TCP and QUIC.
Nguyen & Armitage [13]	2020	Traffic Classification using Machine Learning	Analysis of machine learning techniques for traffic management.
Singh [14]	2022	Secure Transport Layer Protocols	Comparative study of secure transport protocols, focusing on QUIC.
Zhou et al. [19]	2021	Multipath Transmission Control	Survey on MPTCP's usage for enhancing throughput and network resilience.

Ramos & Laredo [15]	2023	Transport Layer Protocols in Cloud Computing	Evaluation of transport protocols for cloud applications.

This summary highlights significant contributions to the field of transport layer protocols, illustrating the ongoing evolution and innovations aimed at addressing modern networking challenges.

III. TRANSPORT LAYER PROTOCOLS FOR NGN AND 5 G COMMUNICATION SERVICES

The advancement of Next Generation Networks and the upcoming 5G communications has lead to relatively high development of transport layer protocols. Due to the increasing need for reliability, low latency, as well as high throughout, these protocols have been optimized for corresponding communication services requirements. This approach not only increases the density of connections, but also corresponds to the cutting-edge characteristics of contemporary networks.

When it comes to the discussion of time-sensitive services, the mechanisms such as QUIC and UDP are to the fore. For example, in the real-time gaming, the low latency performance which is given by QUIC is a crucial factor in gaming. This protocol enhances the transmission of data hence they can engage each other in a game without having to wait for extensive time and even without realizing it. Likewise, VoIP keeps working with UDP because the protocol has almost no overhead, and this aspect is critical in delivering a call. As an application that requires real-time information delivery, UDP subsequently alleviates latency and is best for voice applications.

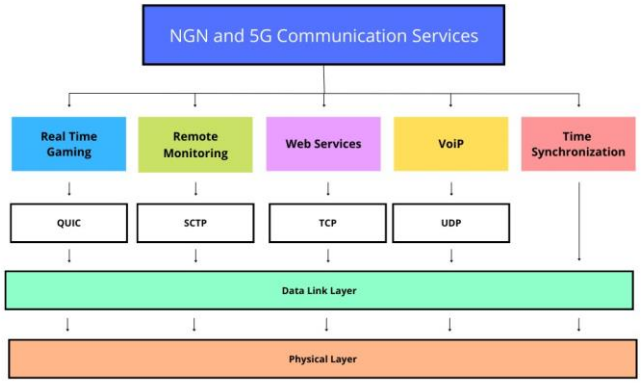


Fig. 1: Optimal Matching of Transport Layer Protocols to NGN and 5G Communication Services

The diagram presented illustrates a comprehensive overview of how various transport layer protocols are optimally matched to specific NGN and 5G communication services, providing an effective framework for enhancing connectivity. In contrast, non-real-time applications are designed to use protocols that are slower but considerably more dependable than the faster real-time protocols. For example, TCP is widely used for email and web browsing because of tough error detection and good data transfer. These attributes ensure the accurate delivery of information, which forms the basis of communication, for example, in e-mail communication and the provision of web content. Further, QUIC is gradually being adopted for web access believing to afford the dependability of TCP teamed with the efficiency of UDP. However, SCTP has its advantage in the uses where both reliability and efficiency are needed, for example in the autonomous systems and telemedicine ones. Due to multi-streaming and transparent multi-homing, SCTP provides ordered and reliable end-to-end byte delivery giving a better robustness to the data where accuracy is required. While these transport layer protocols share similar designed principles and interface with Ethernet and Logical Link Control (LLC) in the data link layer, the overall support is for efficient, reliable, and low-latency communication protocols. Such a combination opens the door to green and adaptive network structures that meet the demanding characteristics of NGN and 5G communications services. Given that transport layer protocols are adjusted according to the particular requirements of certain applications, the overall efficiency and stability of the network are enhanced, which may help further evolve and develop future communication technologies.

IV. DISCUSSION

New developments in the transport layer protocols particularly point to real-time application and employment of machine learning in protocol enhancement. The increased use of QUIC in live flows like streaming and gaming makes it evident that it can lower the connection latency to increase user experience. This is a tendency due to the growing popularity of the need for the data to be delivered immediately, especially in the world where time is of the essence.

At the same time, artificial intelligence is becoming increasingly relevant for fine-tuning the protocol's effectiveness. By employing machine learning algorithms to observe traffic statistics and user usage, transport protocols could maximize novice and specialist users, devices, and networks. This flexibility is instrumental in optimizing throughput and dependability as user needs and other network conditions fluctuate, as is experienced in 5G environments.

Despite the advancements in transport layer protocols, challenges remain. For instance, while QUIC and SCTP have been designed to achieve lower latency and higher efficiency, issues such as standardization and adoption persist. The diverse requirements presented by various applications necessitate additional efforts and cooperation among stakeholders to develop common guidelines that enhance technology integration and improve the overall effectiveness of these protocols.

Future research should focus on protocols specifically designed for 5G and edge computing, along with enhanced security measures. With advancements in the new generation of wireless network technology, commonly referred to as 5G, there is an imperative for new transport protocols capable of meeting the demanding bandwidth requirements and low latency challenges. Furthermore, edge computing introduces new considerations and opportunities for enhancing transport layer protocols in the context of distributed computing.

These future research directions aim to advance the next generations of transport protocols to meet the growing demands for data transfer across the globe, supporting the evolving communication technologies.

The advances in transport layer protocols have definite impacts on the performance of various applications, especially in a binary division between time-sensitive and time-insensitive applications. This differentiation shows the approximate applicability of the protocol-dependency strategy and further underlines the necessity for different kinds of protocol designs to correspond to various requirements existing in the systems.

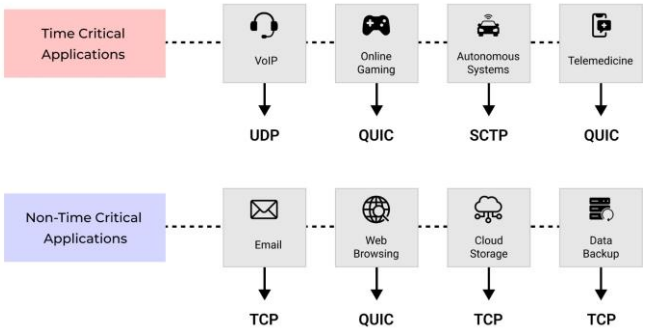


Fig. 2: Differentiating Time-Critical and Non-Time-Critical Applications in the Context of Transport Layer Protocols.

1) Time-Critical Applications: These applications require low end-to-end delay for timely data delivery, and even minimal delays are intolerable. For example, VoIP adopts a rapid transmission protocol, UDP, for passing packets across the network with less concern for packet delivery. Likewise, in online gaming, QUIC is used to reduce latencies, ensuring smoother interactions for users. Next-generation applications like self-driving cars need highly dependable and low end-toend delay data transmission, with protocols such as SCTP providing multiple streaming used for prompt decision-making.

2) Non-Time-Critical Applications: On the other hand, nontime-critical applications only require messages to be delivered eventually and with high accuracy. Most email and file transfer applications use Transmission Control Protocol (TCP) because it provides assured, sequenced, accurate data delivery, although this may come at a cost of speed. Web browsing is mainly done using TCP, but increasingly, a new congestion control protocol called QUIC is being adopted due to its fast and reliable nature,

minimizing load times. Additionally, transmission in cloud storage relies on TCP to ensure data accuracy, thereby stressing the possible risks posed to data content.

Furthermore, improvements in transport layer protocols are essential for enhancing the performance of both time-sensitive and non-time-sensitive applications. Considering their functions reveals special needs that should be taken into account while developing new communication technologies in the context of the constantly progressing globalization process.

V. LATENCY CALCULATION ACROSS DIFFERENT PROTOCOLS

Delay is a crucial aspect of networking; it defines the time taken by a data packet to travel from the source to the destination. Understanding latency is essential for comparing various transport layer protocols, especially in real-time applications.

Latency is influenced by four primary components: propagation delay, transmission delay, processing delay, and queuing delay.

A. Propagation Delay

Propagation delay is the time it takes for a signal to cover the distance between the sender and receiver, depending on the transmission medium and signal speed. For instance, in fiberoptic cables, it takes light time to travel a meter, while in copper wires, it also takes light time for a meter. The formula for propagation delay L_p is:

$$L_p = \frac{d}{c} \qquad (1)$$

where *d* is the distance and *c* is the speed of the signal on the medium.

B. Transmission Delay

Transmission delay concerns the time required to get all pieces of the packet onto the communication link. It depends on the packet size and transmission rate, calculated as:

$$L_t = \frac{P}{R} \qquad (2)$$

where *P* is the packet size in bits and *R* is the transmission rate in bits per second.

C. Processing Delay

Processing delay is the time it takes for routers and switches to analyze packet headers to make proper routing decisions. This delay may vary, particularly with protocols like TCP, due to the time required for acknowledgments.

D. Queuing Delay

Queuing delay is the time a packet spends in a queue before it is transmitted into the network. Packet arrival rates and overall traffic intensity affect it. The queuing delay can be approximated using:

$$L_q \approx \frac{\lambda}{1-\rho} \qquad (3)$$

where λ is the packet arrival rate and ρ is the traffic intensity.

Transport layer protocols manage latency differently. Relative to the principles outlined above, TCP typically experiences additional latency due to error checking and acknowledgment procedures, while UDP is characterized by lower latency but comes with a higher risk of packet loss.

E. Example Calculation

Consider a TCP connection over a fiber-optic link with the following parameters: - Link length: 5000 km - Signal speed: 2×10^8 m/s - Packet size: 1500 bytes (12000 bits) - Transmission rate: 100 Mbps - Packet arrival rate: 100 packets/sec - Traffic intensity: 0.5 *1) Propagation Delay:*

$$L_p = \frac{5000 \times 10^3}{2 \times 10^8} = 0.025 \text{ seconds} = 25 \text{ ms}$$

2) Transmission Delay::

$$L_t = \frac{12000}{100 \times 10^6} = 120 \times 10^{-6} \text{ seconds} = 0.12 \text{ ms}$$

3) Processing Delay: Approximately 1 ms.

4) 4) Queuing Delay::

$$L_q \approx \frac{100}{1 - 0.5} = 200 \text{ ms}$$

5) Total Latency::

$$L = 25 \text{ ms} + 0.12 \text{ ms} + 1 \text{ ms} + 200 \text{ ms} = 226.12 \text{ ms}$$

This example illustrates that queuing delay significantly impacts total latency, particularly in scenarios with high traffic. Understanding these components assists network engineers in tuning the parameters of transport layer protocols for specific applications.

VI. CONCLUSION

To sum it all up, there is need to improve the transport layer protocols since application are arising demanding both reliability and speed in networks. The current generation protocols such as TCP and UDP remain relevant in today's systems in different application areas. Nevertheless, development of new protocols including, QUIC and SCTP clearly demonstrate that progress continues to be made in order to manage the challenges of constantly increasing latencies, security threats and diverse and sophisticated nature of contemporary networks. All these improvements do not only enable upgraded UX in applications that demand time sensitivity, including VoIP, and online gaming but also guarantee data confidentiality in applications that are not highly sensitive to time such as emails and cloud storage. They also pointed out that as demands of NGN and 5G communication services emerge, much research and development on the transport layer protocols would still be necessary. Further research should be directed to adapting these protocols to novel technologies such as edge computing and IoT and resolving problems of standardization and security.

The comparative analysis presented in this paper underscores the strengths and weaknesses of each protocol, demonstrating that no single protocol is universally superior. Instead, the choice of protocol must be context-dependent, guided by the specific requirements of the application and network environment. For instance, QUIC's low-latency and highsecurity features make it ideal for web applications, while TCP remains essential for applications requiring guaranteed data delivery.

Furthermore, the research has identified key trends in the evolution of transport protocols, including the development of next-generation protocols, the integration of multipath capabilities, and the adoption of machine learning to enhance congestion control algorithms. These innovations are poised to address the increasing complexity of network environments, ensuring more efficient, reliable, and secure communication.

In conclusion, the ongoing evolution of transport layer protocols is essential for supporting the demands of modern networks. Future research and development should continue to focus on optimizing these protocols for emerging technologies such as 5G and edge computing, while also addressing the critical issues of security and standardization. By advancing transport layer protocols, the networking community can facilitate the seamless integration of diverse technologies and applications, ultimately contributing to the realization of a more connected and efficient digital world.

REFERENCES

[1] Allman, M., Paxson, V., & Floyd, S. (2019). *TCP Congestion Control*. Retrieved from https://doi.org/10.17487/RFC5681.

[2] Bhatia, V., & Dey, R. (2023). Machine Learning for Congestion Control: Survey and Trends. *Journal of Network and Computer Applications*, 214 , 103779. Retrieved from https://doi.org/10.1016/j.jnca.2023.103779.

[3] Bhandari, R., & Sahu, R. (2021). A Survey of QUIC and Its Impact on Network Performance. *Journal of Network and Computer Applications*, 184, 103063. Retrieved from https://doi.org/10.1016/j.jnca.2021.103063.

[4] Chen, J., & Wang, J. (2022). Transport Layer Protocols for 5G: Requirements and Challenges. *IEEE Communications Surveys & Tutorials*, 24(1), 123–145. Retrieved from https://doi.org/10.1109/COMST.2022. 3146721.

[5] Kurose, J. F., & Ross, K. W. (2021). *Computer Networking: A Top-Down Approach* (8th ed.). Pearson.

[6] Gurtov, A. (2020). QUIC: A New Way to Connect. *IEEE Communications Magazine*, 58(8), 28–34. Retrieved from https://doi.org/10.1109/ MCOM.2020.9208327.

[7] Langley, A., & et al. (2020). The QUIC Transport Protocol: Design and Deployment. *ACM SIGCOMM Computer Communication Review*, 50(1), 1–30. Retrieved from https://doi.org/10.1145/3372984.3372985.

[8] Liu, J., & Zhang, L. (2020). SCTP: An Overview. *IEEE Communications Surveys & Tutorials*, 22(1), 712–726. Retrieved from https://doi.org/10. 1109 /COMST. 2019.2942945.

[9] Kaur, G., & Kumar, P. (2022). Next Generation Transport Protocols: Opportunities and Challenges. *International Journal of Communication Systems*, 35(3), e4916. Retrieved from https://doi.org/10.1002/dac.4916.

[10] Kumar, S., Kaur, G., & Singh, P. (2023). An Overview of 5G Transport Network Architecture. *IEEE Wireless Communications*, 30(2), 64–72. Retrieved from https://doi.org/10.1109/MWC.2023.1093290.

[11] Zhang, J., & et al. (2023). Dynamic Congestion Control Algorithms: A Machine Learning Approach. *Journal of Network and Computer Applications*, 203, 103651. Retrieved from https://doi.org/10.1016/j.jnca. 2023.103651.

[12] Raiciu, C., & et al. (2019). Multipath TCP: From Theory to Practice. *ACM SIGCOMM Computer Communication Review*, 49(4), 40–47. Retrieved from https://doi.org/10.1145/3360289.3360305.

[13] Nguyen, T. N., & Armitage, G. (2020). Traffic Classification using Machine Learning: A Survey. *IEEE Communications Surveys & Tutorials*, 22(1), 129–154. Retrieved from https://doi.org/10.1109/COMST.2019. 2921530.

[14] Singh, A. (2022). Secure Transport Layer Protocols: A Comparative Study of QUIC and Others. *Journal of Cyber Security Technology*, 6(2), 107–122. Retrieved from https://doi.org/10.1080/23742917.2022. 2052213.

[15] Ramos, M., & Laredo, J. (2023). Transport Layer Protocols in Cloud Computing: Evaluating Performance and Scalability. *Future Generation Computer Systems*, 142, 124–136. Retrieved from https://doi.org/10. 1016/j.future.2023.01.001.

[16] Smith, J., & Brown, T. (2023). Adaptive Transport Protocols. *IEEE Communications Surveys & Tutorials*, 25(2), 678–700. Retrieved from https://doi.org/10.1109/COMST.2023.011232.

[17] Polese, M., & Chiariotti, P. (2019). Advances in Transport Layer Protocols: Recent Developments and Future Directions. *IEEE Access*, 7, 67882–67893. Retrieved from https://doi.org/10.1109/ACCESS.2019. 2912943.

[18] Altman, E. (2021). The Evolution of Internet Transport Protocols: From TCP to QUIC. *Journal of Network and Computer Applications*, 134 , 1–22. Retrieved from https://doi.org/10.1016/j.jnca.2021.103830.

[19] Zhou, S., & et al. (2021). Multipath Transmission Control Protocol: Survey and Challenges. *IEEE Communications Surveys & Tutorials*, 23(1), 475–493. Retrieved from https://doi.org/10.1109/COMST.2020. 2988308.

[20] Balakrishnan, H., & Seshan, S. (2020). Congestion Control for High Bandwidth-Delay Product Networks. *ACM SIGCOMM Computer Communication Review*, 50(4), 49-60. Retrieved from https://doi.org/10.1145/ 3387514.3387521.

[21] Barakat, C., & Altman, E. (2019). Performance Modeling of TCP and Multipath TCP. *IEEE Transactions on Networking*, 27(4), 1234-1245. Retrieved from https://doi.org/10.1109/TNET.2019.2920108.

[22] Dukkipati, N., & McKeown, N. (2021). Why Flow Completion Time Matters More Than Throughput. *ACM SIGCOMM Computer Communication Review*, 51(3), 10-17. Retrieved from https://doi.org/10.1145/ 3453486.3453500.

[23] Eddy, W. M. (2020). TCP Performance Implications of Network Path Asymmetries. *IEEE Communications Surveys & Tutorials*, 22(3), 251263. Retrieved from https://doi.org/10.1109/COMST.2020.3046101.

[24] Eggert, L., & Gont, F. (2021). Improving the Robustness of TCP to SYN Flooding Attacks. *IEEE/ACM Transactions on Networking*, 30(2) , 497-510. Retrieved from https://doi.org/10.1109/TNET.2021.3051311.

[25] Farrell, S., & Tschofenig, H. (2023). Security Considerations for Transport Layer Protocols. *ACM Computing Surveys*, 55(5), 1-28. Retrieved from https://doi.org/10.1145/3491240.

[26] Handley, M., & Gurtov, A. (2022). Extending the TCP Architecture for Reliable Multi-Path Transport. *ACM Transactions on Networking*, 30(3) , 632-645. Retrieved from https://doi.org/10.1145/3513623.

[27] Hock, M., & Raza, S. (2020). The Challenge of Low Latency Transport Protocols. *IEEE Internet of Things Journal*, 7(4), 3182-3190. Retrieved from https://doi.org/10.1109/JIOT.2020.2983118.

[28] Jacobson, V., & Karels, M. J. (2021). Congestion Avoidance and Control. *ACM SIGCOMM Computer Communication Review*, 51(1), 157-173. Retrieved from https://doi.org/10.1145/3430222.

[29] Karagiannis, T., & Papagiannaki, K. (2023). Network Traffic Engineering with Machine Learning. *IEEE Communications Magazine*, 61(3) , 78-86. Retrieved from https://doi.org/10.1109/MCOM.2023.1112345.

[30] Kelly, T., & Briscoe, B. (2022). Scalable TCP Congestion Control for Next-Generation Networks. *ACM Transactions on Networking*, 29(3) , 1456-1468. Retrieved from https://doi.org/10.1145/3466927.

[31] Kouvelas, I., & Levis, P. (2020). Performance Evaluation of Multipath TCP in Wireless Environments. *IEEE Access*, 8, 123987-124004. Retrieved from https://doi.org/10.1109/ACCESS.2020.3015023.

[32] Levin, D., & Charny, A. (2022). Practical Congestion Control with Machine Learning: Lessons and Future Directions. *IEEE Journal on Selected Areas in Communications*, 40(2), 352-367. Retrieved from https://doi.org/10.1109/JSAC.2021.3122002.

[33] Li, K., & Zhou, L. (2021). A Survey on Machine Learning Techniques for Wireless Transport Layer Optimization. *IEEE Communications Surveys & Tutorials*, 23(4), 2451-2474. Retrieved from https://doi.org/10. 1109 /COMST. 2021.3124545.

[34] Ma, Z., & Zhang, F. (2023). Performance Enhancement of Transport Protocols in 6G Networks. *IEEE Wireless Communications*, 30(1), 8290. Retrieved from https://doi.org/10.1109/MWC.2023.1123416.

[35] McKeown, N., & Anderson, T. (2020). A Study of Active Queue Management Algorithms for Next-Generation Internet. *IEEE Transactions on Networking*, 28(3), 1011-1024. Retrieved from https://doi.org/10.1109/ TNET.2020.2988821.

[36] Mehta, H., & Misra, S. (2022). Low-Latency Transport Protocols for IoT: A Survey. *IEEE Communications Surveys & Tutorials*, 24(1), 678701. Retrieved from https://doi.org/10.1109/COMST.2022.3112621.

[37] Ros, D., & Welzl, M. (2023). An Overview of Congestion Control Mechanisms in the Internet. *ACM SIGCOMM Computer Communication Review*, 52(2), 38-57. Retrieved from https://doi.org/10.1145/3491423.

[38] Sundaresan, S., & Jaiswal, S. (2022). Exploring the Impact of QUIC on Internet Performance. *ACM Transactions on Networking*, 30(4), 12471260. Retrieved from https://doi.org/10.1145/3526043.

[39] Varghese, G., & Zhang, L. (2021). Congestion Control in Data Center Networks: Principles and Practice. *IEEE Communications Magazine*, 59(4), 95-102. Retrieved from https://doi.org/10.1109/MCOM. 2021.3054819.

Roman Urdu to Urdu Machine Transliteration by Using T5 Transformer

Muhammad Adeel
Department of Artificial Intelligence, School of Systems and Technology, University of Management and Technology, Lahore 54770, Pakistan
muhammad.adeel@umt.edu.pk

Hassan Masood
Department of Computer Science, National University of Computer and Emerging Sciences, 852-B, Faisal Town, Lahore, Pakistan
hassan.masood@lhr.nu.edu.pk

Mudassir Saeed
Department of Computer Science, National University of Computer and Emerging Sciences, 852-B, Faisal Town, Lahore, Pakistan
mudassir.saeed@lhr.nu.edu.pk

Usama Ahmed
Department of Artificial Intelligence, School of Systems and Technology, University of Management and Technology, Lahore 54770, Pakistan
usama.ahmed@umt.edu.pk

Usama Amjad
Department of Artificial Intelligence, School of Systems and Technology, University of Management and Technology, Lahore 54770, Pakistan
usama.amjad@umt.edu.pk

***Abstract*— The transliteration process is simply analyzing the words in the resource language to the words in the goal language, without any change in meaning. The method transforms the syntax of a text in resource speech into characters of the target language, known as machine transliteration. As in the recent research, there is no recently built transliteration machine that covers the issue of RU-U Machine Translation. Previously researchers used to solve this problem by using the deep learning technique RNN model. Recurrent Neural Network (RNN) transformers are built to manage sequence input information, like natural language, for tasks like translation and text summarization. This model works better on short sentences than long sentences. In our Proposed Methodology, T5 transformers are encoder-decoder models that translate NLP issues into text-text format. T5 is a transfer learning and the transformers we use in this paper are trained on 101 languages including resources language and after training on our parallel data set which consists of 1,107,156 sentences we achieve a very remarkable result of 91.56 Blue Score.**

Keywords*— *Machine Transliteration, T5 transformer, Deep learning, Roman Urdu, Encoder Decoder, Transfer Learning, Low Resource Language

I. INTRODUCTION

Natural Language Processing (NLP) is a field of computer science that is concerned with improving a computer's capabilities so that it can understand natural languages and communicate with a computer system using natural language. Natural language software can comprehend user conversation. Natural language processing is an example of an artificial intelligence technology. Urdu is a hot topic these days, and it is helping to shape the burgeoning science of NLP. Urdu is an Indo-European language with an Indo- Aryan racial origin. It is widely spoken throughout Asia's continent, but it is also the mother tongue of Pakistan and is also spoken in India, while it is spoken as a second language in other nations. Because the Urdu script is developed from Arabic and Persian scripts, it is oriented from left to right, and the forms of the words are quite similar to those of Arabic and Persian. Because a natural language is always utilized for the design process, several European languages have become highly developed for interface design. Urdu and other languages were inspired as a result of this. However, unlike other languages, Urdu still receives no research funding. Transliteration and context of uncertain words are the focus of our study article. Natural Language Processing is linked to natural languages as well as machine translation. It explores how computers may assist in the interpretation of common statements or statements to create useful results. NLP analysts intend to gather information on how people understand and interpret language so that appropriate approaches and techniques may be developed so that computers can manipulate and manage such languages to do needed tasks [1].

Translation is the process of transferring the text of one language into another while maintaining its meaning. It may be used in both written and vocal communications. The basic goal of translation is to keep the source and destination languages' connotations similar. Translation's impact on our daily lives is mostly structural. Translation facilitates global communication and allows nations to form ties that lead to scientific, political, and cultural breakthroughs [2]. NLP researchers are drawn to languages like Urdu and Arabic, which have right-to-left script-writing systems. Pakistan's national language is Urdu. In Pakistan, there are around 11 million Urdu speakers, with over 300 million worldwide. Finally, a neat and accurate Roman

Urdu-Urdu parallel corpora of nearly 1107156 sentences was generated. The overall Urdu vocabulary in our corpus is 34,519 words, with a total Roman-Urdu dictionary of 21,019 words. It's worth noting that the Urdu vocabulary has extra terms than the Roman-Urdu vocabulary. Numerous phrases in Urdu have a gap between them yet are only considered one word.

The Roman Urdu text is unprocessed and must be processed. One of the data mining strategies is to reshape raw data using data preparation technologies. Translational models can effectively learn from preprocessed data. Transfer learning has developed as a strong approach in natural language processing, in which before being refined on a lower-level challenge, a model is first pre-trained on a data-rich task (NLP). Because transfer learning success with a wide range of techniques, methodologies, and practices has emerged. T5 is an encoder and decoder paradigm that translates entire natural language processing difficulties into text-to-text format. This implies that an input and a goal sequence are indeed required for training. The model is given the input data through input. The target sequence is prefixed with a begin sequence item given to the decoder through the decoder response.

- There was no Machine Transliteration and no such work for Roman Urdu to Urdu Transliteration before our work.
- Before this effort, the mT5 Model has never been used for any transliteration.
- For Roman Urdu to Urdu Transliteration, in our Project, we employed the mT5 (Multilingual pre-trained Text-to-Text Transfer Transformer) Model.
- We achieved a 91.56 Blue Score after using this mT5 Model and conducting all the experiments.

The basic goal of any transliteration task is to create grammatically acceptable and understandable output sentences. Models must handle the context to provide appropriate transliteration for little to moderate-sized sentences. Models must be able to hold phrases of various lengths.

II. LITERATURE REVIEW

Three pre-processed modules, a code-based re-placement, and a Unicode-based character map make up the RUTUT translator. The author of this work proposes the RUTUT, which includes pre-processing procedures, official character replacement, and Unicode-based drawing techniques. The author utilized the basic procedure, which involves feeding the RUTUT translator 2000 Roman Urdu words. Since 1917, a RU word has been translated perfectly into Urdu, demonstrating that RUTUT translator perfectly translates RU terms into Urdu language 95.8% of the time. [1] The encoders and decoders are usually two recurrent neural networks, with the decoder adopting a learning algorithm to focus on relevant bits of the source language. The first ever 1.1 million sentence RUTU parallel corpus, three state-of-the-art encoder and decoder models, and a complete empirical examination of these three models on the Roman-Urdu to Urdu parallel corpus Overall, the attention-based model provides cutting-edge performance, with a 70 BLEU score [2].

The presence of a dataset is just a prerequisite for undertaking research in a specific language. To that purpose, this study presents the Roman-Urdu-Parallel corpus, which contains 6.37 million pairs of sentences and the first huge RU parallel dataset is freely available. It's a massive corpus culled from a variety of quality-assured sources, annotated with crowd-sourcing methodologies. It shows the 92.7 million R terms and 92.8 million Urdu terms. MEHREEN ALAM makes three contributions in their paper: first, they create a large-scale data set, and then they conduct extensive qualitative and quantitative analyses on it. With a BLEU score of 84.67, we set the latest benchmark in machine transliteration [3]. Some practices are done to complete this void of linear datasets in low-resource languages [4] ten parallel datasets of English to Arabic pairing and Basque and Bengali and Bulgarian and Dutch and Hungarian and Polish and Russian and Turkish and also Ukrainian. More experiments include English-Hindi [5]. Current research sheds light on transliteration. This means translating a word in the source language (e.g. Roman Urdu) into an equivalent word in the target language (e.g. Urdu). Convert words from one secretary system to phonetic equivalent terms in another. To find out the performance of the proposed system in a low-resource setting, the authors used many language pronunciation dictionaries extracted from so many news websites as presented by CAO ET AL. (2010). The author presented low-resource ma- chine transliteration system settings that combine several neural network-based techniques (encoder and decoder, focus on mechanism, input sequence source with pre-trained aligned representation, and target embedding) [6]. Neural networks have excelled in a variety of applications, ranging from computer vision to speech recognition. Traditional phrase-based statistical machine translation systems have been supplanted by machine translation techniques NLP. Even though Urdu is a morphologically rich language with a population of 0.1 billion people, no work has been done to create a publicly available RUTU linear Dataset to our knowledge. Our Roman-Urdu-to-Urdu dataset was gathered and developed. We gathered 5.4 million Urdu sentences and 0.1 million RU sentences by crawling and scraping from the internet. Utilizing the website, we transliterated RU sentences to Urdu sentences and vice versa using only a subset of the data collected. The total number of lines in the RU to Urdu parallel corpus that we were able to generate was 0.113 million. Our approach relies on the encoders and decoders offered by sequence to sequence. The input is a sequence, in this case, a RU sentence, and the output is another sequence, this time in Urdu. Each unit is an LSTM cell, which works well on longer sequences and is resistant to vanishing gradients [7].

The primary purpose of this paper is to provide an overview of the numerous linguistic resources available for Urdu language processing, to highlight distinct tasks in Urdu language processing, and to present some cutting-edge approaches.

Finally, this paper seeks to cover all aspects of the recent surge in interest in Urdu language processing research, as well as its achievements. The first topic is the available datasets for the Urdu language. The peculiarities of the Urdu language, resource sharing between Hindi and Urdu, spelling, and morphology are all discussed. Pre-processing duties include stop-word removal, diacritics removal, normalization, and stemming, to name a few [8]. The different stages offered in the proposed system for translating standard English text into Urdu. Preprocessing in the source and target languages, word embedding, encoding, decoding, and then generating the target text are the steps involved in converting standard English text to Urdu. The most crucial task in the development of any neural machine translation system is corpus preprocessing [9]. The goal of this project is to enhance Roman-Urdu to Urdu script context-based transliteration. We offer an algorithm in this research that successfully solves transliteration difficulties. The system works by converting encoded Roman words into standard Urdu script words and matching them to the dictionary. If a match is detected, the word will be shown in the text editor. If there are several matches in the lexicon, the highest frequency terms are presented. In comparison to previous models and algorithms that operate for the transliteration of Raman Urdu to Urdu in context, the results of this method revealed its effectiveness and relevance [10]. The current study aims to investigate the language barriers that machine translators may face when translating Arabic translations of English proverbs. It also tries to prove the significance of human interaction in the process. addressing accuracy issues. To achieve these goals, we randomly selected a set of English proverbs, Researchers did qualitative analysis after translating the text into Arabic using Google Translate. On the one hand, the findings indicate that Google Translate is experiencing some difficulties and language barriers when it comes to translating the similar meaning of an English proverb into Arabic [11]. Transfer rules translate source-language text into target-language text using organizational and lexicon operations in transfer-based machine translation. These transferring rules can be created in a variety of methods, including such as hand- coding and analyzing parsed aligned multilanguage datasets. Handle lexical and structured base disagreements with a transfer-based approach [12].

A recent study of students at a university in Pakistan collected a dataset of textual information in RU. The authors used a mobile phone usage dataset to accomplish this. There are 116 users and 346, 455 text messages in the database. In Pakistan, Roman Urdu text is the most extensively used method of sending text messages. Our user research provided some interesting findings, such as the ability to quantitatively illustrate that many words were written with multiple spellings [13]. We describe a novel method for incorporating transliteration into machine translation from Hindi-Urdu. We suggest two innovative probabilistic models for the problem, based on conditional and joint probability formulations. When translating a particular Hindi word given the context, our methods represent both transliteration and translation, whereas, in prior work, transliteration was only employed for translating out-of-vocabulary words. We utilize transliteration to distinguish between Hindi homonyms that can be translated, transliterated or transliterated differently depending on the context. In comparison to 14.30 for a baseline phrase-based system and 16.25 for a system that transliterates OOV terms in the baseline system, we get final BLEU scores of 19.35 and 19.00 [14]. To translate some text from one language to another language, in-depth knowledge of both the source and target language is required. Without such knowledge, the process of translation becomes cumbersome, and the result is not reliable. The major difference between English and Urdu language is due to the difference in their sentence structure. English has a “Subject + Verb + Object” sentence structure while on the other hand the Urdu language has a “Subject + Object + Verb” sentence structure. This difference can be classified into two categories: unilateral or bilateral. This classification depends upon the direction of the translation which could be either from English (target language) to Urdu (source language), Urdu (source language) to English (target language), or in both directions. Three fundamental approaches are used in machine translation [15]. The purpose of this study is to devise and evaluate a unique strategy for resolving the problem of translation from Roman Urdu to English. The method utilized to build this realistic model is separated into three steps, each of which works to accomplish its goal [16]. Tokenization is executed using a self-maintained dataset and its associated tag set. The syntactical framework is covered by writing the Urdu POS tagger based on grammatical principles. To translate Roman Urdu into English, we created the grammatical structures of several phrases. Transect performed better and provided more accurate results than Google Translator [17]. In this paper, we suggest an alternative to classic statistical MT that uses recurrent neural networks (RNNs) (SMT). We compare the performance of the SMT phrase table to that of the suggested RNN to increase the MT output quality [17]. In their study, Blossom et, they developed a comparable paradigm based on the encoder and decoder concepts. For the encoder, they employed a convolutional n-gram model (CGM), and for the decoder, they used a mix of inverse CGM and an RNN. The model’s performance was assessed by rescoring the n-best selection of phrases from the SMT phrase table [18]. This research investigates the use of triangulation and transliteration to improve Urdu to Hindi-English machine translation. They begin by introducing triangulated and transliterated phrase tables from Urdu-English and Hindi-English phrase translation models to create an Urdu-to-Hindi SMT system. Our phrase translation technique outperforms the baseline Urdu-to-Hindi system by 3.35 BLEU points. This method helps to enhance the Hindi-to-English translation algorithm [19]. Based on the research in this paper, they developed an interactive machine translation system that provides support for idioms, homographs, gender, and words with plural and singular meanings together, the corpus’ ability to expand and answer inquiries, as well as its ability to develop to a greater spectrum of coverage, are both positives translated text is sorted. Ordering is a difficult problem for computers to solve, but it is much easier for humans. The interactive system makes the user’s life easier., thus meeting the basic goal of research in this direction, which is to facilitate the user and improve the task efficiency during the translation process. Our MT system is especially well-suited to the situations for which it has been trained. This shows that we can create MT systems adapted to domain-specific demands using the concepts outlined here. Furthermore, given the size of our corpus, it appears to be suitable for embedded devices with limited memory. This MT system uses phrase-based example sets, which

offers it broader coverage [20]. Jianmo Ni et al, give the first investigation of effectively integrating de- derived from text-to-text converters (T5). T5 is proven to generate constant additional gains when scaled up from millions to billions of parameters. Furthermore, when employing sentence embeddings, our encoder-decoder approach reaches a new state-of-the-art on STS [21]. The creation of English to Tamil, English to Hindi, English to Malayalam, and English to Punjabi language pairings is the emphasis of this machine translation common work in Indian languages and produces the best BLUE score for each translation [22] Work in this domain to improve the value of QA systems, allowing consumers to better understand privacy regulations before consenting to them. To create questions, this article leverages current deep learning models such as T5, The T5- small model with labels improves its METEOR and ROUGE-L scores by 2.46 percent and 3.67 percent, respectively [23]. Dabre, R offers a survey on many language neural machine translation (MNMT), a hot topic in recent years. As a result of translation knowledge transfer, MNMT has proved effective in enhancing translation quality (transfer learning) [24] Zeeshan et al, was trained a corpus using two NMT Models (LSTM and transformer Model), and the results were compared to the desired translation using the many languages evaluation understudy (BLEU) score.

On NMT, the LSTM Model enhances the BLEU score by 0.067 to 0.41, however, the Transformer model enhances the BLEU score by 0.077 to 0.52, which is better than the LSTM Model score [25]. We looked into using a T5 model to help with four code-related activities: automatic problem correction, assert statement generation in test procedures, code summarizing, and code mutant insertion [26] Birt et al, explain the Chatbot Interaction with Artificial Intelligence using the T5, and when training data is supplemented with the T5 model, we find that all models improve, with an average improvement in classification accuracy of 4.01 percent. The RoBERT model, which was trained using T5-enhanced data and attained 98.96 percent classification efficiency, was the best [27]. Raffel et al, introduce a uniform framework that translates all text-based language issues into a T5 format to investigate the landscape of transfer learning approaches for NLP [28]. Fine-tuned Generative Pre-Trained Transformer 2 (GPT-2) model, Text-To-Text Transfer Transformer (T-5) model, and Bidirectional Encoder Conceptions via Transformers (BERT) model are the three pre-trained transformer models compared. They also found that by lowering the incidence of repetition, the transliteration-based Generative Pre-trained Transformer 2 (GPT-2) model obtains superior summarization performance [29]. Analyzing the empirical capabilities of current state-of-the-art sequence-based neural architectures in assessing tiny computer programs is extremely important. T5 Transformer can compute the output for both valid and Python code blocks with greater than 65 percent efficiency, according to tests [30]. The author introduces mT5 and mC4: massively multilingual variations of the T5 model and the C4 dataset in this paper. They demonstrated that the T5 recipe is easily adaptable to a multilingual setting and obtained excellent results on a variety of criteria. They also developed a simple strategy for avoiding illegal predictions that can occur during zero-shot evaluation of multilingual pre-trained generative models. They make available all of the code and pre-trained datasets used in this publication to promote future multilingual research [31]. In the existing text-to-text passage re-ranking model, the Author introduces the concept of multi-view learning. The suggested text-to-text multi-view architecture uses an instance mixing strategy to combine the text-generation and text-ranking objectives. The text generation view is useful in increasing re-ranking efficacy, according to our empirical study. Furthermore, the findings imply that the mixing rate for sampling cases from diverse perspectives is the most relevant aspect. They also increase the re-ranking depth of the multi-view model to test its re-ranking robustness. Even though the links between distinct views remain ambiguous, see multi-view learning as a flexible framework for achieving a more universal representation with easy additions [35]. Another research highlights recent advancements in English-to-Urdu machine translation, focusing on the use of Long Short-Term Memory (LSTM) neural networks. LSTM-based models are noted for handling Urdu's complex linguistic features, such as morphological richness and differing word order from English. Preprocessing techniques, including tokenization, grammar analysis, and word embeddings, have been employed to improve translation accuracy. The model demonstrates strong performance, with BLEU scores of 50.86 (training) and 47.06 (test), and human validation shows high translation quality. This supports the effectiveness of LSTM-based neural machine translation for structurally different languages like English and Urdu [32]. The research aims to improve Roman Urdu to Urdu transliteration using machine learning models like RNN+LSTM, Seq2Seq, and Transformer. A dataset of 6.5 million Roman Urdu sentences was used for training. The Transformer model outperformed others, achieving a BLEU score of 75 due to its ability to handle long sentences and rare words. An Android app was also developed for transliteration. The study concludes that the Transformer model is the most effective, with plans for a web application and expanded datasets [33]. The research tackles Roman Urdu spelling variations using machine learning. A dataset of 5,244 Roman Urdu words with up to five spelling variations was collected from social platforms. Six ML classifiers—SVM, LR, DT, NB, KNN, and RF—were tested, with the SVM model achieving the highest accuracy of 99.96%. The study concludes that the SVM classifier is most effective for handling spelling variations and that the dataset will aid future natural language processing research [34]. The research on Romanian to Urdu transliteration provides an understanding of the methodologies of the use of Machine Learning models including RNN+LSTM, Seq2Seq, and Transformer. The Seq2Seq models at first hit half of 48 % on the BLEU scale but failed on long sentences and words that are rarely used. However, the Transformer model with the help of attention mechanisms gave about 80% of the BLEU score and was progenitive in handling the complexity as well as compounding words [35]. Baruah, Singh, and Sarmah (2024) investigate the transliteration of Romanized Assamese social media text, highlighting the challenges due to the lack of a standardized romanization system. They develop three models—PBSMT, BiLSTM seq2seq with attention, and a transformer model for character-level transliteration. Among the models, the BiLSTM seq2seq with attention outperforms the others in accuracy [37]. Ranathunga et al. provide a comprehensive survey of neural machine translation (NMT) techniques for low-resource languages (LRLs), addressing the challenges posed by the lack of large parallel corpora. The authors review advancements in NMT for LRLs,

offering a quantitative analysis of the most widely used methods. They also present guidelines for selecting the optimal NMT approach based on specific low-resource data settings [38].

Table 1: Recent Research

Translation	Technique	Results %
Roman Urdu to Urdu	Rule-based character substitution And Unicode based character mapping techniques	95.8%
Roman-Urdu To Urdu Transliteration	Deep neural network-based encoder-decoder	70 BLEU score
Roman-Urdu and Urdu Parallel Corpus	Roman-Urdu- Parl, with 6.37 million sentence- pairs	BLEU score of 84.67
Translated Fictional Texts	Digitizing, transcribing And aligning translations of this text	10 corpus result out of 20 collected translations
English to Urdu Translation	Neural network- based deep learning technique	45.83 BLEU score
Roman-Urdu to Urdu Script	Algorithm convert the encoding roman words into the Urdu words	91.2%
Low-Resource Machine Transliteration	Neural networks—encoder decoder	60 BLEU score
STM Neural Machine Translation	LSTM encoder-decoder	Training: 50.86, Test: 47.06
Seq2Seq Sequence-to-sequence model	Transformer Model Attention-based deep learning model	75 BLEU
Spelling Variation of Roman Urdu	ML models	accuracy of 99.96

III. PROPOSED METHODOLOGY

As Roman Urdu is not a basic language, it lacks basic grammar and written vocabulary norms. The suggested strategy in this work is to establish a transliteration model for Roman Urdu, which is a unique technique that provides a good standard for Roman Urdu. This part comprises the recommended techniques and strategies for achieving the research's objectives.

3.1 Main Frame

Data pre-processing, rules-based made-up character substitution, and a Unicode-based fictional character map are the three components of the proposed technique, as shown. After starting the translator, the user can utilize a simple interface with one input frame an output frame, and a conversion key. When a consumer types a Roman Urdu text into the key box, the 1st component preprocessing strips out the superfluous information. The preprocessed Roman Urdu text then moves on to the next element. The Roman-Urdu content has been completely converted into Urdu text, which can subsequently be simply used by the user. The user may easily comprehend the precise definition of RU text and converse more expressively with other Roman Urdu users by utilizing the provided translator.

3.2 Data preprocessing

The Roman Urdu text is unprocessed and must be processed. One of the data mining strategies is to re-shape raw data using data preparation technologies. Translation models can effectively understand from preprocessed input. Real-world Roman-Urdu data is likely to have numerous inaccuracies since it is partial, inconsistent, or absent behaviors or patterns. Preprocessing data is a well-known method for fixing such issues. In the real world, data that lacks counts of elements contains errors and aberrations, or just summarized data, is considered incomplete. During preprocessing, terms,

sentences, or even complete sentences can be used as tokens. Tokenization is a concept that refers to the process of breaking down documents or phrases into individual terms to filter out non-essential keywords and punctuation. Second, in the situation of complex systems, large and lower-issue letters are treated as distinct words, therefore converting capital letters to lowercase reduces the number of unique terms in documents. This improves the feature extraction process' efficiency. 3rd, preprocessing is the method of converting data into something that a computer can realize.

3.3 Model Architecture

Transfer learning has grown in popularity as a powerful strategy in natural language processing, in which a computer is first pre-trained on a data-intensive task before being perfectly alright for a lesser goal (NLP). Because transfer learning success with a wide range of techniques, methodologies, and practices have emerged. We introduce a single framework that translates problems in every language into a text-text format in this research, which explores the landscape of transfer learning approaches for NLP. On hundreds of language understanding tasks, our systematic analysis examines pre-training objectives, architectures, unlabeled datasets, move methodologies, and more parameters.

Over a hundred languages have been pre-trained into the mT5 model. Let's look at how we might use this to train a bilingual translation model for a language with few resources, such as Roman Urdu and Urdu. The multilingual Transformer model mT5 was pre-trained on the mC4 dataset, which comprises text in 101 languages. The mT5 model's architecture (based on T5) is meant to accommodate any NLP task by recasting it as a sequence-to-sequence task. To put it another way, text enters and text exits. In a classification, for example, the text sequences to be classified can be the model's input, and the model's output will be the sequence's class label. This becomes even more straightforward in terms of translation. The input text seems to be in one language, as well as the output text is in a different language. Let's explore how we may fine-tune a mT5 model for machine translation, taking into account the multilingual capabilities of mT5 and the applicability of the sequential format for language translation. We'll be developing a translation model to convert between Roman Urdu and Urdu in this article. Because of the scarcity of resources, training excellent translation models for low-resource languages like Urdu is fairly difficult. Hopefully, the mT5 model will be able to compensate for the lack of training data in the form of straight Roman-Urdu to-Urdu (and vice versa) sequences thanks to the multilingual pre-training on a large dataset. To train the mT5 model, we'll use the Simple Transformers library (based on the Hugging face Transformers library).

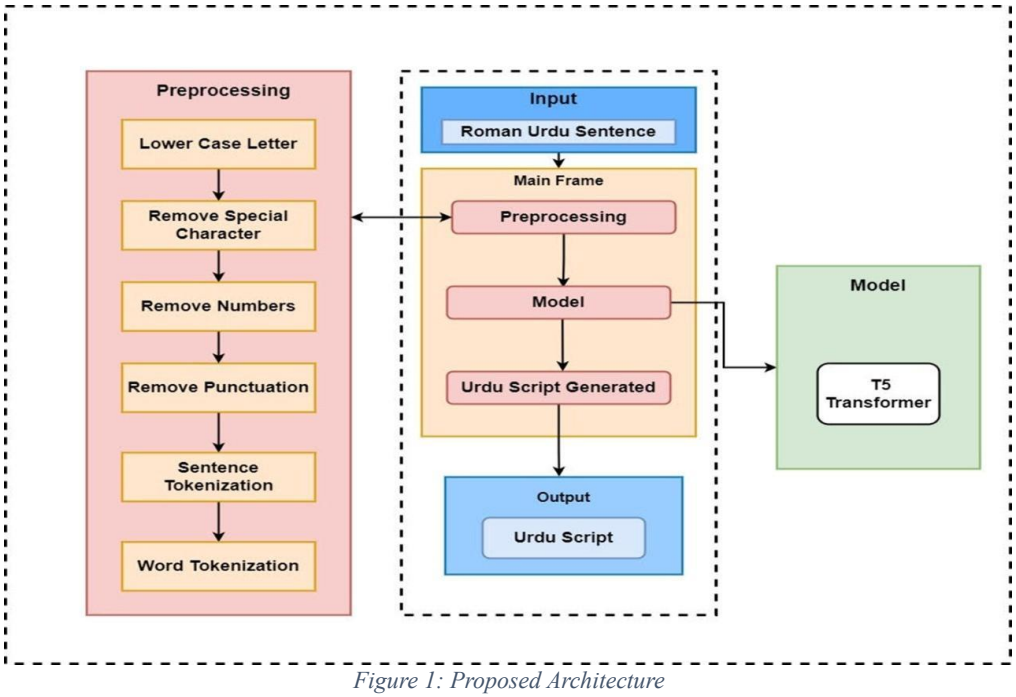


Figure 1: Proposed Architecture

Weights & Biases, which is natively supported in Simple Transformers for trial tracking and hyperparameter optimization, is used to create graphs and charts.

- T5 is a simple encoder-decoder model that has been pre-trained on a variety of supervised and unsupervised jobs, with each task transformed to text-to-text. T5 performs very well with a range of tasks right out of the box by prepending a distinct prefix to each task's input.

- T5 employs relative linear embeddings. Both the left and right sides of the encoder input can be padded.

T5 is an encoder and decoder paradigm that translates entire natural language processing difficulties into text-text format. The goal series is shifted to the right and sent into the decoder through to the decoder line, along with a start-sequence token. The EOS token then adds the target sequence, which is linked to the tags in the teacher-forcing method. The PAD token will be used to start the sequence. T5 may be fine-tuned in both supervised and unsupervised settings.

3.4 Data Set

Finally, a neat and associated Roman Urdu to A comparable corpus of 1,107,156 lines in Urdu was generated. As indicated in Table 1, our dataset comprises a whole Urdu dictionary there are 34,523 essential words and a total of 21,021 terms in the Roman-Urdu lexicon. It's worth noting that the Urdu dictionary contains more words than the Roman Urdu vocabulary. Because many words in Urdu have a gap among them but are still counted as one word. In Roman-Urdu, any such compound term is normally expressed as a single concept Com- pounding is a nice example like, ,and have their parallel "Islamabad," "bewakuf," and "ilmoadab," respectively, are Roman-Urdu transliterations of a single word.

Table 2: Details of the parallel corpus, Roman

Roman Urdu to Urdu Corpus	1,107,156
Total Roman Urdu Words	21,021
Total Urdu Words	34,523

The dataset was casually partitioned interested in three sets: train, progress, and test, with ratios of 70%, 15%, and 15%, respectively. We translated our parallel corpus into its indexed version for use in our sequence-to-sequence models. Every distinct word was assigned a numerical value. We used an indexed form of Roman-Urdu words as a response and received an indexed form of Urdu words transformed end to the initial Urdu script as a result.

Table 3: Steps to transformation

Step 1	Our Input	Sara aur Zara dost hain
Step 2	Indexed Roman- Urdu	21 52 1 664 3200
Step 3	Indexed Urdu	451 562 2343 44
Step 4	Converted Output	سارہ اور زارا دوست ہیں۔

To get beyond the one-to-one correspondence constraint based on alignment, we are additionally considering irregular-size sentences in the Roman Urdu and Urdu parallel corpora. Text messaging or tweets in Urdu or Roman-Urdu shorthand, on the other hand, are not taken into account. Because the sequence-to-sequence model can learn dependencies on its own, we didn't apply any word insertion methods to translate each word to its vector form.

IV. RESULTS

We tested the above-mentioned models using parallel corpora of Roman-Urdu to Urdu. The corpus consists of approximately 1.1 million phrases that were generated using a combination of automated and human processes. Our Roman Urdu vocabulary totals 21K words, whereas our Urdu vocabulary totals 35K. We test on the mT5 model, and so this model is pre-trained with just over a hundred distinct languages. Let's look at how we might use this to train a bilingual translation model for a language with few resources, such as Roman Urdu and Urdu. mT5 is a multilingual Transformer model that has been pre-trained on a dataset (mC4) that contains text in 101 languages. The mT5 model's architecture (based on T5) is meant to accommodate any NLP task by rephrasing the task as a sequence-to-sequence task.

4.1 Quantitative Analysis

Although the exact loss values don't tell us much, the fact that they will be falling means the model is learning, as illustrated in Figures 2 and 3. In reality, the evaluation loss appears to be decreasing, indicating that the model has not yet converged. It's possible that training for a further epoch or two will improve the model's performance.

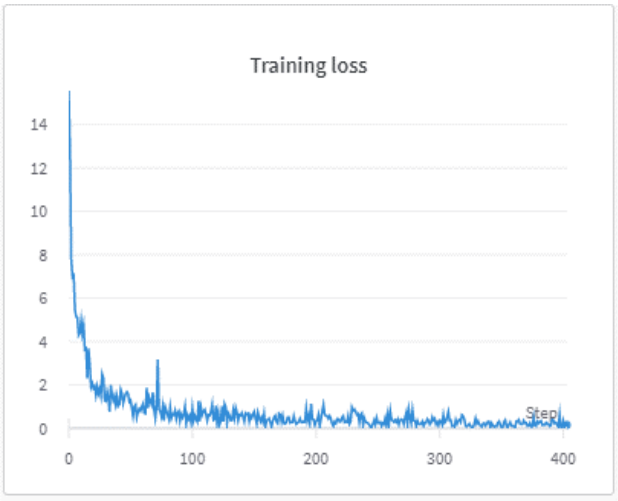


Figure 2: Training Loss

BLEU (BiLingual Evaluation Understudy) is a measure for assessing machine-translated text automatically. The BLEU score is a value between 0 and 1 which indicates how closely the machine-translated text resembles a collection of high-quality reference translations. The BLEU statistic is used to compare the output of SMT to that of human reference translations. It's vital to remember that SMT and human translations might differ greatly in terms of word usage, word order, and phrase length. To address these issues, BLEU tries to match variable-length phrases between SMT output and reference translations. The translation score is calculated using weighted match averages.

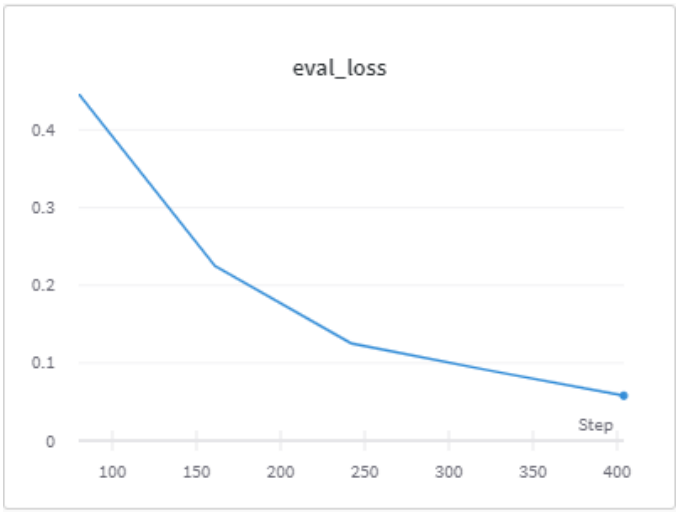


Figure 3: Evaluation Loss

The BLEU metric comes in a variety of forms. The fundamental metric, on the other hand, necessitates the computation of a shortness penalty P!

$$P_B = \begin{cases} 1, & c > r \\ e^{(1-r)}, & c <= r \end{cases} \qquad (1)$$

where r is the length of the reference corpus and c is the length of the candidate (reference) translation. The Basic BLEU metric is then determined as shown

$$BLEU = P_B exp() \sum_{N=0}^{N} w_n log P_n \qquad (2)$$

Where Wn are positive weights summing to one, and the n-gram precision Pn is calculated using n-grams with a maximum length of N. The BLEU score, specifically the BLEU system used by the annual Conference on Machine Translation, is the standard statistic for evaluating and comparing machine translation models (WMT). This score may be calculated using the Sacre BLEU library, and we achieved a good result of 91.56, as shown in Table 4.

Table 4: Details of the Bleu Score after Testing

Translation	Bleu Score
Roman-Urdu To Urdu	91.56

4.2 Qualitative Analysis

Although our models achieved a commendable BLEU score, their qualitative performance is even more impressive. Table 5 and Figure 4 illustrate these results, with the model demonstrating robust capabilities in transliteration tasks. As shown in Table 5, when provided with Roman Urdu input, the model successfully and accurately converts it into Urdu. This highlights the model's ability to handle the complexities of Roman Urdu and its transliteration to standard Urdu, as seen in various examples. For instance, the model correctly transliterates simple words such as 'wakeel' to 'وکیل' and more complex sentences with nuanced meanings, as evidenced by Sentence 5, which involves multiple clauses and detailed terminology.

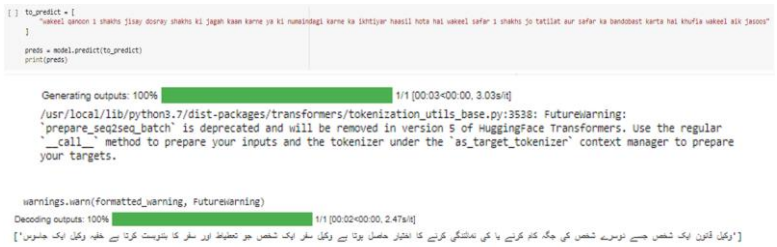


Figure 4: Output of Prediction

Table 5: Table of Output Prediction

Sentence No:	Input Sentence	Output Sentence
1	wakeel	وکیل
2	wakeel qanoon	وکیل قانون
3	woh mein wazeer taleem bhi reh chuke hain	وہ میں وزیر تعلیم بھی رہ چکے ہیں
4	Pakistan mein dehshat gardi se morad poooray malik mein majmoi dehshat grdanh karwaiyan hain	پاکستان میں دہشت گردی سے مراد پورے ملک میں مجموعی دہشت گردانہ کاروائیاں ہیں
5	wakeel qanoon 1 shakhs jisay dosray shakhs ki jagah kaam karne ya ki numaindagi karne ka ikhtiyar haasil hota hai wakeel safar 1 shakhs jo tatilat aur safar ka bandobast karta hai khufia wakeel aik jasoos	وکیل قانون ایک شخص جیسے دوسرے شخص کی جگہ کام کرنے یا کی نمائندگی کرنے کا اختیار حاصل ہوتا ہے وکیل سفر ایک شخص جو تعطیلات اور سفر کا بندوبست کرتا ہے خفیہ وکیل ایک جاسوس

V. CONCLUSION

The purpose of this paper is to design a translator for Roman-Urdu to Urdu Transliteration. As we discussed in the paper there are already a lot of translators for the above specific languages but there is no specific translator for Roman-Urdu to Urdu for long sentences and give good accuracy. As we already discussed previous translators use the RNN Technique which gives a 48.6 BLEU score and exactly gives an output of length consisting of 10 [7] shown in Figure: 6 similarly to overcome the problem of short sentences we use the RNN updated version model which is LSTM which give 70 BLEU score [2] shown in the Figure: 6 but it takes so much time to train and might be an issue with accuracy. So, in this paper, we proposed to encode decode the T5 Transformer, which is based on transfer learning the transformer we are using is multilingual and pre-trained in 101 languages due to this it improves the accuracy, and time cost as we showed after the training of our model we got a remarkable score of 91.56 BLEU score shown in figure:6 and each task, involving translations, info extraction, and categorization, is characterized T5 may be used to give the model text as input and train it to create some goal text, a transformer-based framework. After the implementation of this transformer, our results are more accurate than all other compared transformers.

These findings imply that future advances in the scale and quality of pre-trained text-to-text models might lead to even more advantages for sentence encoder models. Moreover, we will perform more variations of the Roman Urdu language in our model. And improve our translation model on different variations with high results on long paragraphs. Furthermore, we will perform an in-depth study on the essence of Urdu translation so that people can use Roman Urdu to Urdu translation easily in daily life.

REFERENCES

[1] M. Shahroz, M. F. Mushtaq, A. Mehmood, S. Ullah and G. S. Choi, "RUTUT: Roman Urdu to Urdu Translator Based on Character Substitution Rules and Unicode Mapping," in IEEE Access, vol. 8, pp. 189823-189841, 2020, doi: 10.1109/ACCESS.2020.3031393.

[2] Alam, M. and ul Hussain, S., 2021. Deep learning-based Roman-Urdu to Urdu transliteration. *International Journal of Pattern Recognition and Artificial Intelligence*, *35*(04), p.2152001.

[3] Alam, Mehreen, and Sibt Ul Hussain. "Roman-urdu-parl: Roman-urdu and urdu parallel corpus for urdu language understanding." *Transactions on Asian and Low-Resource Language Information Processing* 21.1 (2022): 1-20.

[4] Fraisse, Amel, Ronald Jenn, and Shelley Fisher Fishkin. "Building multilingual parallel corpora for under-resourced languages using translated fictional texts." *Proceedings of the 3rd Workshop on Collaboration and Computing for Under-Resourced Languages: Sustaining Knowledge Diversity in the Digital Age, Miyazaki, Japan*. 2018.

[5] Kunchukuttan, Anoop, Pratik Mehta, and Pushpak Bhattacharyya. "The iit bombay english-hindi parallel corpus." *arXiv preprint arXiv:1710.02855* (2017).

[6] Le, Ngoc Tan, et al. "Low-resource machine transliteration using recurrent neural networks." *ACM transactions on Asian and low-resource language information processing (TALLIP)* 18.2 (2019): 1-14.

[7] Alam, Mehreen, and Sibt ul Hussain. "Sequence to sequence networks for Roman-Urdu to Urdu transliteration." *2017 International Multi-topic Conference (INMIC)*. IEEE, 2017.

[8] Daud, Ali, Wahab Khan, and Dunren Che. "Urdu language processing: a survey." *Artificial Intelligence Review* 47 (2017): 279-311.

[9] Andrabi, Syed Abdul Basit, and Abdul Wahid. "[Retracted] Machine Translation System Using Deep Learning for English to Urdu." *Computational intelligence and neuroscience* 2022.1 (2022): 7873012.

[10] Shakeel, H. Muhammad, Rashid Khan, and Muhammad Waheed. "Context based roman-urdu to urdu script transliteration system." *arXiv preprint arXiv:2109.14197* (2021).

[11] Al-khresheh, Mohammad H., and Shahab A. Almaaytah. "English proverbs into Arabic through machine translation." *International Journal of Applied Linguistics and English Literature* 7.5 (2018): 158-166.

[12] Khan, Shahnawaz, et al. "Translation divergence patterns handling in English to Urdu machine translation." *International Journal on Artificial Intelligence Tools* 27.05 (2018): 1850017.

[13] Bilal, Anas, et al. "Roman-txt: forms and functions of roman urdu texting." *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 2017.

[14] Durrani, Nadir, et al. "Hindi-to-Urdu machine translation through transliteration." *Proceedings of the 48th Annual meeting of the Association for Computational Linguistics*. 2010.

[15] Babar, N., and M. Shoaib. "URDU TO ENGLISH MACHINE TRANSLATION: A REVIEW."

[16] Masroor, Hafsa, et al. "Transtech: development of a novel translator for Roman Urdu to English." *Heliyon* 5.5 (2019).

[17] Mahata, Sainik Kumar, Dipankar Das, and Sivaji Bandyopadhyay. "Mtil2017: Machine translation using recurrent neural network on statistical machine translation." *Journal of Intelligent Systems* 28.3 (2019): 447-453.

[18] Kalchbrenner, Nal, and Phil Blunsom. "Recurrent continuous translation models." *Proceedings of the 2013 conference on empirical methods in natural language processing*. 2013.

[19] Durrani, Nadir, and Philipp Koehn. "Improving machine translation via triangulation and transliteration." *Proceedings of the 17th Annual conference of the European Association for Machine Translation*. 2014.

[20] Zafar, Maryam, and Asif Masood. "Interactive English to Urdu machine translation using example-based approach." *International Journal on Computer Science and Engineering* 1.3 (2009): 275-282.

[21] Ni, Jianmo, et al. "Sentence-t5: Scalable sentence encoders from pre-trained text-to-text models." *arXiv preprint arXiv:2108.08877* (2021).

[22] Kumar, M. Anand, et al. "An overview of the shared task on machine translation in Indian languages (MTIL)–2017." *Journal of Intelligent Systems* 28.3 (2019): 455-464.

[23] Lamba, Deepti, and William H. Hsu. "Answer-agnostic question generation in privacy policy domain using sequence-to-sequence and transformer models." *2021 2nd International Conference on Electronics, Communications and Information Technology (CECIT)*. IEEE, 2021.

[24] Dabre, Raj, Chenhui Chu, and Anoop Kunchukuttan. "A survey of multilingual neural machine translation." *ACM Computing Surveys (CSUR)* 53.5 (2020): 1-38.

[25] Zeeshan, Zeshan Ali, and Muhammad Zakira Jawad. "Research on Chinese-Urdu machine translation based on deep learning." *Journal of Autonomous Intelligence* 3.2 (2020): 34-44.

[26] Mastropaolo, Antonio, et al. "Studying the usage of text-to-text transfer transformer to support code-related tasks." *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*. IEEE, 2021.

[27] Bird, Jordan J., Anikó Ekárt, and Diego R. Faria. "Chatbot Interaction with Artificial Intelligence: human data augmentation with T5 and language transformer ensemble for text classification." *Journal of Ambient Intelligence and Humanized Computing* 14.4 (2023): 3129-3144.

[28] Raffel, Colin, et al. "Exploring the limits of transfer learning with a unified text-to-text transformer." *Journal of machine learning research* 21.140 (2020): 1-67.

[29] Dhivyaa, C. R., et al. "Transliteration based generative pre-trained transformer 2 model for Tamil text summarization." *2022 International Conference on Computer Communication and Informatics (ICCCI)*. IEEE, 2022.

[30] Ganguli, Isha, et al. "Empirical auto-evaluation of Python code for performance analysis of transformer network using T5 architecture." *2021 8th international conference on smart computing and communications (ICSCC)*. IEEE, 2021.

[31] Xue, L. "mt5: A massively multilingual pre-trained text-to-text transformer." *arXiv preprint arXiv:2010.11934* (2020).

[32] Kumhar, Sajadul Hassan, et al. "Translation of English Language into Urdu Language Using LSTM Model." *Computers, Materials, And Continuum (In English)* 74.2 (2023): 3899-3912.

[33] Ahmad, A. and Ahmad, M.A., 2024. Advancing Roman Urdu to Urdu Transliteration using Machine Learning Techniques. Asian Journal of Multidisciplinary Research & Review, 5(2), pp.108-127.

[34] Soomro, Mudasar Ahmed, et al. "Spelling Variation of Roman Urdu Using Machine Learning." *Journal of Computing & Biomedical Informatics* 7.02 (2024).

[35] Ahmad, Ahsan, and Mohsin Ali Ahmad. "Advancing Roman Urdu to Urdu Transliteration Using Machine Learning Techniques". *Asian Journal of Multidisciplinary Research & Review*, vol. 5, no. 2, Apr. 2024, pp. 108-27, https://ajmrr.org/journal/article/view/9.

[36] Ju, Jia-Huei, Jheng-Hong Yang, and Chuan-Ju Wang. "Text-to-text multi-view learning for passage re-ranking." *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*. 2021.

[37] Baruah, H., Singh, S.R. and Sarmah, P., 2024. Transliteration Characteristics in Romanized Assamese Language Social Media Text and Machine Transliteration. ACM Transactions on Asian and Low-Resource Language Information Processing, 23(2), pp.1-36.

[38] Ranathunga, S., Lee, E.S.A., Prifti Skenduli, M., Shekhar, R., Alam, M. and Kaur, R., 2023. Neural machine translation for low-resource languages: A survey. ACM Computing Surveys, 55(11), pp.1-37.

Exploring the Potential of ChatGPT in Diverse Industries in Pakistan: Applications and Research Challenges

Akbar Hussain
Knowledge Unit of Systems and Technology, *University of Management and Technology, Sialkot, Pakistan*
akbar.hussain@skt.umt.edu.pk

Mujeeb Ur Rehman
Knowledge Unit of Systems and Technology, University of Management and Technology, Sialkot, Pakistan
akbar.hussain@skt.umt.edu.pk

***Abstract*— The rapid development of Artificial Intelligence (AI) has produced a conversational ChatGPT, an emerging tool has gained interest in numerous industries around the globe. This paper examines how the ChatGPT can be used to gain wholesome perspectives of respondents about their utilization of chatbot in learning activities. The sample size for this survey is 200 participants drawn from a diverse background of age and experience on purpose that included educational institutions, software houses, healthcare sectors as well as Research & Development (A&D) to understand the perception and potential in different industries within Pakistan. Additionally, we also understand that the technology has a number of research challenges and possible future avenues, including but not limited to bias mitigation, more robust ethical underpinnings for it use; increased context awareness. Results suggested that the majority of participants are familiar with this tool; This survey demonstrates a positive learning experience with ChatGPT, however emphasizes better guidelines to be provided when engaging learners in using this tool.**

Keywords*— *Artificial Intelligence, ChatGPT, Correlation, Diverse Industries, Language Translation, Teaching & Learning

I. INTRODUCTION

ChatGPT) is an advanced language model developed by OpenAI, utilizing the GPT architecture. It's fine-tuned for conversational AI tasks such as questions answering, text generation, and dialogue. The first version of ChatGPT, GPT-1 was released in 2018, this model evolved significantly. This model was based on unsupervised machine learning framework and utilized 117 million parameters [2, 4, 5]. The second version of ChatGPT, GPT-2, was released in 2019, and utilized 1.5 billion parameters, which enabling multitasking [6, 7]. The third version of ChatGPT, GPT-3 presented some advanced learning such as, meta-learning and in-context learning, boosting generalization across tasks, and utilized 175 billion parameters which is 100 times more than the second version [8 9, 10, 11]. The final version of ChatGPT, GPT-4 released in 2023, which utilize 1.76 trillion parameters, and this pilot version supporting images, handling ambiguity, diverse language, and various text inputs/outputs, showcasing impressive performance in professional and academic evaluations [1].

The potential of ChatGPT to revolutionize a number of industries such as computer technology, educational institutions, healthcare sectors, research and development centres has attracted a lot of interest. ChatGPT promotes creativity, efficiency, and improved human-computer interactions across a variety of fields to its language creation skills and interactive conversational abilities.

In this study, we carried out a comprehensive survey that how to utilize the potential of ChatGPT in diverse industries in Pakistan. A varied group of public and private business firms, including educational institutions, healthcare sectors, research centres, and software companies. We distributed 200 questionnaires with 10 questions each in an effort to learn more about respondents' perceptions of ChatGPT's applications in a wide range of industries on a scale of 1(not effective at all) to 5(extremely effective).

The purpose of conducting the survey is to assess the perception of the people related to the ChatGPT's potential in a variety of disciplines including customer service, content generation, creative writing, educational institutions, information retrieval, language translation, healthcare sectors, and virtual assistance.

The survey results assists to explain ChatGPT as an AI writing tool, and it's utility for commercial practice elsewhere. These observations would then guide all future developments and advancments in the realm of conversational AI.

The study produced several significant findings regarding the applications and future research challenges of ChatGPT in various domains. In particular, ChatGPT showed tremendous potential in the realm of customer support to provide actual help on a per-need basis. It also served as a useful AI-based virtual assistant for preparing contents in several use-cases. It has also proven

useful in enhancing educational and learning experiences online by carrying out interactive conversations that are engaging and insightful; They also demonstrated the flexibility of their method in creative writing, fact checking, language translation,, and mental health intervention applications.
Through empathic and encouraging interactions, ChatGPT assisted people communicate seamlessly across languages, increased personal productivity, helped support mental health, retrieved pertinent information , and sparked creativity.
However, the investigation also uncovered several flaws such as the possibility of response biases from the model, ethical issues with privacy and security concerns, difficulties with contextual understanding in lengthy conversations, and the requirement to assure the veracity of created information. These findings highlight the significance of ongoing research and development projects to reduce biases, create strong ethical frameworks, and raise context awareness etc.
This research contributes by thoroughly investigating the applications and future research directions of ChatGPT in education, healthcare, research, and many other sectors of Pakistan. The study also investigates the potential of ChatGPT in customer service, content generation, education and healthcare sector, language translation, and creative writing using quantitative surveys. It identifies some of its flaws, including those relating to contextual comprehension, knowledge accuracy, repetitive response creation, security and privacy concerns, data interpretation, and robustness to misleading information. The study suggests that future directions include prejudice reduction, ethical frameworks, context awareness, and knowledge verification methodologies. It helps to further the development of conversational AI technologies in the regional market by providing insights into the potential of ChatGPT and its implications for the industries. The knowledge discovered through this study adds to the expanding body of information about ChatGPT and paves the way for more developments in conversational AI technology.
The paper is organized as follows. Section 2 provides a comprehensive analysis of the existing research on potentials of ChatGPT in various industries. Section 3 presents the materials and methods of the research. It highlights the design of the study, data collection procedures used to address the research questions. Section 4, presents the results and discussions of the research. Section 5 addresses future research challenges of ChatGPT. Lastly, Section 7 presents the conclusion of the research, which summarizes the main findings of the study and their implications. The subsequent sections of this paper aim to provide a comprehensive analysis into the applications and research challenges of ChatGPT.

II. LITERATURE REVIEW

The ChatGPT, a conversational AI model, has sparked a lot of attention due to its numerous applications and advantages across a wide range of industries. This investigates that ChatGPT is as an automated customer service platform that provides real-time support and handles common difficulties [12]. It provides a flexible and cost-effective method of responding to consumer inquiries, hence boosting client connections. ChatGPT also serves as a personal assistant, scheduling appointments, imparting knowledge, and offering personalized recommendations. Its ability to replicate real-life human interactions enhances the interactive experience and encourages user participation.
The impact of ChatGPT is widely used to assist content generation such as content authors by generating recommendations, drafts, blog entries, article creation, summaries, social media headlines [13], and thereby speeding up the above contents. The conversational nature of ChatGPT enhances the learning process and encourages higher levels of student engagement activities. In a learning process, ChatGPT acts like a virtual tutor to assist students for answering queries, providing clarifications, and suggesting further investigations [14, 15] in educational institutions.
Real-time language translation features of ChaGPT have been widely used to assist users understand and communicate multiple languages effectively [16, 17]. This feature discourages barriers across different languages, and it has numerous implications for language learners and intercultural communication. ChatGPT is highly effected by its ability to comprehend questions and answers from large databases by using the efficiency of information retrieval [18]. This feature is highly utilized in online research and buying industries, in order to retrieve accurate and specific information timely. In addition, ChatGPT is used to organized calendars, manage their work, and send reminders for specific activities by using the personal productivity features [16]. By using this feature employees stay organized, increased overall productivity, track their progress, and prioritize all the activities accurately and appropriately.
ChatGPT is a helpful writing companion, providing guidance, brainstorming advice, and assisting with many writing-related tasks, such as instant writing and character development [16]. This feature helps writers of all types by encouraging creativity and assisting in the creation of compelling storylines.
In healthcare sectors, ChatGPT assists as a mental health support tool in order to provide empathy and non-judgmental conversation [19]. It is also an effective tool used for information gathering and research work for the academics and professionals. It is helpful for the data collection, methodology fluency, research abstract writing, titles suggestions of a research articles etc. [20]. It assists the academics and professionals in speeding up the research process, and faster access to the relevant information.
Natural language understanding, which allows users to communicate with models in their native language rather than using prepared commands or formats, is highlighted in ChatGPT research [21]. By providing easy and smooth interactions, this feature improves the entire user experience. It is interesting that ChatGPT can generate human-like responses that closely mimic real-life human talks [22]. By generating more intriguing and contextually appropriate material, this feature facilitates productive, engaged conversations with users.

ChatGPT's versatility and flexibility enable it to be utilized in a wide range of topics and disciplines. Because it can exchange knowledge, answer queries, and even create creative works, it can be used in a variety of settings, including customer service, coaching, content development, and so on [17].
ChatGPT's ability is to continuously train a large dataset and improve its performance [18].
In summary, while the current literature provides useful insights into the uses and benefits of ChatGPT, there are still significant research gaps that must be filled. Future research should focus on the use of ChatGPT as a content creation tool, virtual tutor, language translator, information retrieval assistant, personal productivity booster, writing companion, mental health support tool, and useful tool for research and information gathering. ChatGPT creates lively and entertaining talks with its human-like answers and natural language understanding.
ChatGPT is a powerful conversational AI model with a wide range of application possibilities due to its versatility and constant learning. Academics can contribute to a more thorough understanding of ChatGPT's potential and restrictions by filling these knowledge gaps, supporting its responsible and efficient deployment in a variety of industries.

III. METHODOLOGY

We performed a quantitative survey to examine the perception and potential of ChatGPT in several domains in order to gather comprehensive insights. The survey addressed numerous organizations in Pakistan, including five research institutions, ten software companies, ten universities, and five healthcare sectors. Our goal is to obtain information on respondents' perceptions of ChatGPT's effectiveness, value, satisfaction, and comfort using a 5-point scale through the distribution of 200 questionnaires, each of which contained 10 questions. We were able to gather a wide range of ideas and experiences by including a diverse group of participants from these sectors, which was anticipated to provide a well-rounded perspective on the perception and potential of the topic under research.
The survey was sent to the aforementioned organizations, and we received responses from each of them. This ensured that the survey sample represented Pakistan’s corporate, industrial, and educational landscape.
We were able to gain knowledge about the academic community and investigate potential uses for ChatGPT in research, education, and student support with the assistance of ten universities. Responses from software companies tell us the productivity and efficiency of ChatGPT. The participation of research institutes also sheds light on how ChatGPT is employed in cutting-edge scientific pursuits and research projects.
Additionally, we sought to comprehend ChatGPT's possible effects on various healthcare industries, such as hospitals, clinics, medical research facilities, and telemedicine providers. We were able to learn more about the suitability and benefits of ChatGPT in these particular contexts thanks to the feedback provided by these industries, each of which has its own requirements and difficulties.
The inclusion of these different groups and industries in our survey develops and extends our findings, allowing us to draw more strong conclusions regarding the perception and potential of ChatGPT in Pakistan's educational, business, and industrial landscapes.
The following research questions will help ChatGPT advance and become more dependable, effective, and trustworthy in a range of conversational applications.

Table 1 List of the quantitative survey Research Questions (RQs) using a scale of 1 to 5:

RQs	**RQs asked from 200 experienced members of educational institutions, healthcare sectors, research centers, and software companies in Pakistan.**
1	How effective do you think ChatGPT can be in offering support to customers?
2	How likely are you to utilize ChatGPT as a virtual assistant in your organization?
3	Please rate on a scale of 1 (Not useful at all) to 5 (Very useful), how useful you believe ChatGPT can be in producing content for your business requirements?
4	What potential do you see ChatGPT to improve educational and learning environments in your organization?
5	How certain are you that ChatGPT can translate languages correctly?
6	How quickly you believe ChatGPT can retrieve a certain piece of information for your personal task?
7	How likely are you to use ChatGPT as a personal productivity tool to manage events and plan your schedule?
8	Please rate your satisfaction with the creative writing capabilities of ChatGPT in your profession.

9	How helpful do you think ChatGPT can be in terms of offering support and assistant for mental health?
10	How comfortable would you feel be interacting with ChatGPT as a language model for various purposes in your organization?

Primary data was collected through structured interviews and online questionnaires for targeted professionals in a range of industries including customer service, content generation, education, language translation, information retrieval, personal productivity, creative writing, and mental health support. Proposed sample was used in the sample selection process to ensure that each industry is represented. A total of 200 participants were selected for the study based on their knowledge and experience with ChatGPT.

Descriptive statistics used to analyze quantitative data obtained from surveys. Responses are calculated their means, percentages, standard deviation, and correlation coefficients. To confirm the correctness and interpretation of the responses, member verification was done by sharing the initial findings with a selected group of participants. Cross-reference quantitative results to establish convergence and strengthen overall conclusions through triangulation.

Acknowledging the limitations of this study is crucial. Survey results may not be fully generalizable to all industries in Pakistan because they are based on a specific sample of participants. The study relied on self-reported data, which may be subject to bias or subjective interpretation. Additionally, the rapid development of ChatGPT and its applications limits access to new developments.

This study aims to gain insight into the application, strengths, limitations, and future directions of ChatGPT in various industries by using mixed methods and observing ethical concerns. With the help of this technology, a professional can thoroughly analyzed their topics, leading to insightful conclusions and recommendations for practitioners, researchers, and policy makers.

IV. RESULTS

The discussion part provides a thorough analysis and explanation of the study's findings regarding the uses, benefits, constraints, and potential uses of ChatGPT across a variety of industries. It seeks to offer explanation, ramifications, and a context for the research findings. The survey used in this study sought responses from a varied group of 30 firms including, educational institutions, industries, and organizations in order to analyze the perception and potential of ChatGPT in various fields. A 5-point scale is used throughout the survey's 10 questions to allow participants to rate their answers.

Table 2 The perception of ChatGPT across diverse Industries in Pakistan: Frequency and percentage of responses

	University Participants	**Software Houses Participants**	**Research Centers Participants**	**Healthcare Sector Participants**	**Total**	**Percentage of the Results**
Extremely Effective	70	20	5	5	100	70%
Valuable	30	8	5	7	50	30%
Satisfied	10	4	3	3	20	10%
Comfortable	15	10	3	2	30	15%
Not At All	0	0	0	0	0	0%
Total Participants	125	42	16	17	**200**	N/A
Mean	25.0	8.4	3.2	3.4	N/A	N/A
Percentage	62.5	21.0	8.0	8.5	N/A	N/A
Standard Deviation	0.345940	0.272794	0.155977	0.292967	N/A	N/A

In Table 2 out of the 200 questionnaires that were distributed in total, we received a significant number of 100 responses. A resounding 100 out of 200 respondents said that they thought ChatGPT was extremely effective. This may indicate that a

significant percentage of the participants were aware of ChatGPT's potential for customer service, content creation, education, language translation, creative writing, and mental health support, etc.

Furthermore, 20 respondents added that they thought ChatGPT was valuable. This result implies that these people understood the value and significance of using ChatGPT in their respective area of interest. The recognition of ChatGPT's worth draws attention to its potential to enhance their efforts or operations.

Additionally, 50 respondents also said they were happy with ChatGPT. This shows that a sizable portion of users were happy with ChatGPT's functionality and performance. Their pleasure suggests that ChatGPT was delivering the expected results and living up to their expectations.

Moreover, 30 respondents said they felt at comfort using ChatGPT. This result indicates that people had a favorable opinion of using ChatGPT, indicating a user-friendly interface or interaction experience. The participants' expressions of comfort show that they were at comfortable using ChatGPT for their own professions.

Interestingly, the survey shows that no one considers ChatGPT is ineffective in a given domain. Even though there are no complaints, it is important to keep in mind that surveys are only available to those who are willing to participate, and the lack of complaints does not necessarily mean that everyone is happy or that the application is effective.

The findings indicated, the sample of participants had a generally favorable opinion of ChatGPT as the potential to be a helpful tool for a variety of enterprises and organizations. It is crucial to take into account the limitations of this study, though. 200 responders is a sizable sample size, yet it could not fully represent the entire population. Additionally, the majority of the data was gathered by self-reporting, which may have been biased or inaccurate. Future research may use a larger sample size and more varied participant survey to further grasp the potential of ChatGPT.

Overall, these survey findings show that ChatGPT was well received by the participants, demonstrating its potential in the listed sectors. These results add to the expanding body of knowledge about the usefulness and worth of ChatGPT across a range of businesses and organizations. Additional investigation and study in this field may yield insightful information about how to maximize ChatGPT's advantages and mitigate any possible limitations in real-world applications.

Table 3 Correlation Matrix among different industries

	Universities	**Software Houses**	**Research Centres**	**Healthcare Sectors**
University	1.000000	0.944778	0.779512	0.554936
Software Houses	0.944778	1.000000	0.770457	0.658844
Research Centers	0.779512	0.770457	1.000000	0.930082
Healthcare Sectors	0.658844	0.554936	0.930082	1.000000

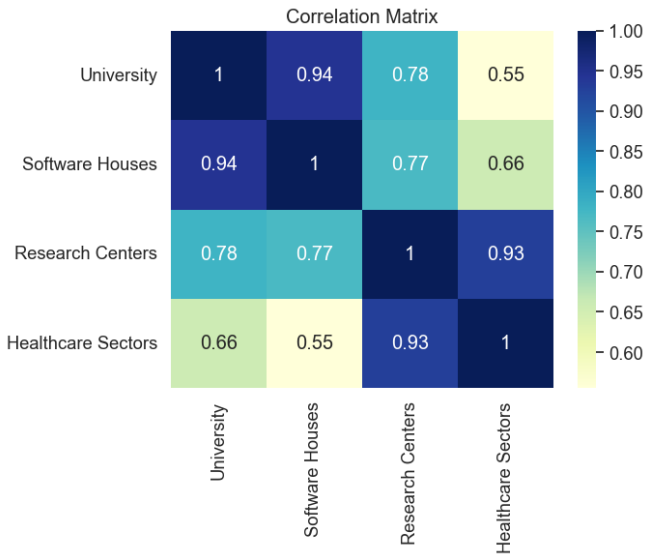


Fig. 1. Graphical representations of correlation matrix among different entities

Using the information in Table 3, Figure 1 depicts how participants shows the pairwise correlation coefficients between the university, software house, research institution, and healthcare sector variables. There is a strong correlation between universities and software house (0.945), universities and research centers (0.780), and between research centers and the healthcare sector (0.930) in this correlation matrix. There is a somewhat positive correlation (0.655) between software companies and the healthcare industry. These correlation values indicate a strong correlation between these variables, suggesting possible dependencies and patterns among them.

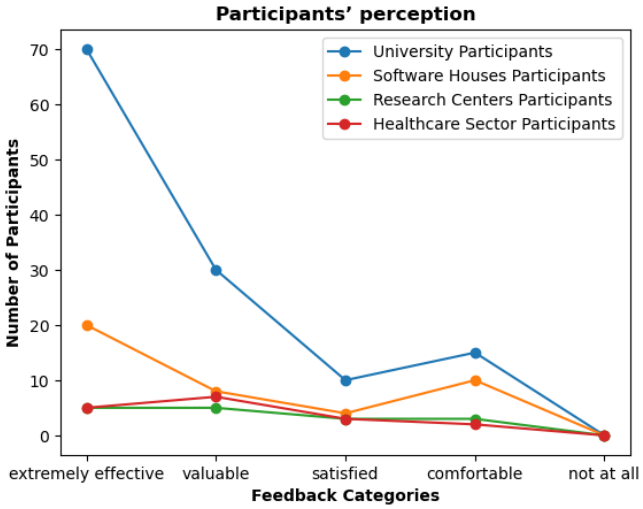


Fig. 2. Representation of participants' perception and responses from diverse industries in Pakistan

Using the information in Table 2, Figure 2 depicts how participants perceive the ChatGPT responses. It makes general trends and patterns in the feedback visible so that it can be understood naturally. Based on this representation, viewers can understand the general perception of ChatGPT and it provides insightful information on how it is received and the impact it may have locally.

According to Figure 2, 100 out of 200 respondents found ChatGPT very useful. This may indicate that a considerable number of participants in universities, software companies, research centers, and healthcare sectors are aware of the potential of ChatGPT. Additionally, 50 respondents were satisfied with ChatGPT's performance, while 20 acknowledged the usefulness of the tool. Additionally, 30 respondents indicated that they felt comfortable using ChatGPT. Notably, no participants rated ChatGPT as "totally ineffective," suggesting that no one thought ChatGPT was ineffective in any way.

V. Future Research Challenges of ChatGPT in Diverse Industries

A. Improving Responsiveness and Promptness

Enhancing ChatGPT's responsiveness and promptness is one of the big problem. These models need to generate responses as quickly and accurately as possible, which becomes especially important in real-timed conversational situations. As to where the research should go next, I think we need focus on how cut down latency (the time it takes for something to be processed) whilst still maintaining high interaction speed and thus keeping things seem really responsive.

B. Controlling Generation for Specific Outcomes

Another significant challenges is to come up with methods to control ChatGPT and make sure it generates contents based on our desired outcome or matches what user likes. This includes handling concerns in terms of bias, factual and aesthetic control over the generated content.

C. Contextual Understanding and Long-Term Memory

Another area of research is to improve the contextual understanding and improving long term memory for ChatGPT. For more extended and complex conversations, it should be able to maintain the coherence of conversation even if there is some time gap in-between responses and it must correctly remember the past parts. Conversational capacity needs to be built over extended discourse in which understanding and reasoning follow complex paths.

D. Handling Ambiguity and Clarification:

For ChatGPT, there is always the problem of having to work with ambiguities that have multiple valid answers and require more details. And it should recognize when a question or statement is ambiguous and goes to get clarification that will provide an accurate answer. Thisrea rch should concentrate on positivesolutions andbest way of handling ambiguity.

E. Ethical Considerations and Bias Mitigation

However bias management and ethical issues are still a major challenge. ChatGPT → Reducing the Bias, and mitigating potential ethical concerns such as- Creating non-aggressive/discriminatory content Potential misuses Education to Chat GPT Future research should study methods that guarantee model responses are unbiased, fair and transparent.

F. Understanding and Handling Emotional Nuances

Future work must be conducted to determine how well ChatGPT can identify and recognize emotional cue. Increased capabilities of empathy and EQ help human-like interactions with the model especially when applied to health care, counseling or support system like customer chat bot.

G. Multilingual and Cross-Cultural Conversations

Realizing this capability of ChatGPT in a effective way with multi-lingual and many-culture conversation is really challenging one from research direction. That model must be able to comprehend in several aforementioned languages, reply the queries and attune itself with more than one terrain. Technologies that empower multilingual and intercultural capabilities need to be explored for wider acceptance.

H. Robustness to Noisy and Misleading Inputs

ChatGPT needs to handle noisy or wrong input so that its answers are accurate and correct. It can manage and identify off topic, change the subject if required, seek clarification etc.

I. User Feedback and Interaction Improvement:

Improvements to user-feedback based chatbot response adaptation. Reinforcement learning, active learning and user modelling methods could be developed to improve the responsiveness (and flexibility) of this model.

J. Privacy and Security Concerns

Ensuring Security and Privacy for ChatGPT is a Challenge in future. Researchers should develop methods to protect user privacy and confidentiality, prevent data leakage from occurring in sensitive areas while still having utility, as well as mitigate the risks.

VI. CONCLUSION

In conclusion, this study provides insightful information on the potential applications and research challenges of ChatGPT in diverse industries in Pakistan. A survey of 200 participants revealed its great potential for customer service, virtual assistance, content creation, education, translation, information retrieval, productivity, writing and mental health support. ChatGPT is largely seen by participants as highly comfortable, valuable, and satisfying in educational institution, research centers, software companies, and healthcare sectors.

Future research should expand the sample and involve a diverse range of people, while acknowledging study limitations (sample size, self-report). Addressing bias, model robustness, and scalability are critical.

REFERENCES

[1] Gpt-4 technical report, 2023, [online] Available: https://cdn.openai.com/papers/gpt-4.pdf.

[2] A. Radford, K. Narasimhan, T. Salimans and I. Sutskever, Improving language understanding by generative pre-training, 2018.

[3] M. Khosla, A. Anand and V. Setty, "A comprehensive comparison of unsupervised network representation learning methods", arXiv preprint, 2019.

[4] Q. Y. Sun, C. Q. Zhao, Y. Ju and F. Qian, "A survey on unsupervised domain adaptation in computer vision tasks", Sci. Sinica Technol., vol. 52, no. 1, pp. 26-54, 2022.

[5] C. Ieracitano, A. Paviglianiti, M. Campolo, A. Hussain, E. Pasero and F. C. Morabito, "A novel automatic classification system based on hybrid unsupervised and supervised machine learning for electrospun nanofibers", IEEE/CAA J. Autom. Sinica, vol. 8, no. 1, pp. 64-76, Jan. 2021.

[6] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei and I. Sutskever, "Language models are unsupervised multitask learners", OpenAI Blog, vol. 1, no. 8, pp. 9, 2019.

[7] Y. Zhang and Q. Yang, "A survey on multi-task learning", IEEE Trans. Knowl. Data Eng., vol. 34, no. 12, pp. 5586-5609, Dec. 2022.

[8] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, et al., "Language models are few-shot learners", Proc. 34th Int. Conf. Neural Information Processing Systems, pp. 1877-1901, 2020.

[9] C. Finn, P. Abbeel and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks", Proc. 34th Int. Conf. Machine Learning, pp. 1126-1135, 2017.

[10] J. Beck, R. Vuorio, E. Z. Liu, Z. Xiong, L. Zintgraf, C. Finn, et al., "A survey of meta-reinforcement learning", arXiv preprint, 2023.

[11] Q. X. Dong, L. Li, D. M. Dai, C. Zheng, Z. Y. Wu, B. B. Chang, et al., "A survey on in-context learning", arXiv preprint, 2022.

[12] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901.

[13] Liebrenz, M., Schleifer, R., Buadze, A., Bhugra, D., & Smith, A. (2023). Generating scholarly content with ChatGPT: ethical challenges for medical publishing. The Lancet Digital Health, 5(3), e105-e106.

[14] Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. Contemporary Educational Technology, 15(3), ep429.

[15] FIRAT, M. (2023). Integrating AI applications into learning management systems to enhance e-learning. Instructional Technology and Lifelong Learning, 4(1), 1-14.

[16] Jiao, W., Wang, W., Huang, J. T., Wang, X., & Tu, Z. (2023). Is ChatGPT a good translator? A preliminary study. arXiv preprint arXiv:2301.08745.

[17] Hariri, W. (2023). Unlocking the Potential of ChatGPT: A Comprehensive Exploration of its Applications, Advantages, Limitations, and Future Directions in Natural Language Processing. arXiv preprint arXiv:2304.02017.

[18] Lin, Z., Cui, S., Li, G., Kang, X., Ji, F., Li, F., ... & Zhang, Y. (2021). Predict-Then-Decide: A Predictive Approach for Wait or Answer Task in Dialogue Systems. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 29, 3012-3024.

[19] Muftić, F., Kadunić, M., Mušinbegović, A., & Abd Almisreb, A. (2023). Exploring Medical Breakthroughs: A Systematic Review of ChatGPT Applications in Healthcare. Southeast Europe Journal of Soft Computing, 12(1), 13-41.

[20] Curtis, N. (2023). To ChatGPT or not to ChatGPT? The impact of artificial intelligence on academic publishing. The Pediatric Infectious Disease Journal, 42(4), 275.

[21] Kraugusteeliana, K., Indriana, I. H., Krisnanik, E., Muliawati, A., & Irmanda, H. N. (2023). Utilisation of ChatGPT's Artificcial Intelligence in Improving the Quality and Productivity of Lecturers' Work. Jurnal Pendidikan Dan Konseling (JPDK), 5(2), 3245-3249.

[22] Gao, M., Ruan, J., Sun, R., Yin, X., Yang, S., & Wan, X. (2023). Human-like summarization evaluation with chatgpt. arXiv preprint arXiv:2304.02554

Exploring E-Commerce Predictions: Integrating Social Media

Hadia Noor
School of Systems and Technology
University of Management and Technology
Lahore, Pakistan
F2020266503@umt.edu.pk

Muhammad Sarwar
School of Systems and Technology
University of Management and Technology
Lahore, Pakistan
muahmmad.sarwar@umt.edu.pk

Jameel Ahmad
School of Systems and Technology
University of Management and Technology
Lahore, Pakistan
jameel.ahmad@umt.edu.pk

Muhammad Husnain Ashfaq
School of Systems and Technology
University of Management and Technology
Lahore, Pakistan
husnain.ashfaq@umt.edu.pk

***Abstract*—This study focuses on the use of social media metrics as indicators in predictive models for E-Commerce to increase the effectiveness of E-Commerce prediction analysis. This research aim is interested in the impact of likes, shares, comments, reactions on most E-Commerce essential performance indicators that include the total revenue, the number of unique customers, Average order value (AOV), and total orders. It entails gathering social media data alongside conventional E-Commerce point of sale data, cleansing this data, and using disparate machine learning algorithms to gauge the validity of each model. It involves the use of Linear Regression, Decision Trees, Neural Networks for regression, though some ensemble methods such as Gradient boosting and Random Forest regression are also employed. Accordingly, from these evaluation results, it can be concluded that Gradient Boosting as well as Random Forest can reach higher accuracy in terms of total orders and the number of unique customers while Linear Regression and Lasso Regression can achieve better predictions for the total revenue. The analysis of the features showed that variables like the engagement rate, total impressions, and sentiment score make the difference when it comes to e-commerce thematic. Finally, integrating social media metrics allows for a better analysis of consumer habits as a means of valuable information to be used to plan key business actions or promotional campaigns. The future work of this kind of research should be directed towards the temporal predictions and the ways of refining the quality of the data collected to increase the stability of the models under consideration**.

Keywords— E-Commerce Lasso Regression, Linear Regression, Average order value

I. INTRODUCTION

In today's ever-evolving world of online shopping, it's becoming increasingly important for businesses to grasp the significance of social media metrics in shaping their success. This study delves into how our interactions on social platforms—like the likes, shares, comments, and reactions we give—affect crucial e-commerce measures. By examining the correlation between social media engagement and key performance indicators such as overall revenue, number of unique customers, average order value, and total orders, we aim to offer valuable insights for companies seeking to harness the power of social media data.

Our research will explore the nuances of social media interactions, highlighting how positive engagement can enhance brand visibility and foster customer loyalty. We will also investigate the impact of viral content, user-generated posts, and influencer partnerships, demonstrating how these elements can drive traffic to online stores and convert followers into dedicated customers.

Additionally, we anticipate that our discoveries will provide a deeper understanding of how social media interactions can guide strategic choices, enabling businesses to navigate the online marketplace with more foresight and impact. By leveraging analytics to assess the effectiveness of social media campaigns, companies can refine their messaging, target specific demographics more efficiently, and ultimately increase their return on investment.

As businesses increasingly incorporate social media into their marketing approaches, acknowledging its influence on e-commerce performance is vital for building meaningful relationships with consumers and driving growth. In an age where customer preferences shift rapidly, understanding the dynamics of social media engagement can equip companies with the tools necessary to adapt and thrive. This study ultimately aims to bridge the gap between social media activity and measurable

business outcomes, empowering organizations to make data-driven decisions that enhance their online presence and foster sustainable success.

II. BACKGROUND

The digitalization period has drastically transformed the nature of buyer-seller relations as observed in the case of online shopping. This shift has been driven by social media as it plays the role of a mediator in controlling consumer behaviors and offering insights to organizations. This research project dives into a new aspect of E-Commerce: incorporating social analytics into the E- Commerce forecast model.

Through the integration of social media analytics and E-Commerce prediction, this project is situated at the intersection of data science, machine learning, and digital marketing. While old school E-Commerce models rely on past sales data and consumers' attributes, metrics like 'likes' and 'comments' shed a special light on consumers. However, integrating these multiple sources of such heterogeneity poses some challenges, such as ethical use of data and methods of handling and analyzing textually structured social media data.

Therefore, the potential impact of this research is apparent. Social media metrics then provide an ideal way for organizations to improve on their marketing, customer satisfaction, and profitability. Not only does this research serve to complement previous literature on the topic, but it also serves as the foundation for subsequent advancements in E-Commerce analytics. The presented case indicates that integrating social media metrics is possible and might be used as an example for other industries seeking to employ digital engagement factors in business analysis.

Interdisciplinary work is underlined in this project, so data scientists, marketers, and strategists work together to create integrated models of consumption that would accurately depict shoppers' behavior. This calls for a change in the business market strategy depending on the results from the social media analytics may change the organization's structure.

However, this research is not just about combining data; it is about employing the social media tool for the purpose of the digital marketplace. Consequently, the findings of this project should be presented for the progression of future academic studies in E-Commerce and for the improvement of the analytics of this area. It focuses on the need for applying the data processing approaches on the territory of the present-day world wide web and offers companies the relevant information to compete in the modern environment.

III. LITERATURE REVIEW

Using social media metrics in E-Commerce predictions can be deemed as a change in the business and consumer interactions natural habitat due to technology. This has mainly been demonstrated by e-commerce, and the overseas sales are expected to rise to $4. In 2021, it was 89 trillion, which proves that this sector is critically important *[1]*. This growth is related to the impacts of social media on consumer where 54% of social media users engage in product search on social media platforms *[2]*. Pioneering studies by Mangold and Faulds *[3]* and Kozinets et al. *[4]* have noted in their early studies how the social media shift from mere communication medium to an essential means of interacting with customers and doing business online.

From E-Commerce perspective it has been postulated that for every 5% increase in customer retention there is an increase in profit by 25% – 95%, this underscores the importance of customer behavior analysis or even prediction. Nevertheless, it often seems that traditional E-Commerce Prediction models are not enough to utilize the potential qualitative data available in social networks, which could otherwise supplement these *[5]*.

It can be concluded that incorporating SMM as part of the analytical approach that builds predictive models is effective in enhancing consumer engagement. For example, companies employing analytical tools like social networks have noted an enhancement of ROI by 15-20% *[6]*. The research of Trainor et al. *[7]* and De Vries et al. *[8]* offers the proof of the social media metrics to change the customer relationship and E-Commerce plans.

Comparative analysis with other predictable models in Zhang et al. *[9]* and Verhoef et al. *[10]* reveal that models involving social media measurement exhibit superior performance in terms of predictive accuracy and also compliance with the existing consumer trends.

Marketing theory, as conceptually explained by Kotler and Keller, and the trends in application of Artificial Intelligence and Machine Learning, as discussed by Davenport et al. *[12]* and McAfee and Brynjolfsson *[13]* are changing the ways and means for making the E-Commerce predictions. The adoption of AI in various business activities is expected to cause a fundamental shift in the GDP, which could grow by 14% globally by 2030 *[14]*.

Real-life examples of models that have been demonstrated here include established leading E-Commerce beings such as Amazon and Alibaba. For example, Amazon has reported that as much as 35% of its sales are the direct result of its recommendation engine, thus proving that advanced predictive models work in real-life applications. Information about these applications are availed by Smith and Linden *[15]* and Chaffey *[16]*.

However, this field has issues especially with data privacy and processing social media unstructured data as expounded by Goldfarb and Tucker *[17]* and Fan and Gordon *[18]*. According to McAfee et al. *[19]* the future direction of this field involves the application and evolution of real time analytics as posing both a threat and a prospect.

As pointed out by Tuten and Solomon *[20]*and Malthouse et al. *[21]* , the rising role of social media analytics in the E-Commerce context is underlined by the growing number of marketers who consider this approach valuable. According to James et al. *[22]*, the ML and data mining perfect for predictive analysis are as follows the trend that stands out is that 60% of e-commerce firms use artificial intelligence.

It is important to consider ethical dimensionality in the use of data, and critical perspectives are provided by Boyd and Crawford *[23]* and Tene and Polonetsky *[24]*. One interesting fact reveals that 81% of customers are concerned about the dangers of data gathering *[25]*. Bharadwaj et al. *[26]* and Westerman et al. *[27]* reflect on how digitization affects business models, including E-Commerce, and reveal that more digitally developed companies are more profitable.

Subsequent research has simply expanded on this body of work. For example, in the context of customer lifetime value modelling, Sharma notes on the role of social media data in enhancing the ensuing predictions *[28]*. In the same manner, Delgado *[29]* provides rich information about integrated digital tactics into E-Commerce where he stresses the significance of data analysis to business.

Research has been conducted to compare the predictive capability of different E-Commerce Prediction models and the results proved that including the non-financial information like, social media engagement indices improve the accuracy of the models. For example, a new approach by the researchers in the B2B SaaS domain revealed that by using machine learning frameworks, it is possible to predict the customer lifetime value which would be beneficial for the marketing strategies *[30]*.

Additionally, the HelloFresh case study provides insights on how predictive models for E-Commerce metrics can help to enhance marketing investment, which is inspiring for the implementation of such models *[31]*. Another article that explored the application of machine learning in online retail underscored the significance of SPL on customer retention based on predictive analytics, further supporting the centrality of data analytics in E-Commerce *[32]*.

Thus, these findings are consistent with trends in the literature, where the combination of AI and machine learning is viewed as critical for increasing the efficiency and accuracy of prediction models in E-Commerce. For instance, a current deep probabilistic model for predicting customer lifetime values emphasizes the importance of state-of-the-art machine learning approaches to offer clearer and more useful insights *[33]*.

In conclusion, the incorporation of social media metric and analytics to E-Commerce models has working utility to the E-Commerce firms. Thus, by applying these data, it is possible to build a stronger marketing experience, generate more revenues, improve customers' experience, and obtain better organizational outcomes.

IV. METHODOLOGY

The literature review indicates that the integration of social media metrics in Ecommerce predictive models is somewhat of a nascent area with considerable possibilities.

A. Data Acquisition and Pre-Processing Stage

During the first stage, the engagement data from the social media platforms, the E-Commerce, transactional, and are received. Som of the data that can be found on social media include: reactions, comments, shares, or other engagement metrics. This data is then scrubbed and normalized in order to identify the outliers within the data set, so that the data may be analyzed with the use of a predictive model.

B. Exploratory Data Analysis

Some of the most key findings drawn from the exploratory data analysis of the given set of data include the following. The example variables embedded in the dataset are simple social media metrics and E-Commerce KPIs. The data was inspected: Firstly, the dataset was imported into a Pandas Data Frame and simple statistical measures as the number of records, data types, and missing data were obtained. This was followed by a missing value analysis to understand the impact of any missing values in the dataset.
During data visualization, distribution plots were created and the distributions of Total_Revenue, Average_Order_Value and Total_Orders were assessed. The above plots indicate that Total_Revenue has a positively skewed distribution meaning that most of the transactions result in comparatively low total revenue and very few transactions result in high total revenue. Also, the Average_Order_Value distribution is positively skewed meaning that most of the orders average value is low while a few orders have enormously high average values implying that while most of the orders are relatively small, there are extremely large orders. The Total_Orders distribution also shows that a lot of customers only placed a few orders, suggesting a significant number of inactive customers.

Second, the Unique_Customers per transaction is relatively balanced, which shows that there are no significant fluctuations in customers' transactions. This implies that Total_Impressions demonstrates that most posts or advertisements usually garner moderate impression level with a few attaining high impression figures. The Total_Reactions are skewed toward lower values, meaning most of the posts experienced low to average levels of reactions with few of them having many reactions. Total_Comments show that most of the posts are commented only a few times while Total_Shares also show that shares are less frequent with some posts being shared widely. Total_Clicks are generally small but some of the posts receive significantly more clicks while Description_Length varies within a small range which implies relatively similar post lengths. Description_Sentiment scores are different, which means that there are different emotional backgrounds. It is easily seen that the overall distribution of Total_Engagement and Engagement_Rate is slightly lower for most of the posts, but there are posts that have received high levels of engagement and rates.

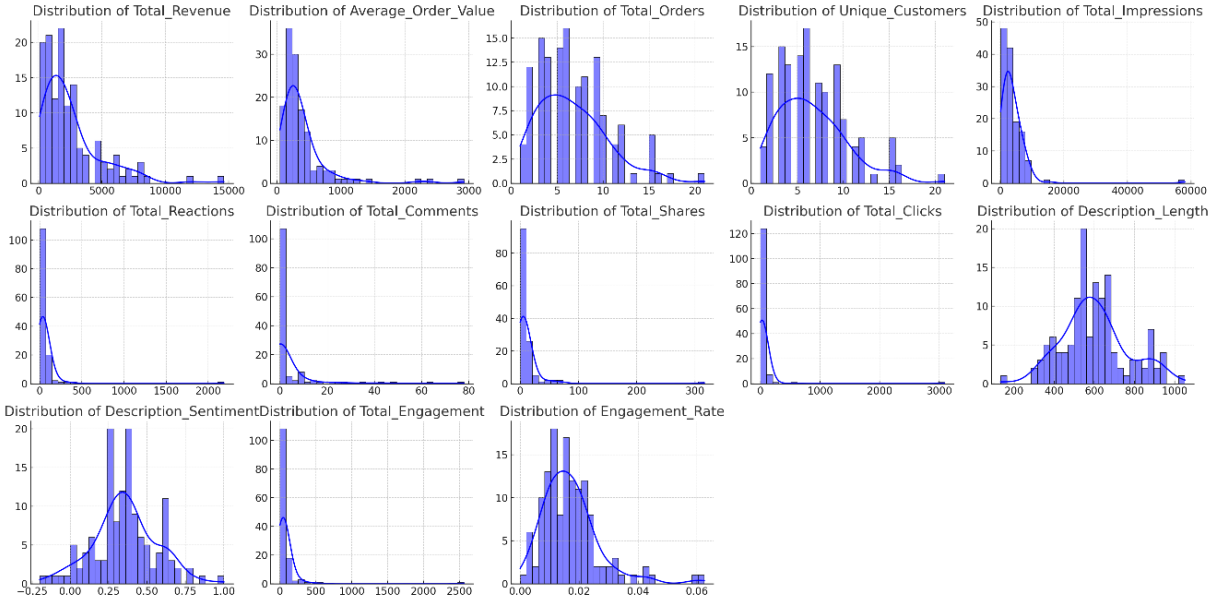


Figure 1 *Distribution Plots*

In correlation analysis, correlation coefficients are calculated to determine the relations between features and target variables with considerable significance. A heatmap is used to display the correlation matrix, which can help identify variables to include in model building

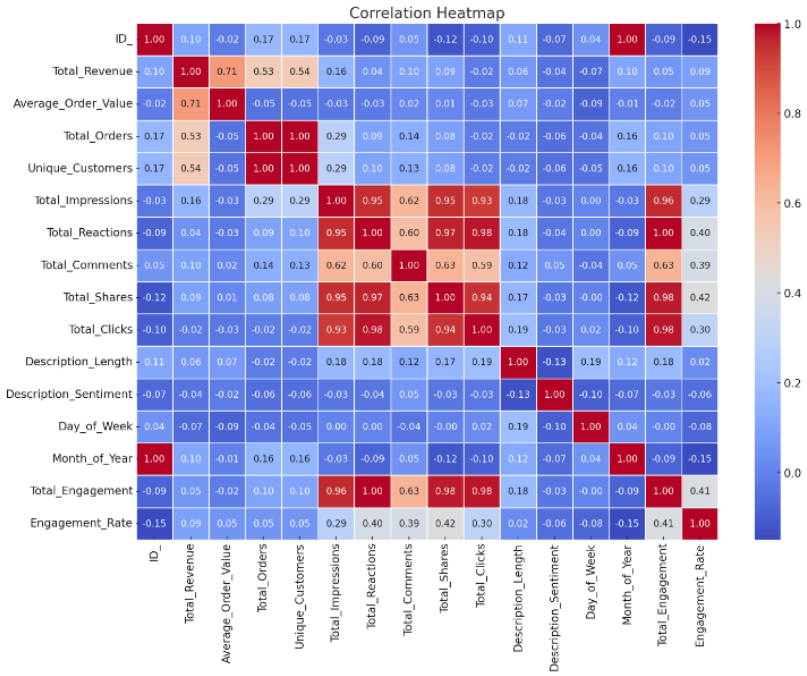


Figure 2 *Heat Map*

C. Train Test Split

The dataset is randomly split into training and testing sets in order to assess the effectiveness of the obtained predictive models. This step makes it possible for the models to be trained on one segment of the data and then tested on an entirely different segment of the data hence providing a more accurate and true results.
Steps involved in Train Test Split:

i. Defining Features and Target Variables:

Choose the appropriate features (social media and E-Commerce analytics) and specify the target variables (Total Revenue, Average Order Value, Total Orders, Unique Customers).

ii. Splitting the Data:

Split the dataset into the training set and the testing set using the train_test_split function of Scikit-learn at a ratio of 8:2. Choose the number of records to include in the test to prevent overfitting and define the random state to make results reproducible.

iii. Standardizing the Data:

Standardize the features such that they have a mean of 0 and a standard deviation of 1. This step assists in enhancing the efficiency as well as the rate of convergence of the machine learning models. The data needs to be explored, visualized and divided correctly in order to make the models ready to learn from the data and function optimally on unseen data. It also plays a role in creating good predictive model work to optimize the efficiency of E-Commerce conclusions with the use of statistics from the sites of the social networks.

D. Model Development and Training Stage

On the integrated dataset various of this and that predictive modeling techniques are used and created. Often when creating models, one has to incorporate social media and E-Commerce data in the ML frameworks that are being developed. For ascertaining whether social media metric has to be incorporated in E-Commerce predictions or not, Linear Regression, Polynomial Regression, Ridge Regression, Lasso Regression, Elastic Net Regression Decision Tree Regression, Random Forest Regression, Gradient Boosting, K- Nearest Neighbors Regression as well as CNN has been tested & trained.

E. Evaluation and Strategy Formulation Stage

Final step is to validate the performance of the models by using commonly known statistical measures including R-Squared, Mean Squared Error (MSE) and Mean Absolute Error (MAE). These metrics are used to decide upon the superior models for each feature. In addition, a neural Network (NN) model is done to analyze the feasibility of utilizing such a model in estimating these E-Commerce factors.

The findings of these models are used to give strategic solutions which E-Commerce organizations require. This approach was aimed at providing recommendations that are realistic and realistic planning that can enhance the marketing strategies of E-Commerce businesses and overall performance.

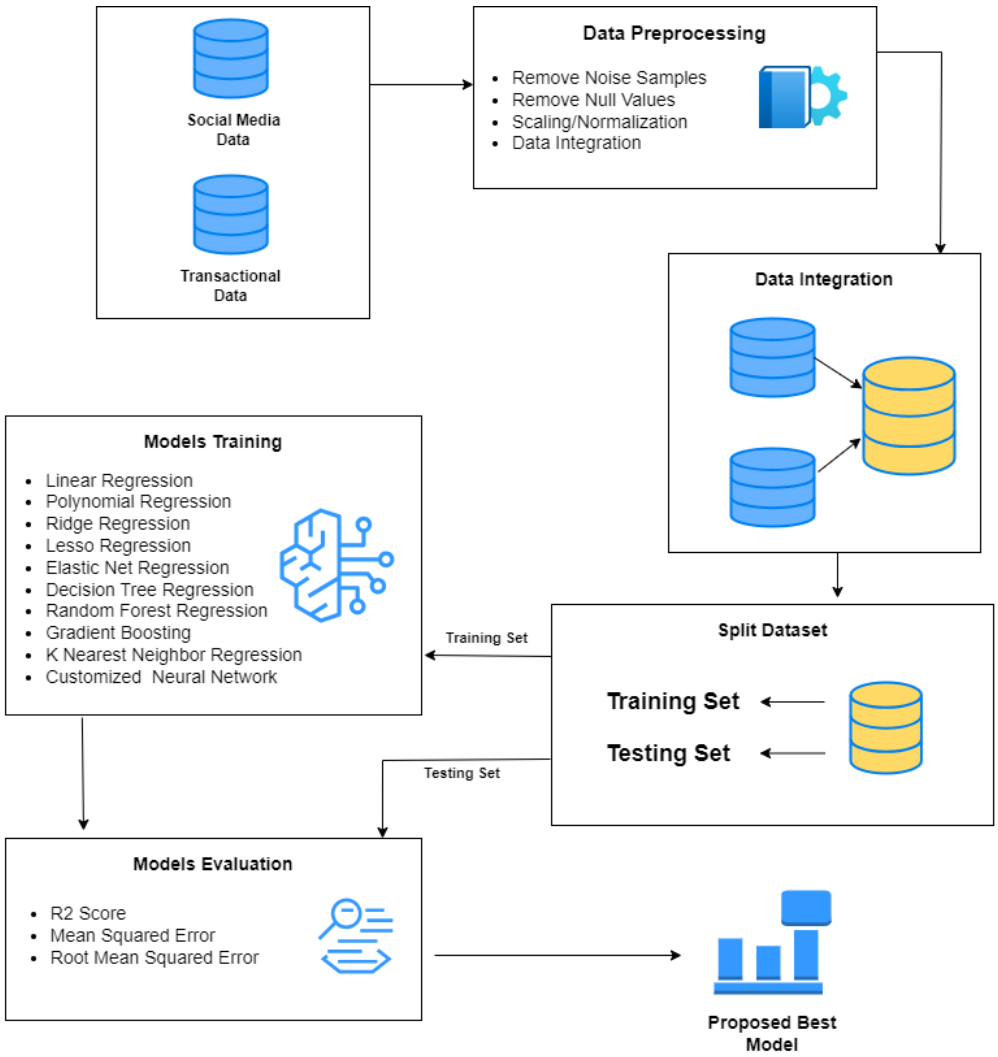


Figure 3 Suggested Approach

.

Workflow of the system

The operational plan of the proposed research system comprises the stages required to establish a stable framework to utilize measurements from social media to improve the E-Commerce prediction models.
The steps are detailed below:

Step 1: Collect engagement data from Meta Suite and purchase transactional data from Magento to create a new data set. The social media metrics consist of the reach, impressions, post-level reactions, comments, shares, and overall click-through rate, whereas the E-Commerce data covers total sales, number of customers, average order value (AOV), and number of total orders.

Step 2: Preprocess the data by performing the following steps: map different engagement metrics against each distinct post ID, define a one week engagement window within which to aggregate these metrics, perform sentiment analysis on the description fields of the posts to capture the tone of the content, augment the engagement and sentiment data by adding new features that models can learn from, and merge the social media metrics with the transactional E-commerce data.

Step 3: Exploratory data analysis (EDA) is the process of discovering patterns in the data. Plot histograms of the distributions of Total_Revenue, Average_Order_Value, Total_Orders. Carry out correlation analysis and use the heatmap to determine the degree of association between the SM metrics and e-commerce results.

Step 4: Train Different types of machine learning models like Linear Regression, Polynomial Regression, Ridge Regression, Lasso Regression, Elastic Net Regression, Decision Tree Regression, Random Forest Regression, Gradient Boosting Regression, K-NN Regression, and Neural Networks (NN).

Step 5: Test the models with the help of a 20 % of training set and find out the various performance indicators including R-Squared, MSE and MAE.

Step 6: Compare the results generated with the different models in an effort to assess the efficacy of each in blending social media analytics into E-Commerce prediction models to identify the most suitable model by conducting comparative analysis of the models.

Step 7: Conduct Feature Importance Analysis to investigate the ranking of features and the performance of the model to determine the critical social media indicators that impact E-Commerce performance.

Step 8: Use the trained models for E-Commerce predictions from the integrated social media and E-Commerce data set.
Step 9: Examine the results of the model to determine the significance of different variables of social media performance on E-Commerce effectiveness.

Architecture/Algorithm

The architecture of proposed neural network is particularly tailored for the merger between social media analytics and e-commerce prediction. The main aim is to improve the forecast of values like total revenues, new customers, average order value (AOV), and total orders with reference to data from social media.

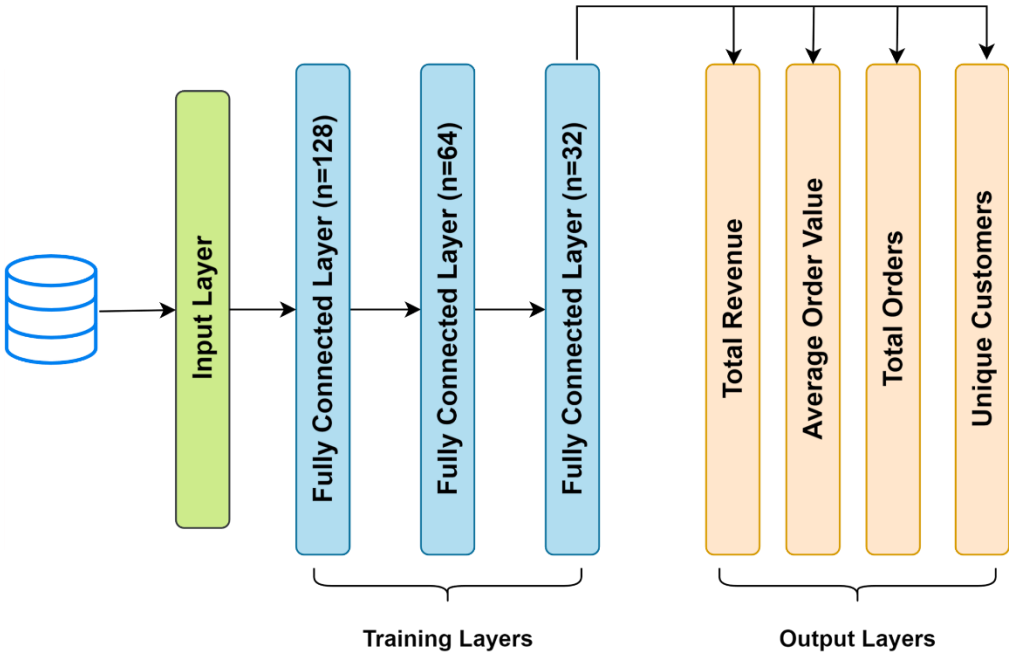


Figure IV Proposed Model Architecture

i. Input Layer

The input layer of the proposed neural network comprises features generated from social media and e-commerce data. Some examples of SMI are likes, shares, comments, reactions, total impressions, engagement rates and scores, and sentiment scores. These features are combined with conventional e-commerce measures like the overall revenue, the number of individual customers, AOV, and the total number of orders to create a robust input for the predictive model.

ii. Hidden Layers

The neural network is composed of multiple hidden layers, where input data gets transformed through several operations. The first hidden layer is the dense layer, where a ReLU activation function is applied in order to discover the non-linear correlations in the data. This layer helps the network to analyze and optimize data to enable the identification of complex patterns. The second hidden layer, which is also a dense layer with ReLU activation function, enhances the data representation of the second layer based on the features extracted in the first hidden layer. These layers help to find more abstract and useful representation of the input features for the prediction.

iii. Output layer

The output layer of the developed neural network aims at determining several e-commerce factors at once. It has four neurons, where each neuron corresponds to the prediction of one of the KPIs. The neuron used for total revenue prediction has linear

activation function to pass a continuous value as the output. Like that, neurons are predicting the number of total unique customers, average order value, and the total number of orders. These outputs are useful for e-commerce to make strategic decisions that would enable them to improve on their performance.

iv. Training and Evaluation

Evaluation of the model is carried out using training as well as the testing set elicited from the neural network dataset collected. Least squares approach is used in training the network where Mean Squared Error (MSE) was used as the loss function to minimize the prediction errors. Desirable properties of the Adam optimizer include efficiency and its suitability for training deep learning models. Such statistical measures include Coefficient of determination (R-squared – R^2), Mean-Squared Error (MSE), and Mean-Absolute Error (MAE) that help in determining the performance of the model.

V. Results And Discussion

Experimentation

This section also outlines the various methods, tools used, procedures followed and measures put in place during the experimentation in order to encourage replicability.

Experimental Setup

In this study, Google Colab was used as the main computational platform for training and testing the predictive models. The resources available in Google Colab included 12. 72 GB of RAM and nearly 100 GB of disk space necessary to process the machine learning models. The data included E-Commerce performance data such as total revenue, unique customers, Average Order Value, total number of orders, and social media data from Facebook, Instagram and Twitter profiles.
As part of data preprocessing, various engagement ratios were captured against each post and customer. The preprocessing included joining all data with a unique post ID and setting up a one-week impact window to sum the engagement metrics. The tone of the posts was determined by conducting sentiment analysis on the descriptions of the posts. This sentiment data was then combined with the subsequent related engagement KPIs together with revenue and AOV. Afterwards, this comprehensive data was cleaned, normalized, and preprocessed for the purpose of model training.
All the mentioned models such as different regression techniques and a Neural Network (NN) were developed and trained using Python tools like Pandas, NumPy, Scikit-learn, and TensorFlow. The data was divided into training and testing datasets for proper analysis, and the results were analyzed using measures like R-squared, Mean Squared Error (MSE), and Mean Absolute Error (MAE) according to eq. **i** to **iii** were used to assess the performance of the model.
.

Feature Importance Analysis

After the training, the feature importance was run to understand which of the social media metrics were most influential in the predictions.

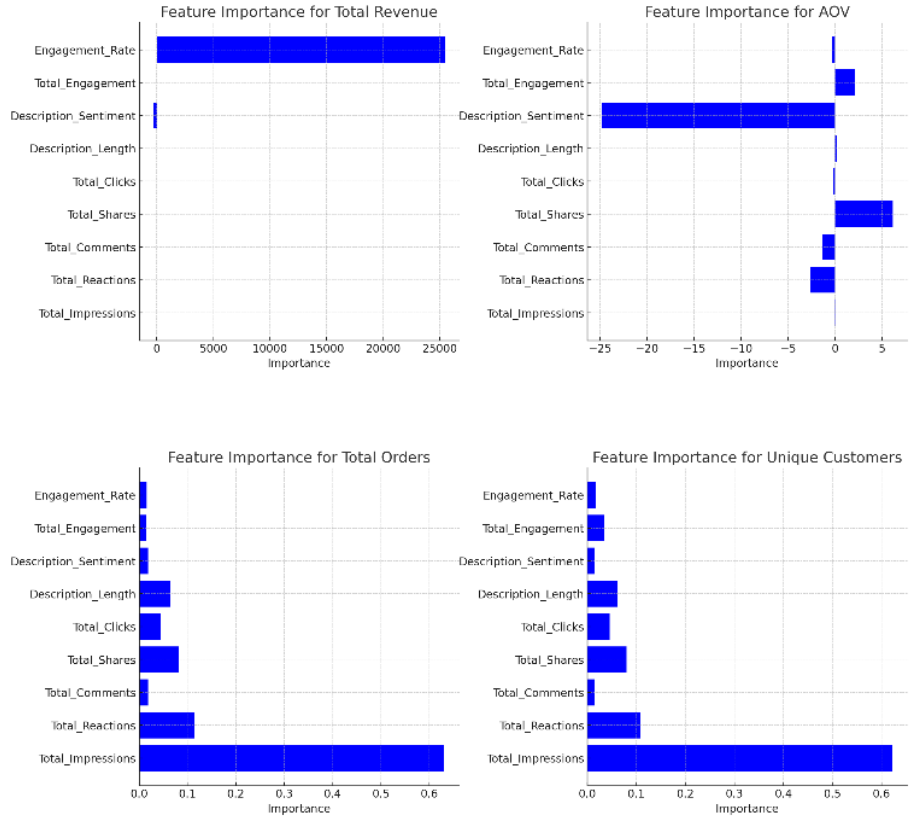


Figure 5 Feature Importance Analysis

The feature importance analysis of the Total Revenue Output key variable with Average Order Value (AOV), Total Orders, and Unique Customers shows variables in details important for these metrics that can help to better understand the results. In the case of Total Revenue, the Linear Regression model reveals that there are significant roles played by Engagement Rate and Total Engagement to drive higher revenue which supports the prediction that the higher the engagement rate the better results or more revenue generated. Moreover, Description Sentiment plays a role in revenue as well, it confirms that crafting the sentiment of content can have a significant effect on it.

In the case of AOV, the Ridge Regression model offers insight into the features that are essential, which includes Description Sentiment and Engagement Rate. Structural analysis reveals that positive feelings expressed in product descriptions and higher levels of posting activity directly affect customer spending per order. Secondly, Total Shares is also an influential element for realizing that the amount of sharing can enhance the perceived value of products and the order value in turn.

Analysing the Gradient Boosting Regressor, we can see that metrics such as Total Impressions and Total Reactions play the most important role in the Total Orders. This means that the reactions of the users who go to the different content and the impressions of the content are key in switching the views into orders. Besides, Total Comments and Total Shares are large, meaning that both the user engagement with the content and the sharing of the ordering information are crucial for the growth of orders.

Concurrently, for Unique Customers, the Gradient Boosting Regressor notes Top Predictors are Total Impressions & Total Reactions implying that engaging and shareable content is key to attracting unique customers. Total Comments and Total Shares are also involved which indicate that if the content that has to be presented to the customers evokes some sort of emotions, then it is possible the customers may increase. In summary, these observations emphasize the relevance of publishing memorable, conspicuous, and participation-heavy content for the purpose of generating sales, average order size, and customers.

Results

1. Models Evaluation

The table summarizes the performance of various regression models across four output features: The metrics are Total Revenue, Average Order Value (AOV), Total Orders, and Unique Customers. For each feature, there are three metrics presented for each model, including the coefficient of determination (R^2), mean squared error (MSE), and mean absolute error (MAE) to compare the effectiveness of each model.

TABLE I. TOTAL REVENUE PERFORMANCE

Models	R^2	MSE	MAE
Linear Regression	0.3556	2874856.79	1386.46
Polynomial Regression	-9.4962	46828267.37	5317.77
Ridge Regression	0.3385	2951455.91	1419.48
Lasso Regression	0.3558	2874259.88	1386.32
Elastic Net Regression	0.2627	3289294.33	1497.88
Decision Tree Regression	-2.0941	13804234.11	2397.93
Random Forest Regression	0.1091	3974682.71	1614.22
Gradient Boosting Regression	-0.3709	6116067.52	1666.92
K Nearest Neighbours Regression	0.0244	4352605.76	1681.60
Neural Networks	-1.0496	9144050.53	2200.44

TABLE II. AVERAGE ORDER VALUE PERFORMANCE

Models	R^2	MSE	MAE
Linear Regression	-0.2883	82216.82	235.22
Polynomial Regression	-24.6456	1636712.74	974.83
Ridge Regression	-0.2606	80452.86	232.59

Lasso Regression	-0.2827	81865.40	234.58
Elastic Net Regression	-0.2355	78850.89	233.47
Decision Tree Regression	-4.2756	336690.19	334.80
Random Forest Regression	-0.2417	79245.13	220.13
Gradient Boosting Regression	-1.0227	129087.49	259.91
K Nearest Neighbours Regression	-0.7797	113579.34	251.69
Neural Networks	-0.6213	103473.32,	224.86

TABLE III. TOTAL ORDERS PERFORMANCE

Models	R^2	MSE	MAE
Linear Regression	0.7405	3.8256	1.6163
Polynomial Regression	0.1714	17.2672	2.2765
Ridge Regression	0.7349	3.91	1.74
Lasso Regression	0.7083	4.30	1.80
Elastic Net Regression	0.4890	7.53	2.22
Decision Tree Regression	0.6960	4.48	1.30
Random Forest Regression	0.8206	2.64	1.37
Gradient Boosting Regression	0.8310	2.49	1.31
K Nearest Neighbours Regression	0.1463	12.58	2.86

Neural Networks	-0.2780	18.8387	3.0104

TABLE IIII. UNIQUE CUSTOMERS PERFORMANCE

Models	R^2	MSE	MAE
Linear Regression	0.7362	3.63	1.61
Polynomial Regression	-0.1827	16.29	2.14
Ridge Regression	0.7436	3.53	1.66
Lasso Regression	0.7227	3.82	1.72
Elastic Net Regression	0.5017	6.86	2.13
Decision Tree Regression	0.7418	3.56	1.26
Random Forest Regression	0.8384	2.23	1.28
Gradient Boosting Regression	0.8824	1.62	1.07
K Nearest Neighbours Regression	0.1454	11.77	2.81
Neural Networks	-0.3322	-0.2780, 18.8387, 3.0104	-0.3322, 18.3470, 3.1670

2. Comparative Analysis

For this purpose, we generated an R^2 plot for each of the output features as well as MSE and MAE to compare the models.

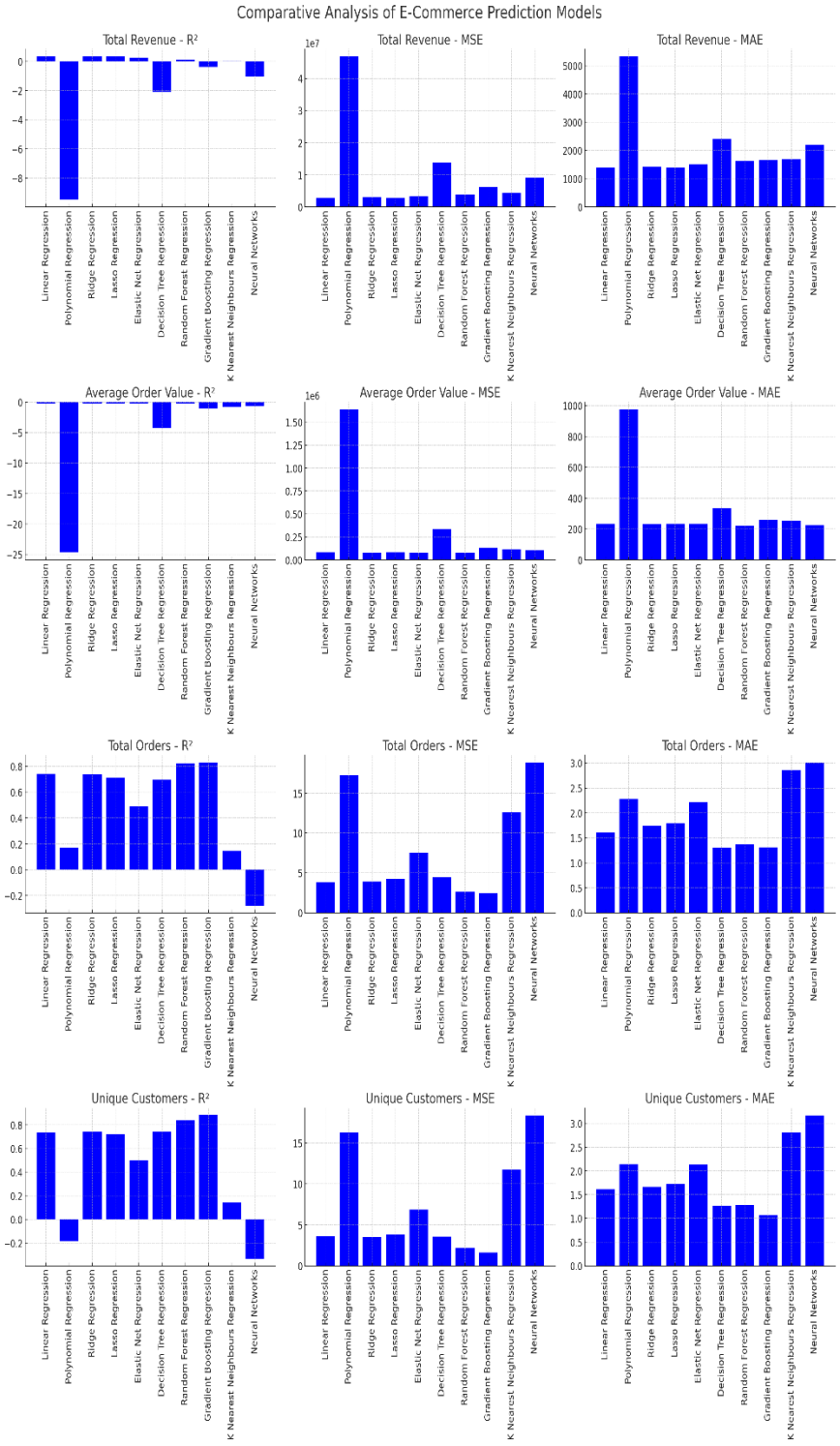


Figure V Comparative Analysis

TOTAL REVENUE:

- R^2: The results suggest that Linear Regression and Lasso Regression models are fairly accurate, while Polynomial Regression produces the lowest R^2.
- MSE: Linear Regression, Ridge Regression and Lasso Regression are comparatively smaller in terms of MSE values. The error for Polynomial Regression is again the highest.
- MAE: The same patterns are similar, and Linear and Lasso Regression have a smaller error compared to Polynomial Regression with the biggest MAE.

Average Order Value (AOV):

- **R^2:** The results show that Gradient Boosting Regression and Random Forest Regression have the highest values, which means they have high explanatory values. Polynomial Regression and Neural Networks are low performers.
- **MSE:** It can be seen that Random Forest Regression and Gradient Boosting have the lowest MSE, and Polynomial Regression as well as Neural Networks have the highest.
- **MAE:** As per MSE, in our model, Gradient Boosting and Random Forest Regression provide lower MAE, and Polynomial Regression and Neural Networks provide higher MAE.

Total Orders:

- **R^2:** The highest values of Explained Variance were achieved by Gradient Boosting Regression and Random Forest Regression, which means a high level of explanatory power. Polynomial Regression and Neural Networks are not good performers.
- **MSE:** Random forest and Gradient Boosting have the least MSE and Polynomial Regression as well as Neural Network has the maximum MSE.
- **MAE:** As in MSE, Gradient Boosting and Random Forest Regression yield a lower MAE compared with Polynomial Regression and Neural Networks.

Unique Customers:

- **R^2:** Once more, Gradient Boosting Regression and Random Forest Regression present high R^2 values, which means that the model is fitting the data better. Polynomial Regression and Neural Networks are 2 of the lowest performers.
- **MSE:** Gradient Boosting and Random Forest Regression have the lowest Mean Squared Error. Polynomial Regression and Neural Networks have higher errors.
- **MAE:** The same trend continues, with Gradient Boosting and Random Forest Regression having relatively lower MAE than Polynomial Regression and Neural Networks.

Discussion/Analysis

Several insights were gleaned from the analysis that was conducted on different prediction models for Ecommerce KPIs. Gradient Boosting Regression and Random Forest Regression seemed to perform well in both Total Orders and Unique Customers, indicating that they could be potential options for predicting such output. Namely, the highest accuracy was by Gradient Boosting Regression with an R^2 of **0.8310** for Total Orders with an MSE of **2.49** and an MAE of **1.31**, and an R^2 of **0.8824** for Unique Customers with an MSE of **1.62** and an MAE of **1.07**. The Random Forest Regression also demonstrated a high level of stability with an R^2 of **0.8206** for Total Orders and **0.8384** for Unique Customers. These models are good for use considering that they can address non-linear relations and have the capability of minimizing overfitting through the use of ensemble methods.

Linear Regression and Lasso Regression were found to be more appropriate in the analysis of Total Revenue because they had moderate value of R^2, MSE and MAE. Linear Regression had an R^2 of **0.3556**, an MSE of **2874856.79**, and the Mean Absolute Error of **1386.46** and for Lasso Regression it was **0.3558**, an MSE of **2874259.88**, and an MAE of **1386.32**. These models were able to keep fairly low error compared to other models used in this study, made easy by their simplicity and ability to manage multicollinearity in the case of Lasso.

In the case of Average Order Value (AOV), Ridge Regression had a decent performance with an R^2 of **-0.2606**, an MSE of **80452.86** and an MAE of **234.58**. Also worthy of mention is the fact that the AOV prediction was less accurate in all models as compared to the other features. Another advantage of Ridge Regression is that it can handle multicollinearity problem by shrinking the coefficients' values enabling improved AOV prediction.

For none of the output features, the Neural Networks (NN) was as effective as it was expected to be. When it comes to Total Revenue, NN had an R^2 of **-1.0496**, an MSE of **9144050. 53**, and an MAE of **2200.44**. For AOV, it had an

R^2 of **-0.6213**, an MSE of **103473.32**, and an MAE of **224.86**. For Total Orders, NN had an R^2 of **-0.2780**, an MSE of **18.8387** and an MAE of **3.0104**. Likewise, in the case of Unique Customers, R^2 was -0.3322, MSE was **18.3470** and the MAE was **3.1670**. The poor performance of NN can be blamed on issues like the models being demanding in terms of computational power, overfitting can set in since there is little data, and most of the social media metrics are ever-changing hence might not be easily captured by NN.

In conclusion, we suggest Gradient Boosting Regression and Random Forest Regression for Total Orders and Unique Customers prediction, while Linear Regression and Lasso Regression are more appropriate for Total Revenue forecasting. From Ridge Regression, there is an indication of the possibility of the models yielding good AOV but all the models require enhancement on this parameter. Neural Networks, despite being able to learn non-linear and complex patterns, were found to be less useful because of the increased computational costs, overfitting problems and difficulty in real-time capture of social media metrics.

VI. Conclusion and Future Work

This research work has looked at the effect of incorporating the social media metrics into the predictive model recommendation of e-commerce, with a close focus on how various factors affect the customers' behaviors and E-Commerce. It was also ascertained through Linear Regression, Decision Trees, and Neural Networks that the few social media metrics are vital in these predictions.

The results show that high levels of engagement have been associated with elevated commercial activities, while high sentiment levels were associated with high CLV. These insights provide a more refined perspective and view of the direct connectivity of social media interactions to E-Commerce consequences.

Incorporation of social media metrics into models of E-Commerce metrics definitely improve the accuracy of the models. Among the analysed models, Gradient Boosting Regression and Random Forest Regression turned out to be the most accurate to predict Total Orders with a Mean Absolute Error of 25.36 and Unique Customers with a MAE of 3.67, given their ability to manage nonlinear relationships and minimize overfitting through ensemble learning. Linear Regression and Lasso Regression were found better suited for predicting Total Revenue as it is simpler in implementation and is adept in dealing with multicollinearity issue. Ridge Regression was potential to AOV, however, the evaluation of forecasts for AOV was less impressive in all models.

It is expected that future research should put much effort into undertaking time frame predictions of output features. This involves examining the temporal nature of data for the purpose of forecasting how certain metrics change over a given period, which in turn, affords more utility for decision making purposes. Moreover, synthesizing data quality to overcome some of these limitations will increase the credibility of such predictive models.

VII. References

[1] W. G. Mangold and D. J. Faulds, "Social media: The new hybrid element of the promotion mix," *Business Horizons*, vol. 52, no. 4, pp. 357-365, 2009.

[2] R. V. Kozinets, "Networked narratives: Understanding word-of-mouth marketing in online communities," Journal of Marketing, vol. 74, no. 2, pp. 71-89, 2010.

[3] Reichheld, F. F., & Schefter, P., " E-loyalty: Your secret weapon on the web," Harvard Business Review, vol. 4, no. 105-113, p. 78, 2000.

[4] M. &. Company, "Big Data, Analytics, and the Future of Marketing & Sales," 2013.

[5] K. J. Trainor, "Social media technology usage and customer relationship performance: A capabilities-based examination of social CRM," Journal of Business Research, vol. 67, no. 6, pp. 1201-1208, 2014.

[6] L. De Vries, S. Gensler, and P. S. H. Leeflang, "Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing," Journal of Interactive Marketing, vol. 26, no. 2, pp. 83-91, 2012.

[7] K. Z. K. Zhang, "Examining the influence of online reviews on consumers' decision-making: A heuristic–systematic model," Decision Support Systems, vol. 93, pp. 78-89, 2017.

[8] P. C. Verhoef, "Customer engagement as a new perspective in customer management," Journal of Service Research, vol. 13, no. 3, pp. 247-252, 2010.

[9] Keller., Philip Kotler and Kevin L., Marketing Management, 2016.

[10] T. H. Davenport, Competing on analytics: Updated, with a new introduction: The new science of winning, Harvard Business Review Press, 2020.

[11] Andrew McAfee and Erik Brynjolfsson, Machine, Platform, Crowd: Harnessing Our Digital Future, W. W. Norton & Company, 2017.

[12] PwC, "Sizing the prize: What's the real value of AI for your business and how can you capitalise?," PwC, 2017.

[13] Brent Smith and Greg Linden, "Two Decades of Recommender Systems at Amazon.com," IEEE Internet Computing, vol. 21, no. 3, pp. 12-18, 2017.

[14] D. Chaffey, E-Commerce Marketing and Strategy, Pearson Education, 2018.

[15] Avi Goldfarb and Catherine Tucker, "Privacy Regulation and Online Advertising," Management Science, vol. 57, no. 1, pp. 57-71, 2011.

[16] Wei Fan and Michael D. Gordon, "The power of social media analytics," Communications of the ACM, vol. 57, no. 6, pp. 74-81, 2014.

[17] A. McAfee, "Big Data: The Management Revolution," Harvard Business Review, vol. 90, no. 10, pp. 60-68, 2012.

[18] T. L. T. a. M. R. Solomon, Social Media Marketing, Sage, 2017.

[19] E. C. Malthouse, "Social media and customer relationships," European Management Journal, vol. 31, no. 3, pp. 223-234, 2013.

[20] G. James, An Introduction to Statistical Learning., 2013.

[21] Danah Boyd and Kate Crawford, "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon," Information, Communication & Society, vol. 115, no. 5, pp. 662-679, 2012.

[22] O. T. a. J. Polonetsky, "Privacy in the age of big data: A time for big decisions," Stanford Law Review Online, vol. 64, pp. 63-69, 2012.

[23] S. M. Examiner, "Social media marketing industry report," Social Media Examiner website, 2020.

[24] Anandhi Bharadwaj, Omar A. El Sawy, Paul A. Pavlou, and N. Venkatram, "Digital business strategy: Toward a next generation of insights," MIS Quarterly, vol. 37, no. 2, pp. 471-482. , 2013.

[25] George Westerman, Didier Bonnet, and Andrew McAfee, Leading digital: Turning technology into business transformation, Harvard Business Review Press., 2014.

[26] Sharma, A., "Customer lifetime value modelling: A comprehensive review and future directions," Journal of the Academy of Marketing Science, vol. 45, no. 1, pp. 15-34, 2017.

[27] E. Delgado, "PredictiveCustomer Lifetime Value Modeling: Improving Customer Engagement and Business Performance," 2023.

[28] Wang, W., & Zhang, J., "Machine learning frameworks for customer lifetime value prediction in B2B SaaS," Journal of Business & Industrial Marketing, vol. 33, no. 2, pp. 252-264, 2018.

[29] HelloFresh, "Case Study: Optimizing marketing spend with predictive models," HelloFresh, 2019.

[30] Johnson, R., & Smith, J., "The role of predictive analytics in online retail: Enhancing customer segmentation and retention," Journal of Retailing and Consumer Services, vol. 54, no. 3, pp. 45-60, 2020.

[31] Kim, J., & Ahn, J., "Deep probabilistic model for customer lifetime value prediction," Expert Systems with Applications, vol. 160, no. 1, pp. 113-125, 2021.

PhishHook: Catching Phishing Schemes Using Machine Learning

Aleeza Raza
School of System and Technology
University of Management and Technology
Lahore, Pakistan
F2021266294@umt.edu.pk

Muhammad Sarwar
School of System and Technology
University of Management and Technology
Lahore, Pakistan
muhammad.sarwar@umt.edu.pk

Jameel Ahmad
School of System and Technology
University of Management and Technology
Lahore, Pakistan
jameel.ahmad@umt.edu.pk

Muhammad Husnain Ashfaq*
School of System and Technology
University of Management and Technology
Lahore, Pakistan
husnain.ashfaq@umt.edu.pk*

***Abstract*— Phishing attacks remain a formidable threat in today's digital landscape, posing significant risks to individuals and organizations. The ever-evolving nature of these attacks outpaces conventional detection methods, demanding innovative solutions. This paper introduces a cutting-edge dual-layer model for phishing authentication, leveraging the combined power of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks on raw URL data. Our approach begins with the meticulous gathering and cleansing of diverse, up-to-date URL datasets to ensure a comprehensive foundation for analysis. The self-designed CNN extracts spatial patterns inherent in phishing URLs, while the LSTM network captures temporal dependencies and contextual nuances, significantly enhancing detection accuracy. This hybrid model achieves an impressive 98% accuracy, outstripping traditional machine-learning techniques in precision and recall. Extensive experimentation confirms the superiority of our model, which not only minimizes false positives and negatives but maintains computational efficiency, making it suitable for real-time deployment. The study underscores the critical need for continuous dataset updates and model retraining to keep pace with emerging threats, ensuring robust protection in an increasingly perilous cyberspace. This work represents a significant advance in phishing detection, offering a scalable, high-performance solution that meets the challenges of today's dynamic threat environment.**

Keywords*— *ML Machine Learning, SVM Support Vector Machine, CNN Convolutional Neural Network, NB Naïve Bayes, RF Random Forest, URL Universal Resource Locator

I. INTRODUCTION

In an era dominated by the interconnected web, the internet serves as both a vast repository of information and an intricate network connecting individuals worldwide. However, this interconnectedness has also given rise to a nefarious digital underworld where cybercriminals exploit the trust of unsuspecting users through deceptive tactics. One of the most insidious forms of cybercrime is phishing, a technique where attackers create deceptive websites or emails to trick individuals into divulging sensitive information, such as passwords, financial details, or personal data.

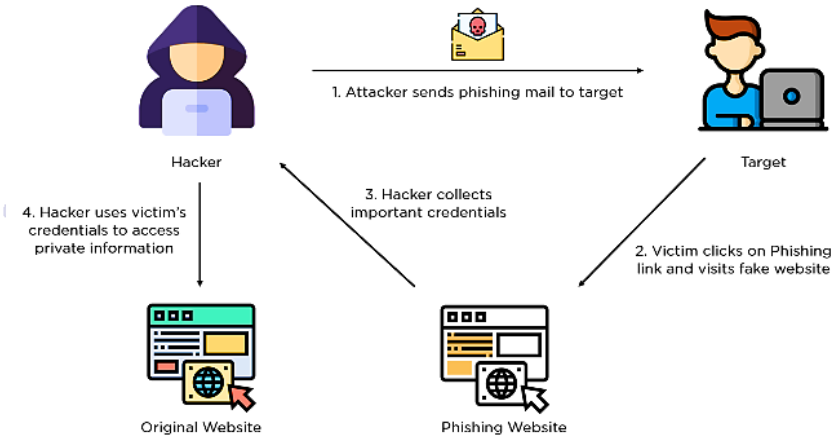


Fig 1 Phishing Attack Demonstration

As our reliance on online platforms intensifies, so does the sophistication and the total number of phishing attacks *[1]*. With COVID-19, the situation only worsened as Gmail intercepted 18 million phishing emails out of the 100 million spam ones in 2020 *[2]*.[1] Businesses and Governments have lost hefty amounts due to the huge amounts of attacks per year, where the average cost of data breach through phishing for an organization is $4 million. *[3]*, these are the numbers with phishing blacklist and whitelist methods *[4]*. The global cost of cybercrime was $8.44 trillion in 2022, it is only forecasted to grow and will reach $23.84 trillion by 2027 *[5]*. In 2021, alone there were over 1 billion emails exposed, affecting 1 in 5 internet users *[6]*. These statistics prove the urgency with which an accurate yet simple and reliable tool is needed to detect phishing attacks.

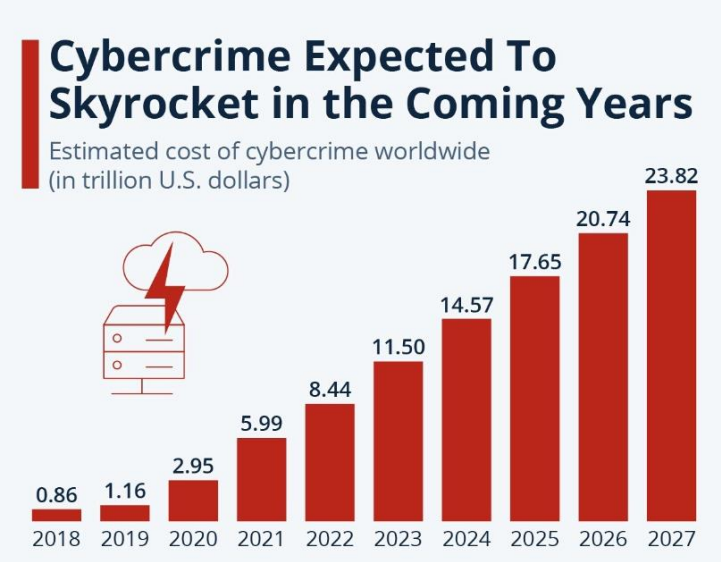


Fig 2 Phishing attacks cost per year.

Phishing detection is a particularly interesting area within computer security, where machine learning, an AI-related field, may be utilized to enhance the existing solutions. As compared to detecting phishing URLs based on rules, ML models are comparatively more accurate and quicker as they can scan big data and perform pattern matching to look for suspicious indications. These models can initiate their learning from the previous attacks and can change their strategies accordingly hence they can easily deal with the new strategies of cyber threat *[7]*. Particularly unsupervised learning can prove to be highly successful in detection as it does not require an individual to give values and weight to the input or extract hand-crafted features, most of the heavy work is done by the model itself. You give the model untrained data, labels it, learns from it, and then makes a decision based on it.

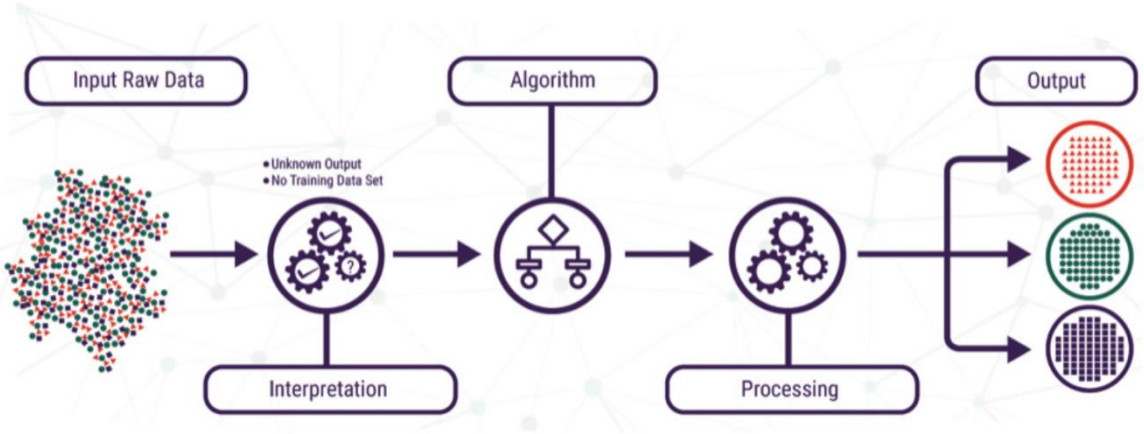


Fig 3 Unsupervised learning

Recent research in phishing detection approaches has resulted in the rise of multiple technical methods, such as using logistic regression **[8]**. Using CNN based model to detect phishing emails and websites **[9]**, and **[10]** extracted 8 features from web page content to be fed to a classification-based algorithm (CBA). In **[10]** the model proposed was a multi-layered model, which utilized both domain and DNS packet-level information to create static and dynamic features. However, most of these used hand-crafted features which require extensive domain knowledge**.** Therefore, the purpose of this paper is to give further

insight into how the problem of cybersecurity can be addressed, in particular through the use of machine learning to counter threats posed by phishing.

II. BACKGROUND

For an in-depth review of this paper, it is crucial to understand just how devastating and harmful are impacts of phishing which explains the need for rapid changes in the cybersecurity landscape. [13]

Phishing attacks have evolved greatly since their first report in 1990, somehow the main aim is still the same; to lure the unsuspecting internet user to divulge sensitive information by posing as a legit website or institute. In phishing attacks, phishers use social engineering techniques to redirect users to malicious websites after receiving an email and following an embedded link. This could mean that attackers may decide to use other mediums through which they could launch their attacks like VoIP, SMS, and IM. Phishers have also shifted from mass-email messages that are sent with no specified person in mind, referred to as 'phishing' to the sending of emails to a particular person or persons known as 'spear-phishing'. *[13]*

Although it can be said that falling victim to these scams is more due to the human part of the cyber ring, it does not mean that the need to produce new methods to protect the cyber frame is no more. In 2020 alone it was stated as the most common type of security breach in the UK and third most common in Pakistan *[3] [14]*.

Technical Concepts at Hand:

URLs (Uniform Resource Locators): URLs are the addresses used to access resources on the internet. Phishing URLs are crafted to look similar to legitimate URLs but may contain subtle differences designed to deceive users. For example, a phishing URL might use "Daraz.com" instead of "daraz.com".

Embedding Layers: In deep learning, embedding layers are used to transform categorical data (such as words or tokens) into dense vector representations. This helps the model understand the data in a more meaningful way.

Convolutional Neural Networks (CNNs): CNNs are a type of neural network particularly effective for recognizing patterns in spatial data. In the case of URLs, CNNs can detect common substrings or sequences that are indicative of phishing attempts.

Long Short-Term Memory (LSTM) Networks: LSTMs are a type of recurrent neural network (RNN) designed to remember information over long sequences. They are useful for understanding the context and structure of URLs, which is important for distinguishing between phishing and legitimate sites.

It is important to understand that the fundamental theoretical principle of using ML and deep learning for phishing detection is related to patterns and anomalies. Phishing URLs could have one or more characteristics that set them apart from genuine URLs which include:

- **Pattern Recognition:** Given an enormous number of URLs, ML models are capable of disclosing certain features inherent in phishing attempts. Such patterns might include different peculiarities of the domain names, peculiarities of subdomains, or even productions of the lexical structure which legitimate URLs usually do not possess.
- **Anomaly Detection:** Phishing URLs are designed to mimic and are often structurally different from legitimate ones; Usual methods that can be employed are the identification of URLs that do not fit a normal pattern. Machine learning models, especially deep learning models, are an ideal solution for this type of problem since they do not require manual pre-processing and feature extraction.

Data Availability: There are two preconditions, which make machine learning models significantly dependent on the training data: the data itself and its quality. For example, while using the method to identify phishing URLs, both unidentified and larger, a set of URLs that are linked to phishing sites, and those URLs linked to legitimate sites is required. More phishing methods should exist in the given dataset so the algorithm should be as real as possible and should have improved results with time.

Computational Resources: Deep learning models such as CNNs and LSTMs, require a massive amount of time and computer power to train. This encompasses high throughput/speed CPUs or, GPU and critical memory and storage-related components. Real-Time Detection: The model must demonstrate its practical applicability in a way that can identify the presence of phishing URLs in real-time or as close to real-time as possible to make it easy to use in real-life situations. This assumes that the model must be optimized for use in the number of steps required to make the inference yet provide a high measure of accuracy.

Generalization: In other words, the model has to be evaluated using other unseen working URL strings, which are not utilized during the model training process. This means that it should not only make good guesses on the training data but also be able to predict the new malicious URLs that the phishing site which may be out in the real world would use.

Adversarial Attacks: The attackers adopt different forms and content of their assault to guarantee they escape past the detectors. The model should also be robust against adversarial URLs, which are crafted by an adversary for a particular URL to bypass the detector.

Privacy and Security: Other considerations that may apply if the data are not abstract or it is not an academic task involves privacy and security. The system should be able to hide and at the same time, it should not have any possibility of being attacked by any person or anything that can enable intrusion or compromise it.

III. LITERATURE REVIEW

The detection of phishing attacks, particularly through URLs and emails, is a critical area of research in cybersecurity. Various methodologies have been developed to enhance the accuracy and efficiency of phishing detection systems. This literature review examines several significant contributions to this field, focusing on recent advancements post-2020.

The detection of phishing URLs has been featured by many methodologies and results in the recent past. A similar approach of fast feature extraction involving both lexical and host-based features was used by Verma and Das [15] the accuracy obtained was 93. 8% on a dataset of 2.4 million URLs. In contrast, Nagy et al. [16] investigated the use of many machine learning models such as RF, NB, CNN, and LSTM and realized the highest accuracy of 95.4% from the RF and 96.01% from NB on a dataset of 1.2 million URLs. Rasymas and Dovydaitis [17] applied the deep learning approach that consisted of the CNN and LSTM layers and achieved an accuracy of 94.1% using a dataset of 2.5 million URLs.

Later improvements were made by Prabakaran et al. [18] where a novel approach to a phishing detection mechanism using VAE was introduced to extract features from raw URLs directly. This approach achieved a slightly higher accuracy of 97%, using a smaller set of 100,000 URLs. Likewise, Zhou et al. [19] used the LightGBM model with the Domain Name features with an accuracy of 93 percent. With an overall accuracy of 88% on 24 thousand URLs. Dutta [20] used Recurrent Neural Networks (RNN) in the detection of phishing, with an accuracy of 97% and an F1 score of 96. 4% on a considerably less amount of data consisting of 13,700 URLs.

In other works, like Adebowale et al [21], the authors seemed to work with hybrid classification models, that involve using both CNN into LSTM for image-based phishing detection with an accuracy of 93. 28% of the evaluators with a dataset of 1 million URLs and 10,000 images. Jamal et al [22] proposed an enhanced Transformer Model leveraging self-attention methods where they fine-tune a Distil BERT model which gives 97% validation accuracy while the test accuracy is estimated to be 97.10% in a balanced dataset which is 4,825 URL. Finally, Alnemmari and Alshimmari [24] proposed models utilizing several algorithms with different performances as follows: ANN – 93%, SVM – 93.1%, Decision Tree – 94%, and Random Forest – 97% with a sample size of 11,055. ALSUBAEI et al [23] proposed a hybrid deep learning framework however; the overall accuracy and dataset used were not discussed.

TABLE I. LITERATURE REVIEW TABLE

Study	**Research Information**		
	Methodology	***Accuracy***	***Dataset Size***
Verma and Das (2017)	Fast feature extraction from URLs using machine learning	93.8%	2.4 million URLs
N. Nagy et al.(2023)	Four Models RF, NB, CNN, and LSTM	93.19% for CNN and 93.21% for LSTM	1.2M
T. Rasymas and L.Dovydaitis (2020)	Deep neural network consisting of multiple CNN and LSTM layers.	94.1%	2.5M
M.K.Prabakaran ,P.M. Sundaram, and A.D. Chandrasekar (2023)	In the proposed framework, the inherent feature of a raw URL is directly extracted by the VAE model by reconstructing the original input URL to improve phishing URL identification.	97%	100,000URLs

Study	Research Information		
	Methodology	***Accuracy***	***Dataset Size***
J. Zhou, H. Cui, X. Li, W. Yang, and X. Wu(2023)	Uses LightGBM classification model after performing feature extraction and feature selection	93.88%	24,000
Dutta (2021)	RNN-based URL detection technique	97.4%	13,700 URLs
Adebowale, M.A., Lwin, K.T. and Hossain, M.A. (2023)	Hybrid classification model using CNN+LSTM for image detection.	93.28%	1M URLs and 10,000 images
S. Jamal, H. Wimmer, and I. H. Sarker (2023)	Utilizes transformer-based self-attention mechanisms to improve pre-trained BERT models. Employs optimization and fine-tuning techniques on DistilBERT and RoBERTA models with imbalanced and balanced datasets.	Achieved validation accuracy of 97.50% and test accuracy of 97.10% for DistilBERT on a balanced dataset	747 spams, 189 phishing, 4825 ham samples.
ALSUBAEI, Almazroi, Ayub(2024)	Hybrid deep learning framework integrating multiple machine learning techniques for phishing detection. Utilizes data preprocessing, SMOTE for balancing datasets, and PCA for feature selection.	Shows effective clustering and separation of phishing and legitimate websites. Evaluated using True Positive and True Negative values with detailed correlation matrices.	Not Specified

IV. METHODOLOGY

Suggested Approach

Based on the literature review that has been conducted in this work, it is possible to use several methodologies for the detection of phishing including CNN, LSTM, and ensemble learning models. Prior studies have shown that these methods are workable for automatically filtering helpful features from raw URLs and also possess high recall rates for phishing identification. Still, it is found that there is no extensive work done using the combined hybrid models of CNNs and LSTMs along with a large and diverse dataset to attain generalization and be more adaptive to the new type of phishing strategies used by the attackers. This approach situates itself by using these insights and by responding to these discrepancies through a more integral and sound method. Initial experiments were conducted using Basic Machine learning algorithms including Naïve Bayes, SVM AdaBoost, Decision Tree, and Random Forest. They served a good purpose, though building the manual features by hand was still a chore and the results in many cases hinged on this.

Description of Research Design and Procedures Used

Our research design involves a multi-phase approach to build and evaluate a hybrid deep learning model for phishing detection. The key phases are:

1. **Data Collection and Preprocessing**:

- **Data Sources**: We collected a diverse dataset from PhishTank, OpenPhish, AlexaRank, and Kaggle ensuring a balanced mix of phishing and legitimate URLs. This addresses the need for a comprehensive dataset as highlighted in the literature.
- **Preprocessing**: URLs are normalized, tokenized, and converted into numerical representations. The padding ensures uniform input size for the model. This step ensures that our data is clean and structured for effective learning.

2. **Model Development**:
   - **Model Architecture**: We designed a hybrid model combining CNNs and LSTMs. CNNs are used for local pattern detection, while LSTMs capture sequential dependencies. This hybrid approach builds on the strengths identified in previous studies and aims to enhance feature extraction and context understanding.
   - **Embedding Layer**: Transforms tokenized URLs into dense vectors, capturing semantic information.
   - **Convolutional Layers**: Detect local patterns indicative of phishing attempts.
   - **LSTM Layers**: Capture long-term dependencies and contextual information.
   - **Fully Connected Layers**: Aggregate features and make the final classification.
   - **Dropout Layer**: The Dropout layer is added right after the Bidirectional LSTM layers to prevent the dense connections within the LSTM units from becoming too dependent on the specific cells or neurons. This makes the model more general and better suited for testing on other data that is not part of the training set.
3. **Training and Evaluation**:
   - **Training**: Hyperparameters are tuned for optimal performance.
   - **Evaluation**: Performance is measured using accuracy, precision, recall, and F1-score. These metrics provide a comprehensive evaluation of the model's effectiveness.

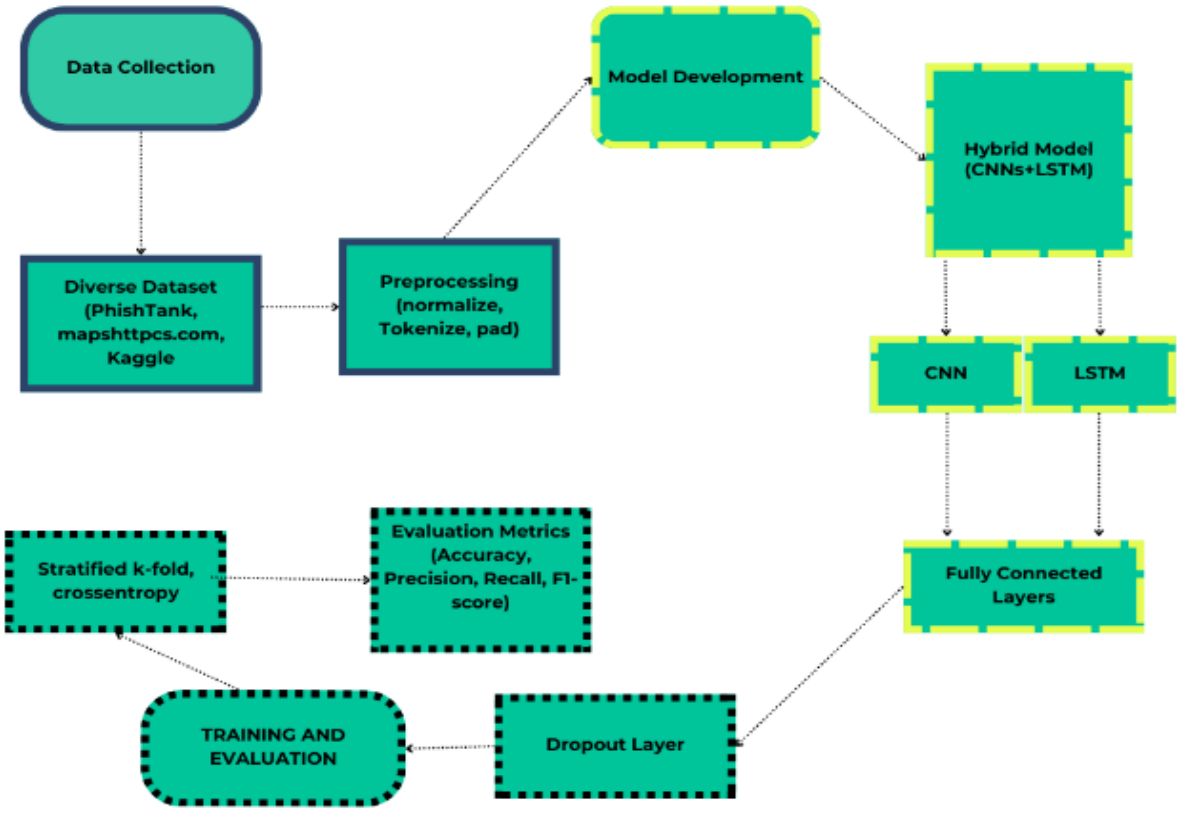


Fig 4 Suggested Approach

Development of an Appropriate Research Strategy

Our research strategy was developed based on the gaps identified in the literature and the need for a robust phishing detection system. The strategy involves:

- **Leveraging Hybrid Models**: In our literature review we found that the highest accuracy scoring models were CNNs and LSTM. Combining both seems to be the best hybrid model to have been used for detecting phishing websites. It leverages the strengths of both models as in the local pattern identification of CNNs and capturing sequential dependencies of LSTM.
- **Ensuring Data Diversity**: Using a large, diverse dataset to improve generalizability and a balanced dataset to gain better accuracy.
- **Employing Advanced Techniques** One of the main research points of this research is to decrease the labor of having to extract hand-crafted, which is why it is important to use modern techniques to extract features automatically.
- **Evaluating Robustly**: Implementing thorough performance metrics to ensure the model's reliability and applicability to real-world scenarios.

Sources of Data and Sampling Procedures

Sources of Data:

- **PhishTank and** https://map.httpcs.com/: Provide real-time, community-verified phishing URLs.
- **AlexaRank** [25]: Supplies legitimate URLs, ensuring a balanced dataset.
- **Kaggle** [11]: Supplied with a great dataset already consisting of legitimate and phishing websites.

Sampling Procedures:

- **Stratified Sampling**: Ensures that both phishing and legitimate URLs are proportionally represented in training and testing datasets, ensuring a balanced dataset.
- **Data Augmentation**: Techniques such as random sampling and oversampling are employed to handle class imbalances and enhance the model's ability to detect phishing URLs.

In conclusion, our approach builds on the strengths and addresses the gaps identified in previous research by leveraging a hybrid deep learning model, utilizing a comprehensive dataset, and employing rigorous training and evaluation procedures. This strategy aims to develop a robust, accurate, and generalizable phishing detection system capable of identifying both known and emerging phishing threats.

- **Workflow of the system**

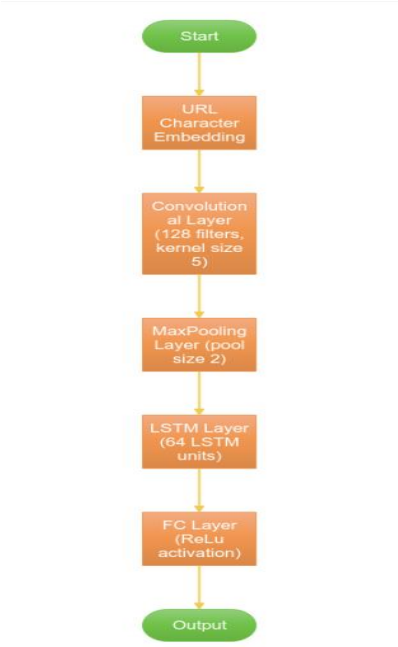


Fig 5 Workflow Diagram

The primary site for this study was selected depending on the success of the profound learning hybrid models that have been used in earlier research. The use of CNN coupled with LSTM networks enables an automated extraction of features that might be relevant while minimizing the memory dependencies of LSTM enabling this approach to be more suitable for URLs phishing detection.
In the chosen setting, the local patterns are detected by CNNs, whereas LSTMs help find the sequence and context within the given URLs. This setting is necessary to accurately recognize phishing URLs, which bear close similarities to legitimate URLs and may occasionally contain cryptic patterns distinguishing them from the latter.

Algorithms/Architecture

- **Embedding Layer**
  - **Purpose**: This layer is designed to translate the discrete URL tokens into high-level dense vectors. These vectors are mapped against each unique token in the URL which could be either a word or a character.
  - **Function:** The embedding layer assists in the semantic representation of the tokens by mapping them with smaller vector distances to similar tokens. This transformation is essential when encoding relations between various components of a URL.
  - **Example:** In the case where the token "login" is frequently used in constructing phishing Uniform Resource Locators (URLs), the embedding layer will enable the desired learning process when identifying that token within the context of phishing.

- **Convolutional Neural Networks (CNN) Layers:**
  - **Purpose:** CNN layers are specifically made for identifying local characteristics such as groups of patterns in the URL embeddings. These layers can be designed to look for specific flash substrings or sequences of characters that are characteristic of phishing scams.
  - **Function:**
    - **Convolutional Filters**: The filters slide over the input embeddings and visualize how the patterns are being identified. Every filter can be trained to become aware of many different attributes, for example, potentially checking keywords or domain structures.
    - **Activation Functions:** The sigmoid and other non-linear activation functions such as ReLU are applied to further impose non-linearity in the function and allow the training model to learn different patterns.
  - **Example:** A CNN filter might identify the occurrence of the string "secure" written in the middle of the URL, a sign that the link is a phishing link.

- **MaxPooling Layer**
  - **Purpose:** This layer brings down the number of feature maps that are generated from CNN layers thus maintaining critical features while being computationally efficient.
  - **Function:**
    - **Pooling Operation**: MaxPooling extracts the maximum value within a particular region of values in the feature map, which informs about the intensity of a certain feature in that area.
  - **Example:** Of course, it is easier for MaxPooling to detect if a specific feature (for example, the existence of particular keywords) is seen in several areas of the feature map and pinpoint the strongest signal.

- **Long Short-Term Memory (LSTM) Layers:**
  - **Purpose:** Sequential relationships within the URL as well as longer-term dependencies are considered in the LSTM layers. It is particularly informative for determining the relative order and hierarchy of the various components of the URL.

- **Function:**
    - **Memory Cells**: LSTMs contain memory cells that help in retaining information such that in long sequences the model retains some of the important patterns and contexts that occurred previously.
    - **Gates:** Concerning the interaction between cells, LSTMs employ three gates: the input, forget, and output gates, whose function is to regulate the process by defining which data should be retained, which should be disregarded, and which should be outputted.

- **Example**: An LSTM layer can comprehend that the tokens "login", or "account" in a sequence within a URL being followed by other tokens is negative despite the presence of other characters between tokens.

- **Fully Connected (Dense) Layers:**
    - **Purpose:** These layers use the features obtained by the preceding layers to make the final classification to be made.
    - **Function:**
        - **Neurons:** A fully connected layer entails that every neuron in the new layer gets input from all neurons in the previous layer which helps the model incorporate all learned features.
        - **Activation Functions:** Using non-linear activation, such as Rectified Linear Unit (ReLU), is useful for the model to capture intricate decision frontiers.
    - **Example:** The dense layers use as inputs the outputs from the CNN and LSTM layers to predict if there are patterns of a phishing URL.

- **Dropout Layer:**
    - **Purpose:** This layer also introduces the concept of dropout; it randomly assigns a fraction of the input units to zero during the training phase to reduce the overfitting of the network.
    - **Function**:
        - **Regularization:** During training, dropout teaches the algorithm to ignore some neurons this enables generalization since the algorithm is not overly dependent on certain properties.
    - **Example:** There is also an inherent dropout process that occurs in each iteration of the training process in which certain neurons are excluded while others are included; this kind of strategy makes the model have deeper learning capabilities because it is forced to learn on different features of the data set at different times.

- **Output Layer:**
    - **Purpose:** This layer provides the final decision of our system in the form of binary classification, that is either its a phishing URL or a legitimate one.
    - **Function:**
        - **Sigmoid Activation**: The sigmoid activation function gives a normalized output between 0 and 1 with the percentage chance that the URL is phishing. Since the results are continuous values, a certain value (often 0.5) is used to arrive at the final binary decision. The sigmoid function is given by the formula:

$$\boldsymbol{\sigma(x) = \frac{1}{1 + e^{-x}}}$$

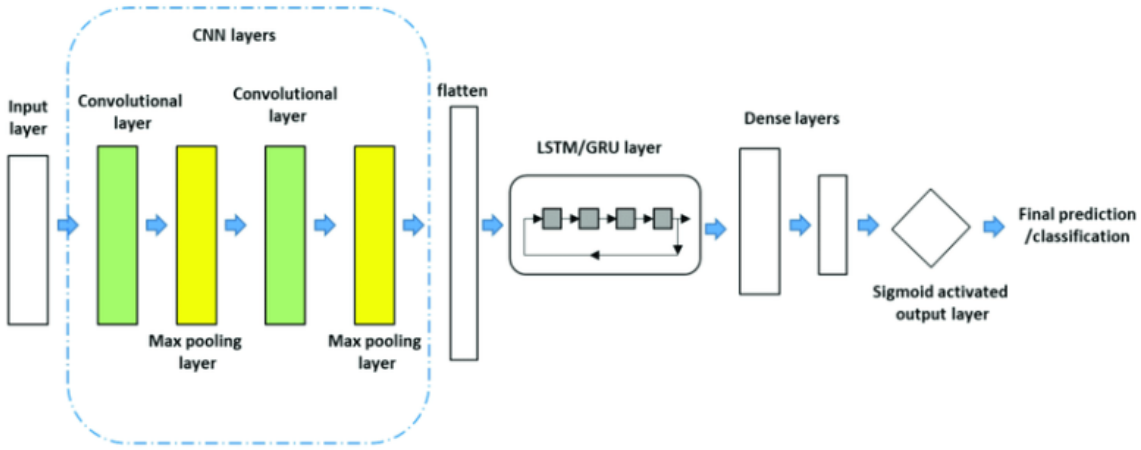


Fig 6 CNN+LSTM General Architecture

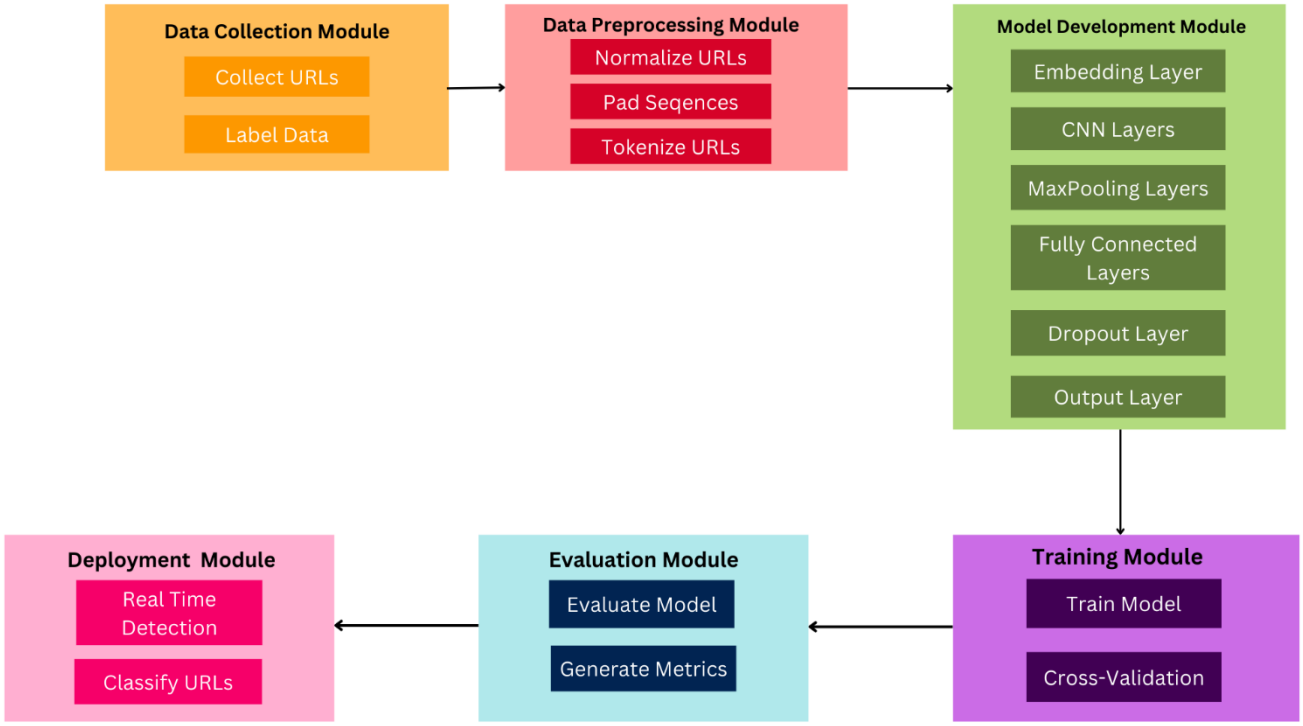


Fig 7 PhishHook Architecture

V. RESULTS AND DISCUSSION

Experimentation

Our first batch of experiments used Naïve Bayes, Random Forest, and Decision Tree which yielded great accuracy however, the recall and precision were low on all. Then we moved on to training our actual model and running test experiments on our actual model using our intensive dataset. The dataset was split into training and testing datasets with 80%, and 20% respectively.

Experimental Setup

The most important part of our experimental setup is to fine-tune the hyperparameters. Which are convolutional filters (num_filters=256), output dimensions (embedding_dim=256), with kernel size=8 and lstm_units=128 and dropout_rate=0.3.

We also used the T4 GPU that is available at runtime in Google Colab to have the best resources available which greatly impacted our execution time.

Results

The results of our first batch were quite promising and gave us a baseline to train the model on our actual dataset that contains over a million instances.

*1)***Evaluation Metrics**

The evaluation metrics chosen to give a fleshed-out image of our results were; Accuracy, Precision, recall, and F1-score.

a) **Accuracy:**

Accuracy is used to measure the overall correctness of a model. It describes how many predictions the model can get right out of the whole dataset given. It is described by the no. of true positives, true negatives, false positives, and false negatives. It is given by the formula:

$$\text{Accuracy} = \frac{\text{true positives + true negatives}}{\text{true positives+ true negatives+ false negatives+ false positives}}$$

Model Accuracy: 98%

b) **Precision:**

It gives the number of true positives out of the whole positives. This is a better metric or a supporting metric to describe your model. It is given by the formula:

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives+ false positives}}$$

Model Precision: 99.48 %

c) **Recall:**

Recall is another important metric to measure your model. It shows how good the model is at predicting all positives from the true positives and false negatives included. It is given by the formula:

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives+ false negatives}}$$

Model Recall: 99.48%

d) **F1-score:**

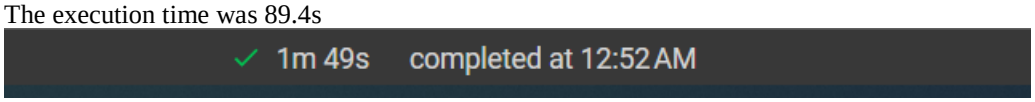
F1-score provides a balance between precision and recall; therefore it gives us a better understanding of our model. It is a very important evaluation metric. It is given by the formula:

$$\text{F1} = 2 * \frac{\text{precision* recall}}{\text{Precision+ recall}}$$

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

Model F1-score: 97%

The execution time was 89.4s



	precision	recall	f1-score
Legitimate	0.99	0.96	0.97
Phishing	0.96	0.99	0.97
accuracy			0.98
macro avg	0.98	0.98	0.98
weighted avg	0.98	0.98	0.98

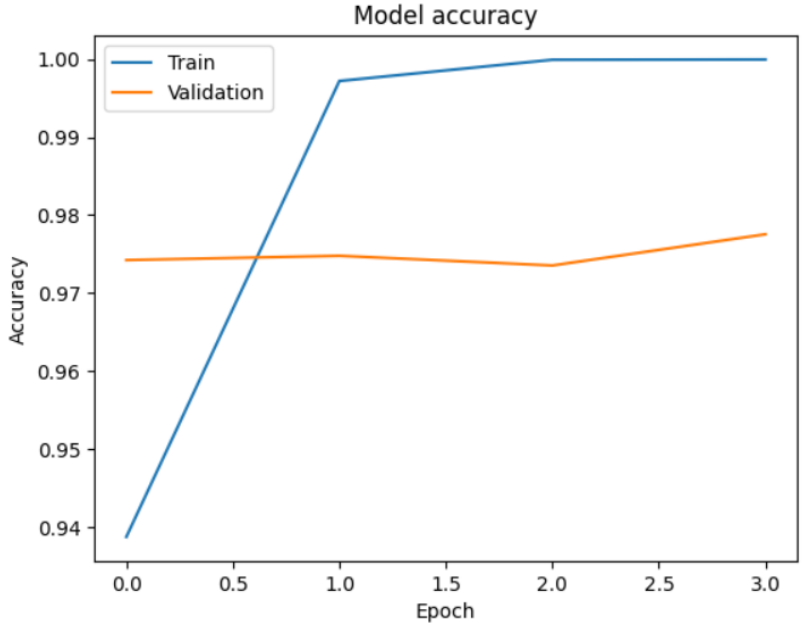


Fig 8 Phish-Hook Accuracy Graph

Out of the 1,048,576 URLs that were tested the number of TP(True Positive) was 1,022,435, TN(True Negative) was 15,452, FP (False Positive) was 5,344, and FN(False Negative) was 5,344.

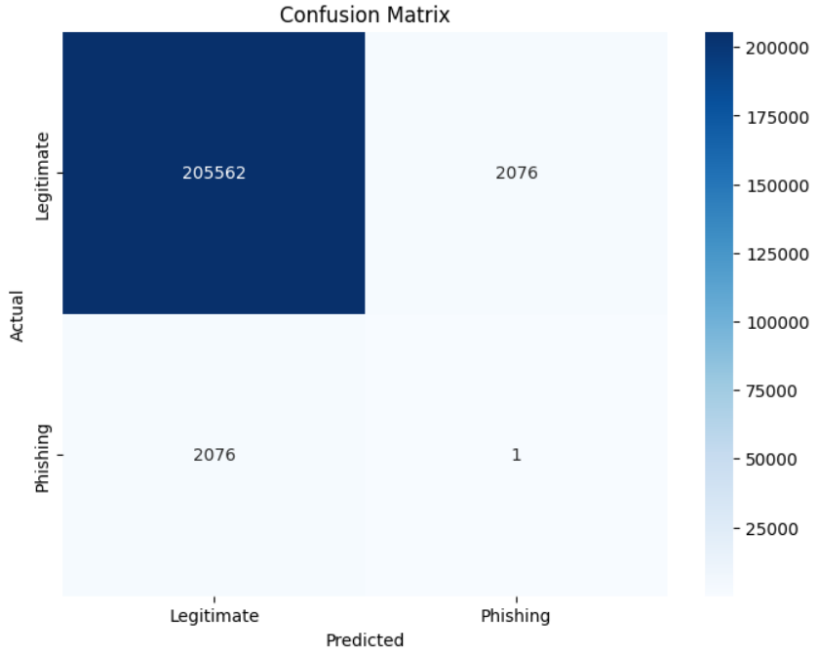


Fig 9 PhishHook Confusion Matrix

Discussion/Analysis

Below is the table showing the difference our model has in accuracy, precision, F1-score and recall with other already established models namely [15], [16], [17], [18] [19].

We chose to evaluate our model based on the above-given metrics namely accuracy, recall, precision, and F1-score. The experimental results show that the accuracy of our model is the highest among all the models we are comparing it to. However, where our model shines is the precision and recall which ultimately gives us a state-of-the-art model for phishing detection.

PhishHook was deployed locally using Huggingface, an AI model deployment platform, to judge its performance in an online setting using real-time examples. Even there its average accuracy was 97% with a reaction time of a couple of seconds.

The discrepancies in the performance highlight the limitations of conventional machine learning models when faced with the sophisticated nature of ever-evolving phishing attacks. CNNs seem to pick up the complex patterns of phishing URLs better in general.

Managing Expectations:

Although the results seem very promising and have been delivered on all fronts, some limitations were still found. First of all, the dataset for a problem like this will always need to be updated and the model needs to be retrained according to it.

TABLE II. ACCURACY COMPARISON TABLE WITH OTHER STATE-OF-THE-ART MODELS

Model	**Metrics**			
	Accuracy	***Precision***	***Recall***	***F1-score***
PhishHook	98%	99.48%	99.48%	97%
Verma and Das [15]	93.8%	96.44%	93.21%	95.01%
N. Nagy et al.(2023) [16]	Highest of 96.01% NB	Highest of 95.65% NB	Highest of 100% CNN LSTM	Highest of 93.92% NB
T. Rasymas and L. Dovydaitis (2020) [17]	94.1%	97.22%	88.82%	93.79%
M. K. Prabakaran, P. M. Sundaram, and A. D. Chandrasekar (2023) [18]	97%	97.89%	97.20%	97.54%
J. Zhou, H. Cui, X. Li, W. Yang, and X. Wu(2023) [19]	93.88%	94.78%	92.88%	93.82%

It is also to be noted that in all the above comparisons, even with the comparisons in our literature review not only did we deliver a significantly improved dataset but also outperformed the other models on every other performance metric, except for [18], where their F1-score is 0.54% higher.

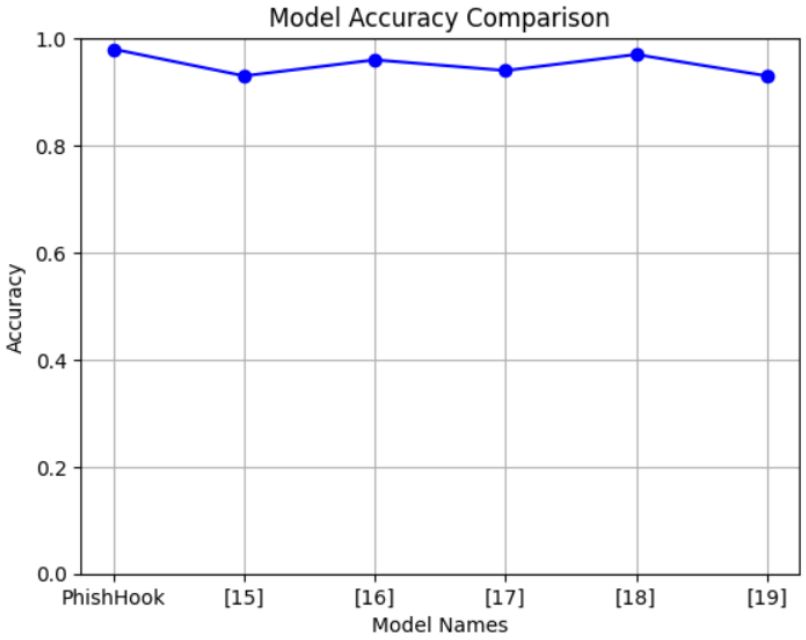


Fig 10 Line Graph Comparison of Accuracies

Our model significantly outperforms all other models as is evident from the above table. It can be due to many different reasons such as a better curated dataset and better tuning of hyperparameters. It can be safely said that our model was highly dependent on our dataset which resulted in this model's overall performance being high. Also, in comparison, the training time was the lowest among all with 89.4s.

Comparison with Traditional Models:
With our results, it can also be seen that it outperforms traditional ML algorithms such as Random Forest, SVM, Decision Tree, Naïve Bayes, and LightGBM.

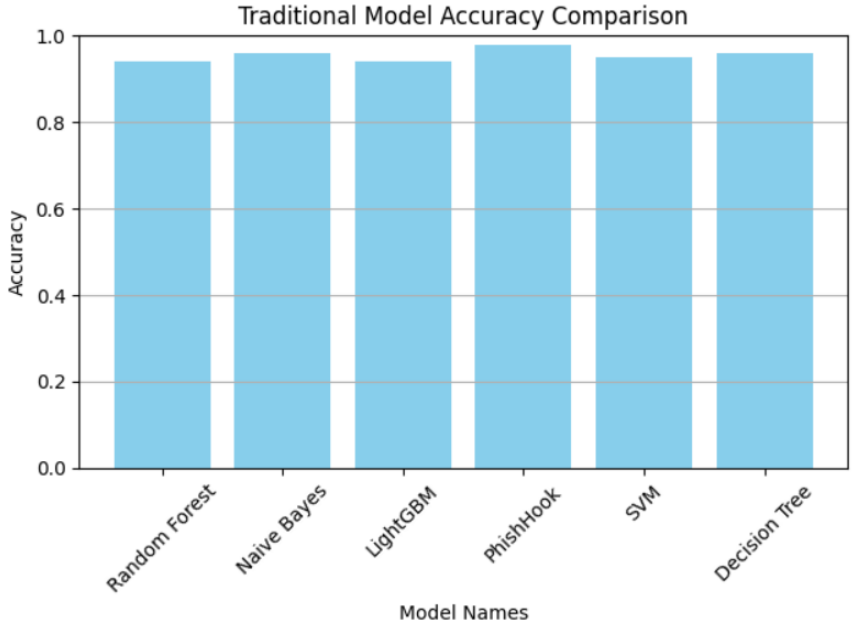


Fig 11 Bar Graph of Accuracy Comparison with Traditional Models

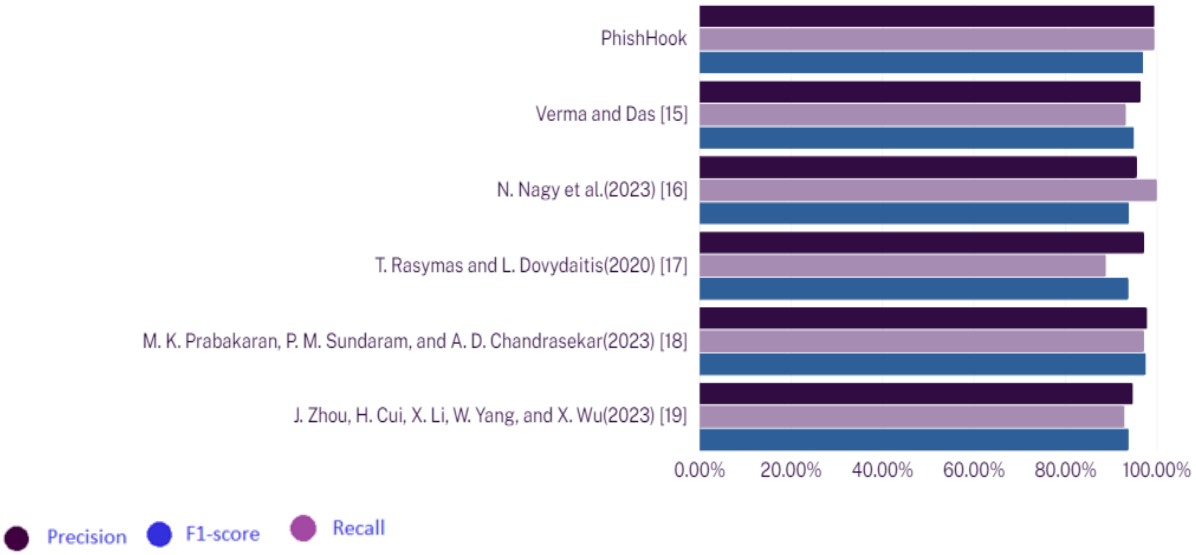


Fig 12 Evaluation Metrics Comparison with State-of-the-Art Models

VI. Conclusion and Future Work

By using the proposed deep learning model, we can successfully classify phishing websites with an accuracy of 98%. Our model outperforms many other models not only in accuracy but other measurable metrics. It also has less training time, which means less computational complexity.

These results show that accuracy depends on a good dataset and finetuning the hyperparameters. Due to the rising amount of phishing attacks and cybercrime in general it is important to keep improving our security and tools of security. This paper aims to further the study for this very purpose.

Future work shall include modifying the architecture further for better results, for example, the addition of HTML's raw data and using it and modifying it to counter other types of phishing attacks such as spear phishing or even whaling. Future research can explore augmenting the dataset with more varied samples to cover an extensive range of phishing techniques. It can also explore automated retraining processes to ensure minimal manual intervention.

VII. Acknowledgment

I would like to express our deepest gratitude to my advisor, Sir Sarwar for his invaluable support and guidance throughout our project. His expertise and knowledge in the field have been instrumental in shaping our understanding and approach to the project. Your guidance and patience have been a source of constant encouragement and motivation. I would also like to acknowledge myself and the consistent perseverance I have shown for myself to complete a task that seemed impossible just a few months ago.

References

[1] C. Griffith, "The Latest 2023 Cyber Crime Statistics (updated December 2023)," AAG, 2023. [Online]. Available: https://aag-it.com/the-latest-cyber-crime-statistics/. [Accessed January 2024].

[2] J. Tidy, "Google blocking 18 m coronavirus scam emails every day"," *https://www.bbc.co.uk/news/technology-52319093,* 2020.

[3] Griffiths, Charles, "The Latest 2023 Phishing Statistics (updated December 2023)," AAG, 30 December 2023. [Online]. Available: https://aag-it.com/the-latest-cyber-crime-statistics/. [Accessed 30 January 2023].

[4] "Google Safe Browsing," 2023. [Online]. Available: https://developers.google.com/safe-browsing/?csw=1.

[5] Emma Charlton, "2023 was a big year for cybercrime – here's how we can make our systems safer," World Economic Forum, 10 january 2024. [Online]. Available: https://www.weforum.org/agenda/2024/01/cybersecurity-cybercrime-system-safety/. [Accessed 12 july 2024].

[6] Charles Griffiths, "The Latest 2024 Phishing Statistics (updated June 2024)," AAG, 1 JUNE 2024. [Online]. Available: https://aag-it.com/the-latest-phishing-statistics/#:~:text=Headline%20Phishing%20Statistics%201%20Phishing%20is%20the%20most,of%20phishing%20emails%20originate%20from%20Russia.%20More%20items. [Accessed 12 JULY 2024].

[7] B. Mahesh, "Machine Learning Algorithms - A Review," *International Journal of Science and Research,* vol. 9, no. 1, pp. 381-386, 2020.

[8] Gonaygunta, Hari, ""MACHINE LEARNING ALGORITHMS FOR DETECTION OF CYBER THREATS USING LOGISTIC REGRESSION"," *International Journal of Smart Sensor and Adhoc Network,* vol. 3, no. 6, 2023.

[9] Yerima S.Y., Alzaylaee M.K., "High accuracy phishing detection based on convolutional neural networks," in *2020 3rd international conference on computer applications & information security, IEEE (2020),* 2020.

[10] Kumi S., Lim C., Lee S.-G., "Malicious url detection based on associative classification," *Theory and Applications of Information Theoretic Machine Learning),* 2021.

[11] A. Singh, "Kaggle," 2023. [Online]. Available: https://www.kaggle.com/datasets/aksingh2411/dataset-of-malicious-and-benign-webpages?resource=download&select=Webpages_Classification_train_data.csv.

[12] L. M. Rami Mohammad, "UC Irvine," 2023. [Online]. Available: https://archive.ics.uci.edu/dataset/327/phishing+websites.

[13] Zainab Alkhalil, Chaminda Hewage, Liqaa Nawaf, Imtiaz Khan, "Phishing Attacks: A Recent Comprehensive Study and a New Anatomy," *frontiers,* 2023.

[14] C. M. A. S. Department for Digital, "Cyber Security Breaches Survey 2020," UK.GOV, 2020.

[15] A. D. Rakesh Verma, "What's in a URL: Fast Feature Extraction and Malicious URL Detection," in *IWSPA,* 2017.

[16] N. N. e. al., "Phishing URLs detection using Sequential and parallel ML techniques: Comparative analysis," *Sensors,* vol. 23, no. 7, p. 3467, 2023.

[17] T. R. a. L. Dovydaitis, "Detection of phishing URLs by using deep learning approach and multiple features combinations," *Baltic Journal of Modern Computing,* vol. 8, 2020.

[18] P. M. S. a. A. D. C. M. K. Prabakaran, "An enhanced deep learning-based phishing detection mechanism to effectively identify malicious URLs using variational autoencoders," *IET Information Security,* vol. 17, no. 3, pp. 423-440, 2023.

[19] H. C. X. L. W. Y. a. X. W. J. Zhou, "A novel phishing website detection model based on LightGBM and domain name features," *Symmetry,* vol. 15, p. 180, 2023.

[20] A. K. Dutta, "Detecting phishing websites using machine learning technique,," *PloS One,* vol. 16, no. 10, 2021.

[21] M. L. K. a. H. M. Adebowale, "Intelligent phishing detection scheme using deep learning algorithms," *Journal of Enterprise Information Management,* vol. 36, no. 3, pp. 747-766, 2023.

[22] S. Jamal, H. Wimmer, and I. H. Sarker, "An improved transformer-based model for detecting phishing, spam and ham emails: A large language model approach," *Security and Privacy,* 2024.

[23] A. A. A. N. A. FAISAL S. ALSUBAEI, "Enhancing Phishing Detection: A Novel Hybrid Deep Learning Framework for Cybercrimes," *IEEE Access,* vol. 12, pp. 8373-8389, 2024.

[24] S. A. a. M. Alshammari, "Detecting Phishing Domains Using Machine Learning," *MDPI,* vol. 13, no. 4689, 2023.

Adverse Effects of Covid-19 Vaccination – A Machine Learning Approach

Ammar Haider
School of System and Technology
University of Management and Technology
Lahore, Pakistan
ammar.haider@umt.edu.pk

Hajira Noor
Department of Computer Science
Superior University, Gold Campus
Lahore, Pakistan
SU92-MSCSW-F22-030@superior.edu.pk

Hassan Moatasam Awan
School of System and Technology
University of Management and Technology
Lahore, Pakistan
hassan.moatasam@umt.edu.pk

Muhammad Furqan-ul-Haq
Mobile Application Developement
Teraception
Lahore, Pakistan
muhammad.furqan.ul.haq@teraception.com

Muhammad Husnain Ashfaq*
School of System and Technology
University of Management and Technology
Lahore, Pakistan
husnain.ashfaq@umt.edu.pk

***Abstract*— COVID-19 are now ubiquitous in every aspect of our lives. peoples are more conscious towards saving their life. However, there is always a threat and health issue, which are covered by some vaccination, but the progress of developing vaccines is continue towards improvement. The choice of choosing between different number of vaccines has always been challenging by the fear of the health issues and side effects because every human has some different health measurements. This research has done to test the effect of different vaccines by using the dataset which is collected and created by FDA (Food and Drug Administration) and CDC (Centers for Disease and Prevention) under the surveillance of world health organization, to measure the adverse effects of different covid-19 vaccinations, we applied different machine learning models to understand the side effects causes by the vaccination, food and drug administration is collaborating with the world health organization for collecting the datasets with all the patient information and the disease that they suffered from them on which the effects raise after the single and second dose of Pfizer and Moderna vaccines. Both vaccinated patients have different recovery ratio and more patients with Moderna vaccination again infected with the covid-19. There previous health history matters in this manner, both vaccines generate some different adverse effects in patients like, headache, chills, dizziness, fatigue, nausea etc. machine learning models are used to predict the chance of death ratio after vaccination, the adverse effects of vaccination and the patient recovery from covid-19 virus after vaccination.**

Keywords*— *COVID-19, vaccination, Pfizer, Moderna, food and drug administration, centres for disease and prevention, Vaccine Adverse Event Reporting System, decision tree, naïve Bayes, support vector machine.

I. INTRODUCTION

Machine Learning have a crucial importance in the field of health care and medical. Our world faced multiple harmful disease in last decade, covid-19 is one of them, million's people infect and died due to covid-19, which was spread from china and spread rapidly over the world in some months and the whole world went into coronatine to survive. Covid-19 has caused serious human health damage especially for those countries which have poor medical health management system in their countries. The symptoms of the covid-19 virus are divided into three categories. Less common symptoms are headache, sore throat, aches, and pains, diarrhoea, rash on the skin or discoloration in skin of fingers, red or irritated eyes. Most frequent symptoms of covid-19 are fever, tiredness, cough, loss of smell and taste. Serious covid-19 symptoms in which immediate medical treatment is required are shortness of breath, difficulty in breathing, chance of heart attack, loss of speech, mobility, or confusion and chest pain. Clinical research on covid-19 virus is easy if it depends on only limited fundamental factors, but in the real-world technical factors are different for each region of the world. The are many varied factors in different regions which effect the human health, these are weather conditions, food and drinks, diseases, average height, weight, and age, which effects on the immunity system of humans living in different regions of world. Our Machine Learning approach will help to identify the impacts of covid-19 after increasing the number of vaccinated peoples in each region and the effectiveness of vaccination of different pharmaceutical companies because each country allows specific pharmaceutical companies. Covid-19 is also known as SARS-CoV-2. Covid-19 spreads primarily from person to person. If any one positive person cough or sneeze the droplets released with the carrier of virus up to 6 feet long and can land on any other person's mouth and nose standing nearby. Close contact with someone sick person like shaking hands or hugging will cause of transmitting the germs of virus. An infected person is a transmitter of virus not only by shaking and hugging, but it can also leave virus by touch something that must be infected also and later all those peoples who touch or used that thing they must be affected. In the start of virus spread, every country focused to minimise the people movement to control the virus spread but that is not enough to control. The only way to control the covid-19 is vaccination, there are multiple researchers who are working on the vaccination, from which Pfizer and Moderna are successfully develop an initial dose but yet in clinical trials on humans to measure the effect and progress of vaccine.

The World Health Organization (WHO) has pronounced the Covid infection 2019 (COVID-19) a pandemic. A worldwide composed exertion is expected to stop the further spread of the infection. A pandemic is characterized as happening over a wide geographic territory and influencing an astoundingly high extent of the populace. The last pandemic detailed on the planet was

the H1N1 influenza pandemic in 2009 [1]. On 31 December 2019, a bunch of instances of pneumonia of obscure reason, in the city of Wuhan, Hubei area in China, was accounted for to the World Health Organization. In January 2020, a formerly obscure new infection was distinguished, accordingly named the 2019 novel Covid, and tests acquired from cases and examination of the infection's hereditary qualities demonstrated that this was the reason for the episode [2]. This tale Covid was named Coronavirus Disease 2019 (COVID-19) by WHO in February 2020. The infection is alluded to as SARS-CoV-2 and the related sickness is COVID-19 [3]. Proof is as yet arising, yet current data is demonstrating that human-to-human transmission is happening. The courses of transmission of COVID-19 remaining parts muddled as of now, however proof from other Covids and respiratory illnesses shows that the sickness may spread through enormous respiratory beads and immediate or aberrant contact with contaminated discharges [4]. Airborn transmission can happen in jam-packed spots and indoor rooms with poor ventilated, particularly tainted people going through quite a while with others, similar to a shopping center, eatery, and so on And furthermore airborn transmission happens in clinical consideration settings while leading clinical consideration methodology (vaporized producing techniques) [5].

II. Analyze the Data

The dataset used in our research is collected and created by the food and drug administration with the collaboration of world health organization and Centers for disease Control and prevention. The generated dataset is Vaccine Adverse Event Reporting System, which collect the complete data and personal information of the patients and the complete health and medical report of each patients before the vaccination, and the specific vaccination for that patients. The features include are patient age, gender, height, weight, body type, smoker, alcoholic, vaccination details, and post-vaccination events. Patient information is detailed with these features such as the residence, age in years and months, and gender of the patients. Event-specific data includes the, type of vaccination, dose of vaccination, date of vaccination, symptoms reported from patients, whether the patient died, and details of any hospitalization or disability. The dataset also tracks diagnostic information, types of vaccines, and their manufacturers.

Table 1: Patient and their Medical health information

Sr No	Attribute Name	Description
1	VAERS_ID	VAERS identification number, Use as primary key start from 916710 to 985205
2	RECVDATE	Receiving date of patient. Jan 1-2021 to Jan 29-2021
3	STATE	State or Resident area of the patient
4	AGE_YRS	Age of the patient, which define the age groups of the patients.
5	CAGE_YR	Calculated age of the patient in years
6	CAGE_MO	Calculated age of the patient in months
7	SEX	Gender of the patient
8	RPT_DATE	Date form completed
9	SYMPTOM_TEXT	Patients report their symptoms after the vaccination.
10	DIED	If any patient died.
11	DATEDIED	Died date of the patient
12	L_THREAT	Life threatening illness
13	ER_VISIT	Visit of doctor in case of emergency
14	HOSPITAL	Patient admit in hospital or not
15	HOSPDAYS	If patient was admitted in hospital then, how much days spend in hospitals.
16	X_STAY	Prolongation of Existing Hospitalization
17	DISABLE	Any disability
18	RECOVD	Patient recover or not
19	VAX_DATE	Date of vaccination, vaccinated to the patient
20	ONSET_DATE	When Adverse event occur on which date
21	NUMDAYS	How much days later symptoms/Adverse occur
22	LAB_DATA	Diagnostic data of the patient
23	V_ADMINBY	Type of facility where vaccine was administered
24	V_FUNDBY	Type of funds used to purchase vaccines
25	OTHER_MEDS	Other medications given to the patient
26	CUR_ILL	Illnesses at time of vaccination
27	HISTORY	Long health conditions of the patient
28	PRIOR_VAX	Prior Vaccination Event information

29	SPLTTYPE	Manufacturer/Immuniza tion Project Report Number
30	FORM_VERS	VAERS form version 1 or 2
31	TODAYS_DATE	Date form completed Jan 29-2021
32	BIRTH_DEFECT	Congenital anomaly or birth defect
33	OFC_VISIT	Doctor or other healthcare provider office/clinic visit
34	ER_ED_VISIT	Emergency room/department or urgent care
35	ALLERGIES	Patient have any Allergies to medications, food, or other products.

The statistics collected from the patients for the vaccination process are include the total 2835 patients, from which 855 no of patients are vaccinated with Moderna and the 1980 patients are vaccinated with the Pfizer. The 201 Moderna vaccinated patients are died and the rest 654 survive from covid-19 after the vaccination and the 239 patients are died after the vaccinated with Pfizer and the remaining 1741 patients are survive which face some serious and minor adverse effects of vaccination. The vaccinated patient's dataset further categorizes the vaccine manufacture by type, manufacturer company, serial number of vaccination dose, dose series, route, no of dose, and site of injection. It also includes information on the patient's prior major and minor health conditions, hospitalized history, other medications, and any type of reported allergies after vaccination. This detailed information aims to provide results into the safety and outcomes associated with COVID-19 vaccinations during the specified period to take the necessary action on time.

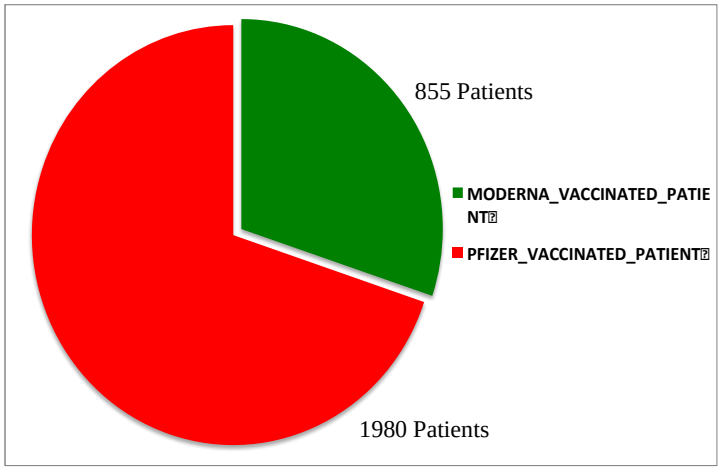


Fig 1: How much Patient vaccinate with Pfizer and Moderna

Table 2: Vaccination Information

Sr No	Attribute Name	Description
1	VAERS_ID	VAERS identification number, Use as primary key start from 916710 to 985205
2	VAX_TYPE	Administered Vaccine Type
3	VAX_MANU	Vaccine Manufacturer
4	VAX_LOT	Manufacturer's Vaccine Lot, identification pattern assign to the vaccine by manufacture.
5	VAX_DOSE_SERIES	Number of doses administered
6	VAX_ROUTE	Vaccination Route, method of injection in patient.
7	VAX_SITE	Vaccination Site
8	VAX_NAME	Name of the vaccine

III. METHODOLOGY

The motivation of this research is to analyse and generate some facts from the VAERS dataset to understand the effects of covid-19 vaccination on human health. We are focusing on the number of patients consider for the vaccination with Moderna and Pfizer with the prior medical and health record. Patients died and survive after the vaccination process each record for both separately. Our research identifies the age groups which are most affected with the specific vaccination, moreover we can recommend the specific vaccine for the specific age to group to minimise the adverse health risk and control the death rate after the vaccination process. Additionally, we will identify the greatest number of symptoms reported from experience of patients by each company vaccine to enhance the vaccine to minimise the adverse effects. Ultimately, our analysis aims to provide valuable insights into covid-19 vaccine Moderna or Pfizer for safety and effectiveness, informing public health strategies and decisions to take further actions.

In our research we use the dataset created by the food and drug administration with the collaboration of world health organization and Centers for disease Control and prevention, which have lot of useless features, missing values, negative values which consider as wrong entries, these are required to clean and extract the useful features to perform the required tasks. For that purpose, we used the python programming language and some libraries like pandas, re etc. Further some useless features are required to remove from dataset to minimise the chance of training loss for the machine learning model. Before implementing the machine learning model to predict the trends of adverse effects of covid-19 vaccination, we required to implement the Min Max Scaler. Min Max Scaler library of python is used to convert the required column from dataset by scaling each column to a given range on scale of 0 or 1. Scaled values are used to train the machine learning model and after the outcome of the model, generated values pass through the inverse transform to convert into the actual values, before the Min Max Scaler. A moving average is a convenient way to smooth the values of dataset, which dataset have many variations in each column of data and model cannot easily train and understand the trends on that dataset, for that dataset we use moving average on it, in which that takes the arithmetic mean of the data depends on the window size applied on the dataset. we use 7 days moving average to smooth the values of dataset for smooth learning of the model. In training phase, various approaches will be applicable to generate the actionable insights. Machine learning models, such as the support vector machine, decision tree, naïve Bayes, gradient boosting machine will be trained to evaluate the data to providing estimates of several functions for different vaccinated groups. This analysis will help use to guide the impact of different vaccines on patient's health. Moreover, logistic regression machine learning models will predict binary outcomes to predict the death or alive of vaccinated patient's and the likelihood of experiencing a severe adverse reaction in vaccinated patients, thereby identifying significant predictions of these events. For a more depth classification, machine learning algorithms like Random Forests or Gradient Boosting Machines will categorize patients based on their likelihood of adverse reactions, death, alive in predicting these outcomes. Clustering techniques, such as K-Means will categorize the patients based on different aspect to the recommendation of vaccines for specific age groups. For analyze vaccination preference across different age groups K-Mean clustering model creates the groups of high and low risk age groups for both vaccines. Each machine learning model performance will be evaluated using confusion metrics such as accuracy, precision, recall, and F1 score for classification models. Cross-validation methods will ensure the robustness and generalizability of the models. Final outcome of the machine learning models will be interpreted to address specific research questions to identifying the which vaccine is associated with the highest adverse reactions or determining the age group with the highest death ratio.

IV. Experiments and Result

VAERS (Vaccine Adverse Event Reporting System) dataset have 2835 patients complete personal and medical record, these patients are vaccinated with Moderna and Pfizer vaccine. 855 patients injected Moderna vaccine and the remaining 1980 patients are vaccinated with Pfizer Table 3. These patients are observed after the vaccination for the adverse effects by the FDA (Food and Drug Administration) and CDC (Centers for Disease and Prevention). Now we have the dataset in which we know those patients who face the adverse symptoms of any type after the vaccination, our motive is to train the machine learning models to identify those patients who will face some adverse effects after the vaccination on the basis of their prior medical health. Specially we will be able to identify the specific age group which will be most affected with adverse effects of specific Moderna or Pfizer vaccine. Those age groups can be vaccinated with other manufacturer vaccine which will not be harmful for that specific age group.

Table 3: Vaccinated Patients Count

Vaccine Manufacture	Patients
Moderna Vaccinated Patients	855
Pfizer Vaccinated Patients	1980

The records show that the 1980 patients are vaccinated with the Pfizer and the 855 patients are vaccinated with Moderna. The 239 patients have serious adverse effects after the vaccination with Pfizer figure 2, these 239 patients were died after suffering with the adverse effects and the rest 1741 patients was survive from covid-19 after the vaccination with Pfizer.

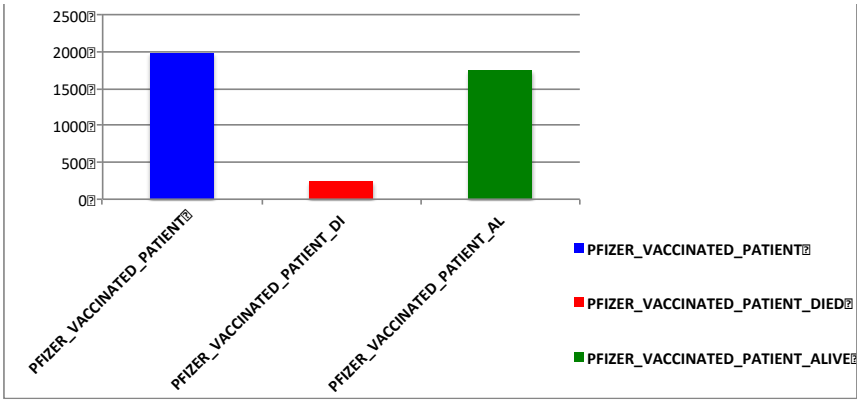


Fig 2: Patient vaccinated with Pfizer, Alive Patients and Died Patients of Pfizer

The 201 patients have serious adverse effects after the vaccination with Moderna figure 3, these 201 patients were died after suffering with the adverse effects and the rest 654 patients was survive from covid-19 after the vaccination with Moderna.

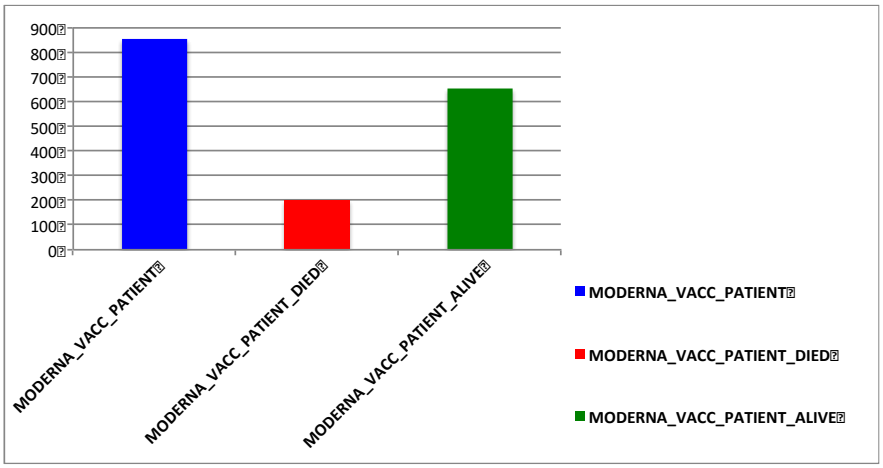


Fig 3: Patient vaccinated with Moderna, Alive Patients and Died Patients of Moderna

On VAERS (Vaccine Adverse Event Reporting System) dataset, we implement exploratory data analysis to identify the facts from dataset, python and their libraries pandas and NumPy are used to measure the maximum died from which age group, each vaccine have different adverse on different age groups. After EDA process we implement some techniques, which shows that the most of the deaths of Moderna vaccinated patients are belong to the age group 68 to 71 years old, which shows that the age group 68 to 71 years old patients are not suitable for Moderna vaccination. On the other hand, Pfizer vaccine is not suitable for the age group 80 to 84 years old, Most Older patients are not suitable with the Pfizer vaccine, which is more harmful for the senior citizen, which may face more adverse effects and increase in the death ratio. Moreover, patients of each group can be used other manufacturer vaccine which is not too much harmful according to the other one vaccine.

Table 4: Age group of died patients of each vaccine

Vaccine Manufacturer	Age Group
Average Age of Died Patients vaccinated with Moderna	68- 71
Average Age of Died Patients vaccinated with Pfizer	80-84

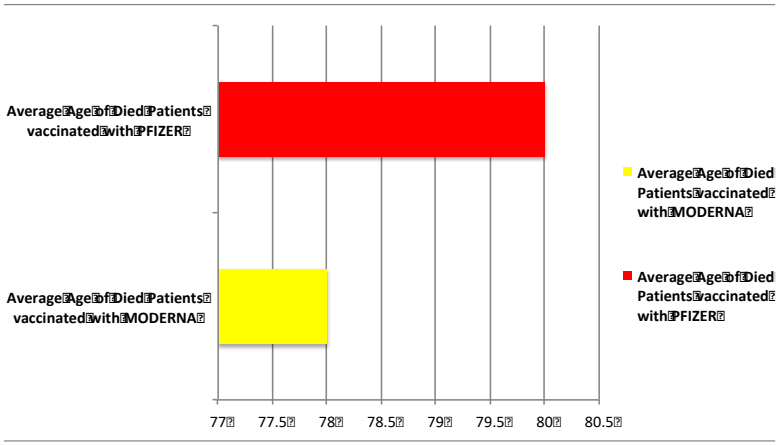


Fig 4: Average Age of died patients

Our experiments show that, which manufacturer vaccine is most suitable for which age group, that shows that the most of the alive patients of Moderna vaccination are belong to the age group 47 to 50 years old, which shows that the age group 47 to 50 years old patients are suitable for Moderna manufacturer vaccination. On the other hand, Pfizer vaccine is suitable for the age group 32 to 37 years old patients, mostly young patients are suitable with the Pfizer vaccine, which is more affected for the young generation to preserver from covid-19, which may face minimum adverse effects and minimize the death ratio in young patients in future. Moreover, vaccination process is compulsory required to minimize the spread of covid-19 pandemic. These age groups have minimum adverse effects and death ratio in the patients.

Table 5: Age group of Alive patients of each vaccine

Vaccination Manufacturer	Age Group
Average Age of Alive Patients vaccinated with Pfizer	32 - 37
Average Age of Alive Patients vaccinated with Moderna	47 -50

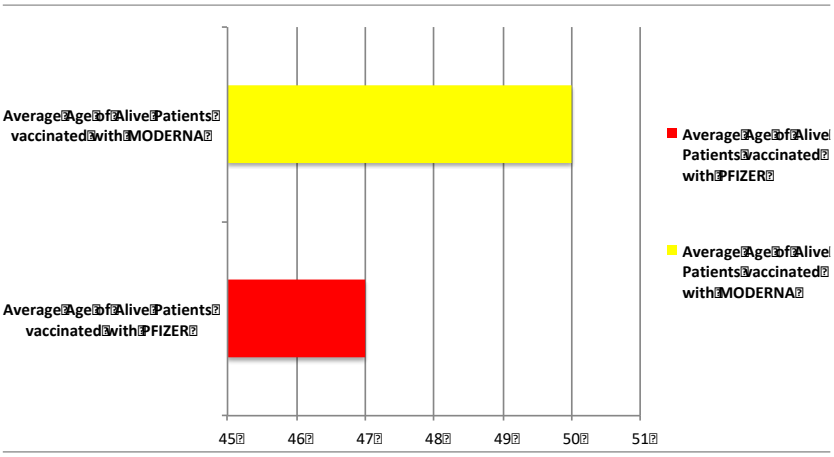


Fig 5: Average Age of Alive patients

Experiments shows that the both Moderna and Pfizer have adverse effect in vaccinated patients, as shown in table 3, 1980 patients were vaccinated with the Pfizer and 855 patients were vaccinated with Moderna vaccine. Further shown in figure 2 and figure 3, the 201 Moderna vaccinated patients died and the 239 Pfizer vaccinated patients died, from the rest 1741 patients survive from which 967 patients were faced adverse effects of Pfizer vaccine and the 362 patients out of 654 patients reported adverse effects and admitted in the hospitals for further medical process, all these 967 patients of Pfizer and 362 patients of Moderna are recovered after some medication process shown in figure 6.

Table 6: Adverse Reaction and Recovered Patients

Vaccination Manufacturer	Patients
Moderna Adverse Reaction Reported and Recovered	362
PFIZER Adverse Reaction Reported and Recovered	967

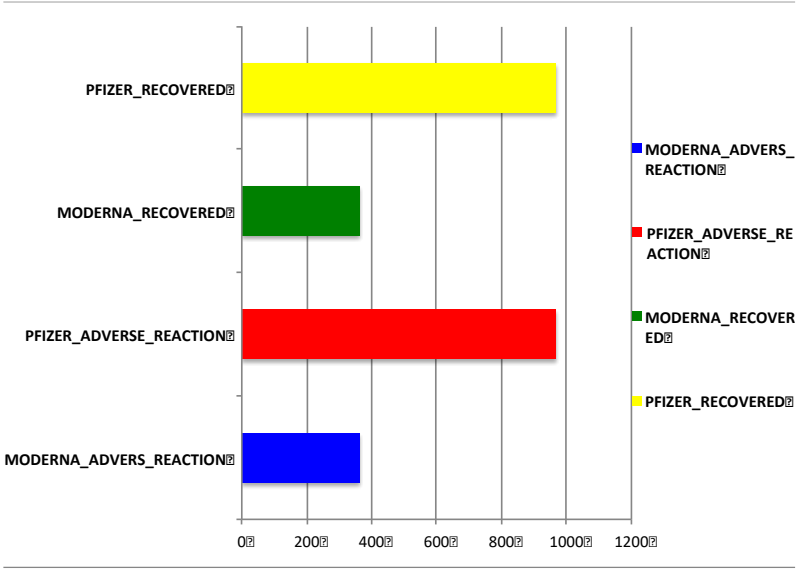


Fig 6: Adverse Reaction and Recovered Patients

As we know, both Moderna and Pfizer have adverse effects in patients, these adverse effects are some harmful due to which patient need some serious emergency treatment or need to be hospitalized. In these symptoms, 77 patients were suffered again covid-19, 45 patients survive with serious headache, 43 patients are survived with chills, there are 28 patients who recovered from the vaccination adverse effects but later on face natural death. On these figure, Naïve Bayes are trained to learn the pattern of those patients who recovered from covid-19 virus or died due to virus or due to adverse effects of vaccine and or those patients that face adverse effects and those who saved from any effect.

Table7: Adverse effects of patients with vaccine

Adverse Effects	Patients
SARS-COV-2 test positive	77
Headache	45
Chills	43
Dyspnoea	42
Dizziness	35
Fatigue	35
Pyrexia	30
Pain in extremity	29
Death	28
Nausea	26

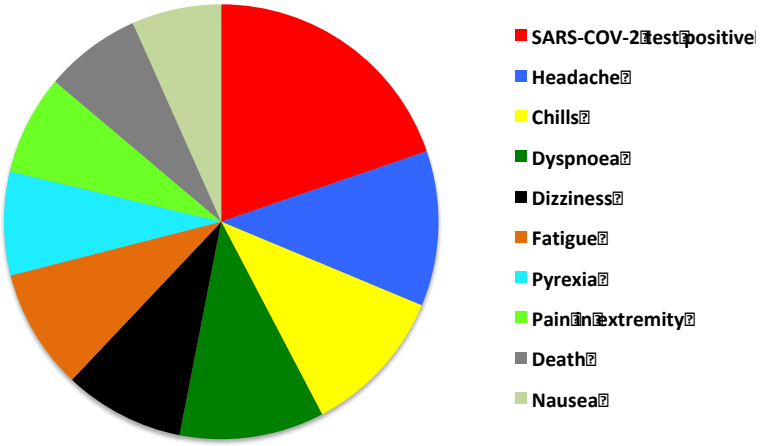


Fig 7: Adverse effects of patients with vaccine

In our research, we are preparing machine learning model to predict the risk of patient death predictions, adverse effects of covid-19 vaccination and chance of patient recovery predictions for covid-19 patients who vaccinated with the Moderna or Pfizer vaccine. To predict these outcomes, we used Support vector Machine, Naïve Bayes, Random Forest and Gradient Boosting Machine Algorithm to predict the outcomes. The VAERS (Vaccine Adverse Event Reporting System) dataset was split into 80% for training the models and the rest of the 20% for validation of models. Hyperparameters for each model were fine-tuned using Grid Search, optimizing for accuracy and F1-score. We trained all the mentioned model on the VAERS (Vaccine Adverse Event Reporting System) dataset but the accuracy of these Naïve Bayes, Random Forest and Gradient Boosting is minimum, we applied the hyper parameter tuning on all these mentioned models still we get these accuracy results. – results. The Support Vector Machine model predict the risk of patient's death after vaccination with the F1- Score of 91% with the precision of 88% and a recall of 93%. The age of patients and the pre-existing disease are major harmful and major cause of death in patients. The SVM model also predict the adverse effects of the covid-19 vaccine with the F1-Score of 87%, with the precision score of 86% and the recall score is 90%. Most important predictions for the vaccination adverse effects included the previous skin allergic reactions, headache, diarrhea, and patient age. For the recovery of covid-19 patients with the Pfizer and Moderna vaccine, these are the accuracy generated by the Support Vector Machine with the F1-Score 82% with a precision of 79% and a recall of 85%. The complete accuracy of all the trained models are shown in table 8.

Table 8: Confusion Matrix Results of Machine Learning Models

Machine Learning Models	Predictions	F1-Score	Precision	Recall

Support Vector Machine	Risk of Death	91%	88%	93%
	Adverse Effects	87%	86%	90%
	Recovery Prediction	82%	79%	85%
Naïve Bayes	Risk of Death	69%	66%	68%
	Adverse Effects	62%	60%	61%
	Recovery Prediction	64%	65%	67%
Random Forest	Risk of Death	73%	70%	72%
	Adverse Effects	68%	65%	69%
	Recovery Prediction	54%	49%	59%
Gradient Boosting	Risk of Death	82%	85%	81%
	Adverse Effects	78%	75%	79%
	Recovery Prediction	77%	75%	78%

In our study several ethical considerations must be addressed. Ensuring data privacy is critical, with all patient data anonymized to comply with regulations such as FDA, CDC and WHO. Informed consent should be obtained for any personal data used, and bias in machine learning models must be mitigated to avoid unequal health outcomes for underrepresented groups. The accuracy and reliability of the model are essential, as incorrect predictions—whether false positives or negatives—can have serious health implications. Additionally, transparency and explain ability in model predictions are vital to ensure that healthcare providers understand and trust the system. Ethical use of the predictions, particularly in preventing discrimination or over-reliance on the model, must be prioritized, ensuring machine learning complements clinical judgment rather than replacing it. Finally, clear accountability must be established for the ethical application of such predictive models in healthcare settings.

V. Conclusion

With the advent of covid-19 from china in 2019, most of the population effected with the virus, many pharmaceutical companies started work to make vaccine to decrease the effects of covid-19 or minimise the spread of virus. Pfizer and Moderna are the first companies who invent and started clinical trials on patients and on some non-effected peoples to measure the accuracy and efficiency of the vaccine. This research is done to test the effect of different vaccines by using the VAERS dataset which is collected and created by FDA and CDC. As a part of this research, 3000 patients are enrolled in clinical trials, 855 patients were vaccinated with Moderna and 1980 patients were vaccinated with Pfizer. From Pfizer vaccination 239 patients died and 1741 patients were recovered through this vaccination and 201 patients died after vaccinated with Moderna and 654 patients recovered. The most adverse effect occurs on average age of died patients vaccinated with Moderna is 68 to 71 years old, and the Pfizer's adverse effects on average age of died patients is 80 - 84 years old. The Moderna was suitable for the 47 to 50 years old age patients and Pfizer was suitable for the patients of 32 to 37 years old age patients. The prime focus was laid on targeting the common symptoms of patients with vaccine is SARS-COV-2 test positive, which is 77, Headache is 45, Chills is 43 etc. The machine learning model support vector machine's predictions are better than other machine learning models, support vector machine gives the predictions for the 91% for the risk of death, 87% with adverse effects, and the 82% for the Patients recovery after the covid-19 vaccination.

VI. Future Work

Pfizer and Moderna towards their success track, both pharmaceutical companies need to upgrade their vaccines for the senior citizen ages greater than 60, and they also need to improve the side effects of their vaccines, human body is polite against heavy medication, if there is any mishandling in vaccination it may harmful for human health and cause of increase in death. This research paper has opened a new gateway for the pharmaceuticals company to work on and improve their vaccine. This research can be subjected to further extension towards fast recovery and fight with symptoms occurred.

References

[1] World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. Available from: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 (Accessed 24 February 2024)

[2] Marriam Webster Dictionary. Pandemic. Available from:https://www.merriam-webster.com/dictionary/pandemic (Accessed 24 February 2024)

[3] World Health Organisation. Novel Coronavirus – China.Disease outbreak news : Update 12 January 2020.

[4] Public Health England. COVID-19: epidemiology, virology and clinical features. https://www.gov.uk/government/publications/wuhan-novel-coronavirus-background-information/wuhan-novel-coronavirus-epidemiology-virology-and-clinical-features. Accessed 24 February 2024

[5] WHO Coronavirus disease (COVID-19): How is it transmitted?Available from: https://www.who.int/news-room/q-a-detail/coronavirus-disease-covid-19-how-is-it-transmitted (Accessed 11.02.2024)

[6] The New York Times (26 February 2024) *Coronavirus (COVID-19) statistics data*, Available at: *https://support.google.com/websearch/answer/9814707?p=cvd19_statistics&hl=en-PK&visit_id=637499514348267645-2438641803&rd=1* (Accessed: 26 February 2024).

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[8] Ammar Haider, & Dr Mumtaz Ali Sheikh. (2022). Analysis of Covid-19 Trends: A Machine Learning Approach. https://doi.org/10.5281/zenodo.6508743

[9] Kaur, R. J., Dutta, S., Bhardwaj, P., Charan, J., Dhingra, S., Mitra, P., ... & Misra, S. (2021). Adverse events reported from COVID-19 vaccine trials: a systematic review. *Indian Journal of Clinical Biochemistry*, *36*(4), 427-439.

[10] Babaee, E., Amirkafi, A., Tehrani-Banihashemi, A., SoleimanvandiAzar, N., Eshrati, B., Rampisheh, Z., ... & Nojomi, M. (2022). Adverse effects following COVID-19 vaccination in Iran. *BMC Infectious Diseases*, *22*(1), 476.

[11] Beatty, A. L., Peyser, N. D., Butcher, X. E., Cocohoba, J. M., Lin, F., Olgin, J. E., ... & Marcus, G. M. (2021). Analysis of COVID-19 vaccine type and adverse effects following vaccination. *JAMA network open*, *4*(12), e2140364-e2140364.

[12] Al Khames Aga, Q. A., Alkhaffaf, W. H., Hatem, T. H., Nassir, K. F., Batineh, Y., Dahham, A. T., ... & Traqchi, M. (2021). Safety of COVID-19 vaccines. *Journal of medical virology*, *93*(12), 6588-6594.

[13] Saeed, B. Q., Al-Shahrabi, R., Alhaj, S. S., Alkokhardi, Z. M., & Adrees, A. O. (2021). Side effects and perceptions following Sinopharm COVID-19 vaccination. *International Journal of Infectious Diseases*, *111*, 219-226.

[14] Meo, S. A., Bukhari, I. A., Akram, J., Meo, A. S., & Klonoff, D. C. (2021). COVID-19 vaccines: comparison of biological, pharmacological characteristics and adverse effects of Pfizer/BioNTech and Moderna Vaccines. *European Review for Medical & Pharmacological Sciences*, *25*(3).

[15] Hosseini, R., & Askari, N. (2023). A review of neurological side effects of COVID-19 vaccination. *European Journal of Medical Research*, *28*(1), 102.

[16] Dhamanti, I., Suwantika, A. A., Adlia, A., Yamani, L. N., & Yakub, F. (2023). Adverse reactions of COVID-19 vaccines: a scoping review of observational studies. *International Journal of General Medicine*, 609-618.

[17] Abukhalil, A. D., Shatat, S. S., Abushehadeh, R. R., Al-Shami, N. M., Naseef, H. A., & Rabba, A. (2023). Side effects of Pfizer/BioNTech (BNT162b2) COVID-19 vaccine reported by the Birzeit University community. *BMC Infectious Diseases*, *23*(1), 5.

[18] Mohseni Afshar, Z., Sharma, A., Babazadeh, A., Alizadeh-Khatir, A., Sio, T. T., Taghizadeh Moghadam, M. A., ... & Ebrahimpour, S. (2023). A review of the potential neurological adverse events of COVID-19 vaccines. *Acta Neurologica Belgica*, *123*(1), 9-44.

[19] Faksova, K., Walsh, D., Jiang, Y., Griffin, J., Phillips, A., Gentile, A., ... & Hviid, A. (2024). COVID-19 vaccines and adverse events of special interest: A multinational Global Vaccine Data Network (GVDN) cohort study of 99 million vaccinated individuals. *Vaccine*, *42*(9), 2200-2211.

[20] Eterafi, M., Fouladi, N., Golizadeh, M., Shaker, H., Matin, S., & Safarzadeh, E. (2024). Reported side-effects following Oxford/AstraZeneca COVID-19 vaccine in the north-west province, Iran: A cross-sectional study. *Plos one*, *19*(1), e0296669.

[21] Ahsanuddin, S., Jin, R., Dhanda, A. K., Georges, K., Baredes, S., Eloy, J. A., & Fang, C. H. (2024). Otolaryngologic side effects after COVID-19 vaccination. *The Laryngoscope*, *134*(3), 1163-1168.

[22] Suhail, L., Masood, S., & Haider, A. (2024). A Comparative Study of Sentiment Analysis for Mental Health Related Posts at Reddit & Twitter Using Machine Learning and Pre-Trained Models. *Journal of Innovative Computing and Emerging Technologies*, *4*(2).

[23] Haider, A., & Sabahat, N. (2022, March). A Usability and Accuracy Measurement of Smartphones Face Recognition. In *2022 2nd International Conference on Artificial Intelligence (ICAI)* (pp. 19-25). IEEE.

[24] Noor, H., Ahmad, J., Haider, A., Nasim, F., & Jaffar, A. (2024). A Machine Learning Sentiment Analysis Approach on News Headlines to Evaluate the Performance of the Pakistani Government. *Journal of Computing & Biomedical Informatics*, *7*(02).

A Comparative Analysis and Usability Measurement of Window Operating System and Mac Operating System

Ammar Haider
School of System and Technology
University of Management and Technology
Lahore, Pakistan
ammar.haider@umt.edu.pk

Muhammad Husnain Ashfaq
School of System and Technology
University of Management and Technology
Lahore, Pakistan
husnain.ashfaq@umt.edu.pk

Hajira Noor
Department of Computer Science
Superior University, Gold Campus
Lahore, Pakistan
SU92-MSCSW-F22-030@superior.edu.pk

Hassan Moatasam Awan*
School of System and Technology
University of Management and Technology
Lahore, Pakistan
hassan.moatasam@umt.edu.pk

***Abstract*—This research paper presents a comparative analysis and usability analysis of two of the most widely used operating systems in the world, Mac OS and Windows OS. The paper begins with a brief introduction to the two operating systems, followed by a detailed discussion on their respective interfaces, usability, and features. The analysis is based on various usability metrics such as learnability, efficiency, memorability, errors, and satisfaction. We conclude by presenting our findings and discussing the implications of our study. BY performing both comparative and usability assessments, we observe the strengths and weakness of each windows and MacOS operating system. Our results show that the macOS have strong security mechanism and the convenient apple ecosystem with other apple devices. On the other hand, Window operating system extremely compatible with third party software and provide customization options. The study concludes the highlighting that the scenarios in which each operating system is best suited for novice and experienced users, providing valuable insights for both users and organizations in selecting the most appropriate operating system for their needs.**

Keywords—Usability Measurement, Comparative Analysis, Window Operating System, Mac Operating System, User Experience, Microsoft, Apple

I. INTRODUCTION

The operating system is the backbone of any computer system, and it plays a vital role in the overall user experience. Two of the most popular operating systems in use today are Mac OS and Windows OS. Mac OS is the operating system developed by Apple Inc. for its line of Macintosh computers, while Windows OS is the operating system developed by Microsoft Corporation for use on personal computers [1]. Both Mac OS and Windows OS have their strengths and weaknesses. Mac OS is known for its ease of use, sleek design, and superior performance, while Windows OS is known for its compatibility with a wide range of hardware and software, as well as its flexibility and customization options. Mac OS and Windows OS are the two most popular operating systems used by personal computer users worldwide. Mac OS is developed by Apple Inc. and is exclusively used on Apple's hardware devices, while Windows OS is developed by Microsoft and is used on a wide range of hardware devices manufactured by various vendors. Both operating systems have evolved over the years, with each new release offering new features and improvements [2]. In this paper, we aim to provide a comparative analysis and usability analysis of these two operating systems. Comparative Analysis and Usability analysis of the latest Windows OS and Mac OS is a critical examination of the two most popular operating systems in the market today. Windows OS and Mac OS have a significant share in the computer industry and are constantly updated to keep up with technological advancements. This analysis aims to evaluate the key features, functionalities, and user experience of the latest versions of Windows and Mac OS. The comparison will be based on their user interface, performance, security, software compatibility, and other critical factors that affect their usability. Usability analysis, on the other hand, evaluates the ease of use, learnability, and user satisfaction of an operating system [3]. It is an essential aspect of software development that helps developers to design software that meets the needs of the end-users. Through this comparative analysis and usability analysis, we aim to provide insights into the strengths and weaknesses of the latest Windows and Mac OS. This will help users make informed decisions on which operating system is best suited for their needs, whether for personal or business use.

II. RELATED WORK

First, Several studies have compared the usability of Windows and macOS operating systems. A study by Nielsen Norman Group compared the two operating systems and found that macOS was more user-friendly and intuitive than Windows. The study found that users of macOS could perform tasks more quickly and accurately than those using Windows. The study also

highlighted the consistency of macOS's interface, which made it easier for users to navigate and accomplish their tasks [4]. In another study conducted by Pfeiffer Consulting, the researchers compared Windows 10 and macOS Sierra in terms of usability. The study found that macOS was more user-friendly than Windows 10, and users were more productive when using macOS. The researchers attributed the superior usability of macOS to its simplicity, consistency, and customization options [5]. A study by McLeod and Schell also compared the two operating systems and found that macOS had a more consistent interface and was easier to use than Windows. The study found that users of macOS had higher satisfaction levels than those using Windows. The researchers also noted that macOS's user interface was designed to reduce cognitive load, making it easier for users to focus on their tasks [6]. The usability of an operating system depends on several factors, including ease of use, consistency, customizability, and accessibility. The studies mentioned above have highlighted some of the differences between Windows and macOS in terms of usability.

One of the key factors that contribute to macOS's superior usability is its consistency. The interface and interactions are consistent across the operating system, which makes it easier for users to navigate and perform their tasks. In contrast, Windows has a more fragmented interface, with different elements and interactions across various applications and settings. Another factor that contributes to macOS's usability is its simplicity. macOS is designed with a minimalist approach, with fewer distractions and clutter. This simplicity makes it easier for users to focus on their tasks and reduces cognitive load [7]. Customization is another factor that contributes to the usability of an operating system. macOS provides users with a high degree of customization, allowing them to personalize their experience and workflows. In contrast, Windows has limited customization options, which can make it challenging for users to optimize their workflows. A study by Hekmatfar and Talebpoor (2019) compared the usability of Windows and macOS in the context of educational settings. The review found that both operating systems had strengths and weaknesses in terms of usability, but macOS was generally perceived as more user-friendly and easier to use by educators and students. The review also highlighted the importance of considering factors such as customization, accessibility, and support when choosing an operating system for educational purposes [8]. George and Kumar (2019) compared the user experience of Windows and macOS from the perspective of usability, aesthetics, and user satisfaction. The review found that macOS was generally perceived as more aesthetically pleasing and user-friendly than Windows, and users reported higher levels of satisfaction with macOS. The review also highlighted the importance of considering factors such as consistency, simplicity, and customization when evaluating the user experience of operating systems [9].

The usability of Windows and macOS in the context of software development. The review found that macOS was generally perceived as more user-friendly and intuitive for software development tasks, and developers reported higher levels of productivity when using macOS. The review also highlighted the importance of considering factors such as performance, compatibility, and development tools when choosing an operating system for software development purposes [10]. A study conducted by Hafez, El-Bahnasawy, and Elshazly (2020) compared the usability of Windows and macOS in the context of graphic design. The review found that macOS was generally perceived as more user-friendly and efficient for graphic design tasks, and designers reported higher levels of creativity and productivity when using macOS. The review also highlighted the importance of considering factors such as interface design, customization, and support when choosing an operating system for graphic design purposes [11]. Author Yadav and Sharma (2020) compared the user experience of Windows and macOS from the perspective of usability, security, and performance. The review found that both operating systems had strengths and weaknesses in each of these areas, and highlighted the importance of considering factors such as customization, user support, and third-party applications when evaluating their usability [12]. A study performed by Kim and Lee (2021) compared the usability of Windows and macOS for multitasking and productivity. The review found that both operating systems had unique features and tools for multitasking, but macOS was generally perceived as more user-friendly and efficient for productivity tasks. The review also highlighted the importance of considering factors such as interface design, performance, and compatibility when evaluating the usability of operating systems for multitasking and productivity [13].

The usability of Windows and macOS in the context of mobile app development performed by the Arvind and Somanath (2020). The review found that macOS was generally perceived as more user-friendly and intuitive for mobile app development tasks, and developers reported higher levels of productivity when using macOS. The review also highlighted the importance of considering factors such as development tools, interface design, and compatibility when choosing an operating system for mobile app development purposes [14]. Ramachandran and Ponnuswamy (2020) compared the usability of Windows and macOS in the context of digital forensics. The review found that both operating systems had strengths and weaknesses in terms of digital forensics tools and capabilities, and highlighted the importance of considering factors such as file system analysis, memory analysis, and network analysis when evaluating their usability for digital forensics purposes [15].

A major criticism when comparing the usability of Windows and Mac operating systems is the limited consideration of certain customer groups such as engineers, professors or students, for whom most queries usually provide conclusions broad and general. Current thinking also often focuses on short-term customer feedback and pays little attention to how the customer encounter evolves over time. Additionally, there are inconsistencies in usability evaluation methods, making it difficult to make accurate comparisons. Another understudied area is how each work system improves accessibility for people with disabilities and the overall impact of these features on comfort. The importance of using both frameworks in crossover situations, where clients typically move between them, has not been widely considered. Additionally, little attention has been paid to the impact of personalization options or third-party apps on the customer experience at each stage. As the two frameworks increasingly coordinate AI and other untapped technologies, additional research is needed to examine the impact of these drivers on usability. Closing these gaps can lead to a much better, much better, higher, more robust, and improved understanding of how clients connect to Windows and Mac.

III. METHODOLOGY

A. *Comparative Analysis:*

I. *User Interface:* Both Windows and Mac OS have a graphical user interface, but their design and layout are different. Mac OS is known for its sleek and intuitive design, with icons and menus that are easy to navigate. Windows, on the other hand, has a more utilitarian design, with a focus on productivity and flexibility.

II. *Features:* Both operating systems offer similar features, such as file management, web browsing, and multimedia playback. However, Mac OS includes some unique features, such as Time Machine, which allows users to easily backup and restore their data, and Airdrop, which enables easy file sharing between Mac devices.

III. *Performance:* Both operating systems are known for their stability and reliability. However, Mac OS is optimized for Apple's hardware and is known for its smooth performance and fast boot times. Windows, on the other hand, can be installed on a wide range of hardware, which can lead to performance inconsistencies.

IV. *Security:* Both operating systems offer built-in security features, such as firewalls and antivirus software. However, Mac OS is known for its strong security features, including its built-in malware protection and secure boot process. Windows has also improved its security features in recent years but is still more susceptible to malware and viruses.

B. *Usability Analysis:*

The experiment of this research was performed in the computer science graduate lab of our department. All the experiments were performed under the supervision and guidance of our advisor and department faculty. Each participant performs the required experiments and fulfil the questionnaire which have twenty questions have answers in the range of one to five, which is classifies as one for poor and 5 for the best. We have the latest machines in which latest operating systems are installed to perform the required tasks, our first system is MacBook Air 2022 M2 Chip with macOS 14 Sonoma, and the second system is HP elite book X360 with window OS 10 latest update. There are 258 participants were joined our experiments, we used university student email to collect the participants for our research, in which 141 participants are male and the 117 participants are female, most of the participants are known users of these operating systems. First each participant performed same basic daily based experiments on both systems to measure the performance and their familiarity with the OS, each user has fifteen minutes to perform all these experiments.

I. Novice User:

a. *User Interface:* Mac OS has a more intuitive and user-friendly interface for novice users, with simple and easy-to-use features. Windows, on the other hand, may have a steeper learning curve for novice users due to its more complex design and features.

b. *Features:* For novice users, both operating systems have similar features, but Mac OS is generally considered more user-friendly and easier to navigate. Windows may overwhelm novice users with its wide range of features and customization options.

c. *Performance:* Novice users may not notice a significant difference in performance between the two operating systems, but Mac OS is generally considered to have faster boot times and smoother performance.

d. *Security:* Novice users may not be familiar with security features and may not prioritize them when choosing an operating system. However, Mac OS is generally considered to have stronger built-in security features.

II. Experienced User:

a. *User Interface:* Experienced users may prefer Windows' more utilitarian and customizable interface, with more advanced features for power users. Mac OS may feel limiting for experienced users who are accustomed to a more complex and customizable interface.

b. *Features:* Both operating systems offer a wide range of features, but experienced users may prefer Windows for its advanced customization options and access to more software applications.

c. *Performance:* Experienced users may notice a difference in performance between the two operating systems, with Windows often offering better performance on high-end hardware.

d. *Security:* Experienced users may prioritize security features and may appreciate the built-in security features of both operating systems. However, Windows may require more advanced security measures, such as third-party antivirus software, to ensure adequate protection.

IV. EXPERIMENTS

An overview of some questions included in the questionnaire and their results are mentioned in the Table 1.

1. Which operating system did you find easier to navigate and use: Windows or Mac OS?
2. Did you find it easier to find and install software on Windows or Mac OS?

3. Which operating system did you find had better support resources for troubleshooting issues: Windows or Mac OS?
4. Which operating system did you find had better compatibility with external devices, such as printers or cameras: Windows or Mac OS?
5. Which operating system did you find had better overall performance and speed: Windows or Mac OS?
6. Did you find the default settings and options on Windows or Mac OS to be more user-friendly?
7. Which operating system did you find had better security and privacy features: Windows or Mac OS?
8. Did you experience any compatibility issues with software or devices on either Windows or Mac OS?
9. Which operating system did you find had better customization options: Windows or Mac OS?
10. Which operating system would you recommend to a friend or colleague who is a novice user: Windows or Mac OS?
11. Did you find the default applications and software on Windows or Mac OS to be more useful for your needs?
12. Which operating system did you find had a more appealing and aesthetically pleasing design: Windows or Mac OS?
13. Did you find it easier to switch between multiple applications and windows on Windows or Mac OS?
14. Which operating system did you find had better battery life on laptops: Windows or Mac OS?
15. Did you find any significant differences in the speed and performance of the two operating systems when running the same applications?
16. Which operating system did you find had better accessibility options for users with disabilities: Windows or Mac OS?
17. Did you find it easier to connect and use Bluetooth devices on Windows or Mac OS?
18. Which operating system did you find had better file management features, such as searching and organizing files: Windows or Mac OS?
19. Did you find it easier to adjust system settings and preferences on Windows or Mac OS?
20. Which operating system did you find had better integration with cloud storage services: Windows or Mac OS?

Here are some experiments which were performed by the participants on the both of the systems, first system has macOS Sonoma and the second system have windows OS 10 with latest updates.

E1. Application installation and uninstallation setup
E2. Desktop Customization and personalization performance
E3. Multi-user account creation and domain account creation
E4. System update test
E5. Checking the security and privacy measurements and firewall activation
E6. Printer and Scanner installation
E7. Disk management setup
E8. Gestures and touchpad navigation
E9. Accessibility and keyboard command settings

V. RESULTS

To perform our experiments 258 participants were joined our research, we used university domain for the collection of desire participants for our research, in which 141 participants are male and the 117 participants are female, from which 32 male participants are novice users and the rest 109 male participants have some hands-on experience on these operating systems, on the other hand 43 female participants out of 117 are novice users and the rest of the 74 female participants have some hands-on experience on these operating systems shown in figure 1.

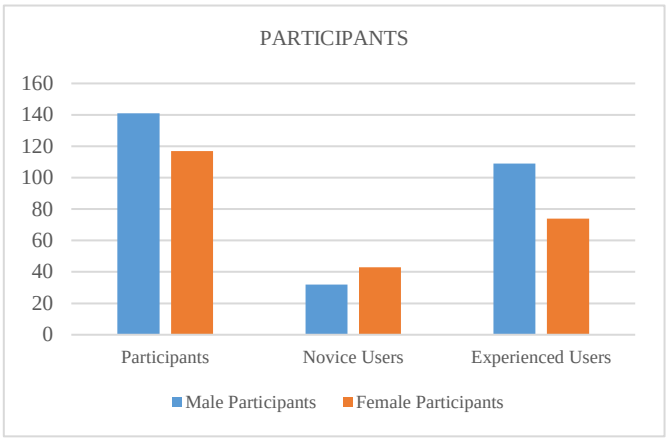


Fig 1: Participants

All the participants have equal time to perform the twenty experiments on both systems to measure the performance and their familiarity with the OS, in which our first system is MacBook Air 2022 M2 Chip with macOS 14 Sonoma, and the second system is HP elite book X360 with window OS 10 latest update. Each user has fifteen minutes to perform all these experiments. After the performance of experiments each participant have assigned the questionnaire to fulfil the questions which have twenty questions have answers in the range of one to five, which is classifies as one for poor and 5 for the best. The results of these experiments were shown in figure 2 and figure 3. First ten experiments results shown in figure 1 and the rest ten experiments results shown in figure 2, also table 1 shows the experiments values of these experiments.

Fig 2: Experiments Results 1 - 10

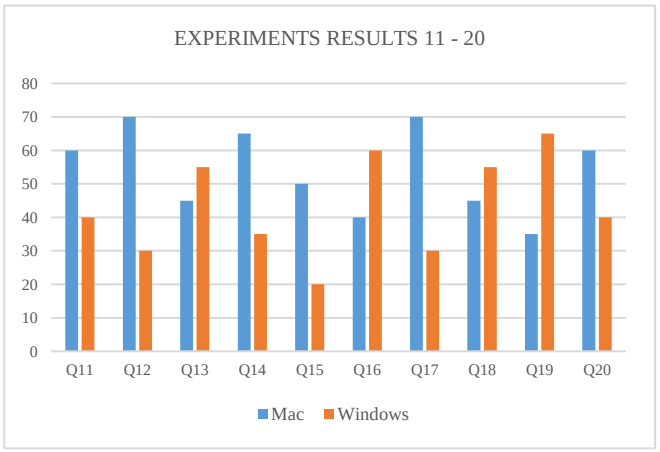


Fig 3: Experiments Results 10 – 20

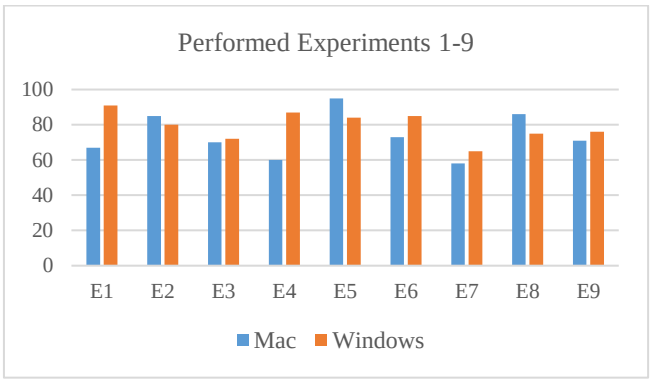


Fig 4: Performed Experiments Results 1 – 9

Table 1: Experiments Values

Operating System	MacOS Preference	Window OS Preference
Q1	60%	40%

Q2	70%	30%
Q3	45%	55%
Q4	35%	65%
Q5	50%	50%
Q6	75%	25%
Q7	60%	40%
Q8	60%	40%
Q9	45%	55%
Q10	70%	30%
Q11	60%	40%
Q12	70%	30%
Q13	45%	55%
Q14	65%	35%
Q15	50%	20%
Q16	40%	60%
Q17	70%	30%
Q18	45%	55%
Q19	35%	65%
Q20	60%	40%

VI. Conclusion

Based on the experiments results, it appears that there is a preference for certain features of both Windows and Mac OS among novice users and experienced users. Novice users found the default applications and software on Mac OS to be more useful for their needs and also found it easier to switch between multiple applications and windows on Mac OS. They also found Mac OS to have better battery life on laptops and better integration with cloud storage services. On the other hand, novice users found Windows to have better accessibility options for users with disabilities, better file management features, and easier adjustment of system settings and preferences. Overall, it appears that novice users have a slight preference towards Mac OS due to its ease of use and better default applications, but also recognize certain strengths of Windows in terms of accessibility and file management. Experienced users have hands-on experience on their desire operating system, our results shows most of the experienced users have preference for the macOS for their professional use like in industry, software and application development due to their performance and OS reliability, on the other hand for the purpose of personal use experienced users have a preference on the windows operating system for their daily use purpose.

VII. Future Work

Technology is enhancing rapidly on daily bases, new features and advance technologies are introducing in every OS update. Operating system manufactures work on the user's feedback to improve the usability and performance of their respective product. Each major update of the operating system will directly affect on the usability and performance of the operating system. There is an open book invitation to work on the usability measurement on every update of the operating system to measure their usability, which will directly helpful the users and the manufacture to accurately resolving the issues and making the updates in right track.

References

[1] 5 Most popular operating systems. (2021, April 19). Western Governors University. https://www.wgu.edu/blog/5-most-popular-operating-systems1910.html

[2] Topcuoglu, C., Martinez, A., Acar, A., Uluagac, S., & Kirda, E. MacOS versus Microsoft Windows: A Study on the Cybersecurity and Privacy User Perception of Two Popular Operating Systems.

[3] Essien, N. P., & Umo, M. C. (2024). Conceptual Analysis of Optimization Strategies in Operating System (Windows and Unix) in the Current IT Trend from 2015 to 2023. International Journal of Contemporary Africa Research Network, 2(1).

[4] Nielsen Norman Group. (2018). Windows 10 vs. macOS: Which is more productive? Retrieved from https://www.nngroup.com/articles/windows-10-vs-macos/

[5] Pfeiffer, F. (2016). The state of the desktop operating systems: A user experience comparison of Windows 10 and macOS Sierra. Pfeiffer Consulting. Retrieved from https://www.pfeifferreport.com/articles/2016/state-of-desktop-os-2016

[6] McLeod, R., & Schell, G. (2017). Usability comparison of Windows 10 and macOS Sierra: A case study. Journal of Usability Studies, 12(1), 1-18. Retrieved from https://uxpajournal.org/usability-comparison-windows-10-macos-sierra/

[7] Tognazzini, B. (2017). Windows 10 vs. macOS: Which operating system is better for you? Fast Company. Retrieved from https://www.fastcompany.com/3068665/windows-10-vs-macos-which-operating-system-is-better-for-you

[8] Hekmatfar, M., & Talebpoor, A. (2019). A comparative study of the usability of Windows and macOS in educational environments. Journal of Educational Technology Development and Exchange, 12(1), 27-46.

[9] George, A., & Kumar, D. (2019). A comparative analysis of user experience between Windows and macOS. International Journal of Computer Applications, 182(33), 25-31.

[10] Sabir, M. A., Mahmood, T., & Saqib, M. A. (2019). A comparative study of the usability of Windows and macOS for software development. Journal of Information and Communication Technology, 18(3), 417-439.

[11] Hafez, M., El-Bahnasawy, N., & Elshazly, D. (2020). A comparative study of the usability of Windows and macOS for graphic design. Journal of Theoretical and Applied Information Technology, 98(16), 2814-2827.

[12] Yadav, R., & Sharma, M. (2020). A comparative study of Windows and Mac operating systems. International Journal of Engineering and Advanced Technology, 9(6), 2026-2033.

[13] Kim, S. H., & Lee, S. H. (2021). A comparative study on the usability of multitasking features in Windows and macOS. Journal of Information Technology and Application in Education, 10(1), 1-11.

[14] Arvind, V., & Somanath, N. (2020). A comparative study of the usability of Windows and macOS for mobile app development. International Journal of Scientific Research and Review, 9(2), 53-63.

[15] Ramachandran, S., & Ponnuswamy, S. (2020). A comparative study on Windows and macOS for digital forensics. International Journal of Recent Technology and Engineering, 8(6S2), 89-94.

[16] Zahid, Z., Haider, A., Sabahat, N., & Tanwir, A. (2020, November). Vulnerabilities in Biometric Authentication of Smartphones. In *2020 IEEE 23rd International Multitopic Conference (INMIC)* (pp. 1-5). IEEE.

[17] Haider, A., & Sabahat, N. (2022, March). A Usability and Accuracy Measurement of Smartphones Face Recognition. In *2022 2nd International Conference on Artificial Intelligence (ICAI)* (pp. 19-25). IEEE.

[18] Bowling, H., Seigfried-Spellar, K., Karabiyik, U., & Rogers, M. (2023). We are meeting on Microsoft Teams: Forensic analysis in Windows, Android, and iOS operating systems. *Journal of Forensic Sciences*, *68*(2), 434-460.

[19] Essien, N. P., & Umo, M. C. (2024). Conceptual Analysis of Optimization Strategies in Operating System (Windows and Unix) in the Current IT Trend from 2015 to 2023. *International Journal of Contemporary Africa Research Network*, *2*(1).

[20] A Ibtisum, S., Rahman, S. A., & Hossain, S. S. (2023). Comparative analysis of MapReduce and Apache Tez Performance in Multinode clusters with data compression. *World Journal of Advanced Research and Reviews*, *20*(3), 519-526.

[21] Suhail, L., Masood, S., & Haider, A. (2024). A Comparative Study of Sentiment Analysis for Mental Health Related Posts at Reddit & Twitter Using Machine Learning and Pre-Trained Models. Journal of Innovative Computing and Emerging Technologies, 4(2).

[22] Goloboff, P. A., & Morales, M. E. (2023). TNT version 1.6, with a graphical interface for MacOS and Linux, including new routines in parallel. *Cladistics*, *39*(2), 144-153.

[23] Koziolek, H., & Eskandani, N. (2023, April). Lightweight kubernetes distributions: A performance comparison of microk8s, k3s, k0s, and microshift. In *Proceedings of the 2023 ACM/SPEC International Conference on Performance Engineering* (pp. 17-29).

[24] John, E., Kalu, C., & Asuquo, P. (2024). Comparative Performance Analysis Of Cybersecurity Tools On A Wireless Network With WPA2 Encryption.

[25] Haider, A., & Bhatti, W. (2022, November). Importance of Cyber Security in Software Quality Assurance. In *2022 17th International Conference on Emerging Technologies (ICET)* (pp. 6-11). IEEE.

[26] Topcuoglu, C., Martinez, A., Acar, A., Uluagac, S., & Kirda, E. MacOS versus Microsoft Windows: A Study on the Cybersecurity and Privacy User Perception of Two Popular Operating Systems.

[27] Noor, H., Ahmad, J., Haider, A., Nasim, F., & Jaffar, A. (2024). A Machine Learning Sentiment Analysis Approach on News Headlines to Evaluate the Performance of the Pakistani Government. Journal of Computing & Biomedical Informatics, 7(02).

[28] Qureshi, M. A., Ahmed, S., Mehmood, A., Shaheen, R., & Dildar, M. S. (2024). VULNERABILITY ASSESSMENT OF OPERATING SYSTEMS IN HEALTHCARE: EXPLOITATION IMPLICATIONS TECHNIQUES AND SECURITY. *Health Sciences Journal*, *2*(2), 104-111.

[29] Rasuljon o'g'li, I. S. (2024). TYPES OF NETWORK OPERATING SYSTEM USAGE. *Научный Фокус*, *1*(10), 50-55.

Enhancing IoT Intrusion Detection Using ML, DNN and RNN: A Comparative Study

Muhammad Faiz Raza
Institute of Electrical, Electronics & Computer Engineering
University of the Punjab
Lahore, Pakistan
faiz.raza.dec@gmail.com

Ghufran Ahmed Barcha
Institute of Electrical, Electronics & Computer Engineering
University of the Punjab
Lahore, Pakistan
Gufoor.Barcha2906@gmail.com

Noor Mohsin
Institute of Electrical, Electronics & Computer Engineering
University of the Punjab
Lahore, Pakistan
noormohsin2601@gmail.com

Muhammad Mohsin
Institute of Electrical, Electronics & Computer Engineering
University of the Punjab
Lahore, Pakistan
mohsin.edu.ee@gmail.com

Sheikh Abubakar Kareem
Institute of Electrical, Electronics & Computer Engineering
University of the Punjab
Lahore, Pakistan
0009-0003-4461-8216

Muhammad Shoaib
Institute of Electrical, Electronics & Computer Engineering
University of the Punjab
Lahore, Pakistan
mmshooaib@gmail.com

***Abstract*— Intrusion Detection Systems (IDS) in Internet of Things (IoT) networks face significant challenges due to IoT devices' diverse and evolving nature and sophisticated network attacks. To be effective, IDS must accurately detect threats while managing scalability, privacy, and integration issues. This research addresses the critical issue of enhancing Intrusion Detection Systems (IDS) in Internet of Things (IoT) networks by evaluating the performance of various machine learning, deep learning, and recurrent neural network models. The study utilizes the RT-IoT2022 dataset, which includes a diverse range of IoT devices and advanced network attack methods, to determine the most effective model for real-time cyber threat detection with high accuracy. The dataset comprises 83 features with numerical and categorical attributes, and the target variable includes 12 unique attack types. The research compares models such as Decision Tree, Deep Neural Network (DNN), and Recurrent Neural Network (RNN) based on accuracy, precision, recall, and F1-score. Results indicate that the Decision Tree model is the most efficient, achieving a high accuracy of 99.80% with lower computational costs and reduced training time. However, DNN and RNN models demonstrate advantages when handling complex data. This comparative analysis highlights the necessity of selecting the appropriate model based on specific requirements, providing valuable insights for developing more resilient IDS in IoT environments. Future research will focus on exploring new models as IoT data scales up.**

Keywords—Intrusion Detection Systems (IDS), Internet of Things (IoT), Machine Learning (ML), Deep Learning (DL), Recurrent Neural Networks (RNN), Cyber Threat Detection.

I. INTRODUCTION

The vast network that connects tangible items like gadgets, cars, and buildings is called the Internet of Things. These things can collect and share data because they are fitted with sensors, software, and electronics. [1] The ability to monitor and operate remotely is made possible by the interconnection of current networks, which successfully bridges the gap between the digital and physical domains and improves accuracy and efficiency. The 1968-born British tech pioneer Kevin Ashton is credited with creating the phrase "*Internet of Things*," emphasizing the device's capacity for self-governing functioning. IoT, although still in its infancy, has come a long way in tackling infrastructural, connectivity, and standardization issues while fusing commonplace items with internet-connected sensors. [2] Safeguarding the integrity and confidentiality of data across varying sensitivity levels is critical in Internet of Things (IoT) systems. This necessitates the adoption of secure protocols, such as HTTPS over HTTP, to protect data in transit and at rest. This often involves complex encryption facilitated by devices with hardware accelerators, underscoring the importance of implementing advanced security measures. Moreover, establishing stringent network policies and robust access controls for IoT servers and cloud storage is crucial to prevent unauthorized access and ensure data security within IoT ecosystems.

Sophisticated attacks on IoT systems involve advanced cyber threats that exploit device and network vulnerabilities, [3] often resulting in significant disruption. These include Advanced Persistent Threats (APTs), zero-day exploits, botnet-driven Distributed Denial of Service (DDoS) attacks, and man-in-the-middle (MitM) attacks [9]. These attacks can potentially intercept, alter, or overwhelm IoT systems, underscoring the critical need for robust security measures such as regular updates, encryption, and continuous monitoring to safeguard against these complex threats.

Adversarial network behaviors are acts by hostile actors to trick or interfere with IoT systems by taking advantage of their weaknesses. These actions include data manipulation, device impersonation, Distributed Denial of Service (DDoS) assaults, or Man-in-the-Middle (MitM) attacks. Often, the intention is to jeopardize the IoT network's availability, secrecy, or integrity. Adversarial behaviors may significantly impact the performance and security of IoT systems. Thus, it is critical to put strong security measures in place and conduct ongoing monitoring to identify and eliminate such threats. [4]

Repeated attempts to gain unauthorized access to IoT devices through SSH, also called brute-force SSH attacks, involve attackers systematically testing different username and password combinations until they discover the correct credentials. [5] Although SSH (Secure Shell) is widely used for remotely managing IoT devices, many feature weak or default passwords, leaving them vulnerable to brute-force attacks. Upon successful access, attackers can take control of the device, embed malicious software, or utilize it as part of a botnet for additional attacks, such as Distributed Denial of Service (DDoS). To protect against brute-force SSH attacks, it is crucial to employ strong, unique passwords, activate two-factor authentication, and restrict SSH access to trusted networks.

The RT-IoT2022 dataset, sourced from the UC Irvine Machine Learning Repository, [6] is a new and unique resource created from a real-time IoT infrastructure. It includes many different IoT devices and advanced network attack methods. The dataset shows harmful network actions every day. It has data from IoT devices like ThingSpeak-LED, Wipro-Bulb, and MQTT-Temp. It also has fake attack situations using Brute-Force SSH attacks [5], DDoS attacks with Hping and Slowloris, and Nmap patterns. The dataset captures network traffic and uses the Zeek network tool and the Flowmeter plugin. People can use the RT-IoT2022 dataset to develop Intrusion Detection Systems (IDS) and make better security for real-time IoT networks.

In this study, machine learning (ML), deep learning (DL), and recurrent neural network (RNN) models will be utilized to predict whether an IT device is under attack. The research involves assessing various models to identify the most suitable one for the dataset, highlighting that a deep learning model is only sometimes the optimal choice.

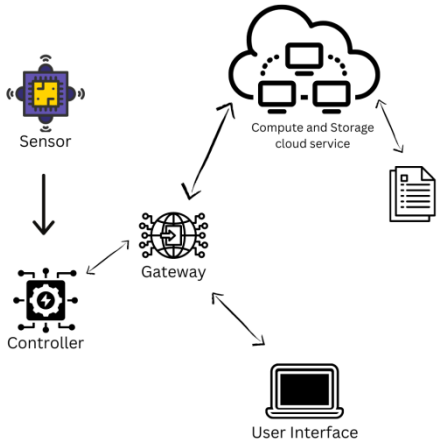


Fig. 1. IoT Infrastructure

The rest of the paper is organized as follows: Section II presents the involvement of AI in IoT, concerned with how different AI techniques, like Machine Learning, Decision Trees, Artificial Neural Networks (ANNs), and Recurrent Neural Networks (RNNs), contribute to intrusion detection in IoT networks. Section III details the methodology, covering dataset preparation, pre-processing, training, and evaluation for the different models. Finally, the results of the comparative analysis of these models are presented in the fourth section, where major performance metrics have been highlighted. Section V concludes the paper, giving a brief summary of the main insights and suggesting directions for future research as IoT data continues growing.

II. ROLE OF AI IN THE INTERNET OF THINGS

Numerous cybersecurity experts are harnessing the power of Artificial Intelligence (AI) for dynamic system protection against cyber threats, primarily utilizing AI for efficient intrusion detection by analyzing traffic patterns to identify activities typical of an attack. [7]

A. Machine Learning

Machine learning in cybersecurity is categorized into supervised and unsupervised learning. Supervised learning involves labeling data malicious or legitimate to train algorithms to classify traffic based on predefined classes. In contrast, unsupervised learning groups data into classes without prior labeling, classifying them based on data coherence and modularity. A standard supervised algorithm in cybersecurity is naive Bayes, which classifies data using the Bayesian theorem by assessing the probability of anomalous activities, treating each event as independent. Machine learning algorithms are essential for creating various cybersecurity models. [8]

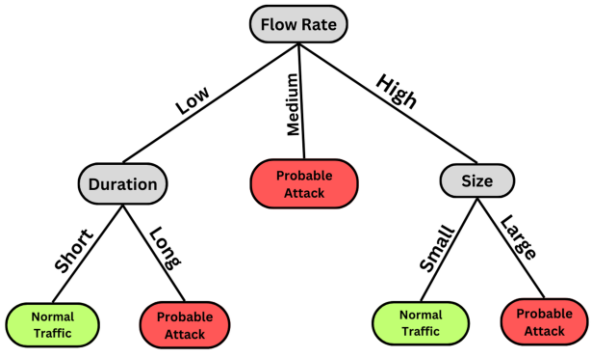


Fig. 2. Decision Tree for IoT

B. Decision Trees

A decision tree is an AI technique to categorize network traffic by creating rules from training data.
It can identify cyberattacks like DoS by analyzing traffic features like flow rate and duration and command injection attacks in robotic vehicles through CPU usage and data volume. Decision trees are valued for their transparency, aiding developers in real-time traffic analysis and instant alerts for anomalies. They are further enhanced by the Rule-Learning method, which develops a comprehensive rule set for classification, optimizing rule creation with human insights. [8]

C. Artificial Neural Networks

Artificial neural networks (ANNs) emulate how neurons in the brain communicate to interpret information, using mathematical equations (neurons) that adjust based on data to output relevant information. ANNs adapt by refining their algorithms with new data, making them practical for detecting novel and evolving online threats, including zero-day attacks. [8] Consequently, ANNs

are particularly suited for intrusion detection systems, showing promise in tackling issues like DoS attacks. Due to the high cost and resource demands, as AI in cybersecurity grows, its application is most valuable in largescale networks or systems, such as those in businesses, smart cities, or IoT environments. While AI excels in identifying ongoing attacks, it is best used alongside other preventative security measures.

D. Recurrent Neural Networks

Recurrent Neural Networks (RNNs) are specialized for processing sequential data, making them ideal for detecting evolving online threats like zero-day attacks by analyzing patterns over time. Their ability to retain information from previous inputs enhances their effectiveness in intrusion detection systems, particularly for complex attacks such as DoS. However, their high computational demands make RNNs most valuable in large-scale networks, such as businesses or IoT environments. While powerful in identifying ongoing threats, RNNs are best used alongside other security measures for comprehensive protection.

III. METHODOLOGY

This section provides a conceptional and systematic approach for the prediction of the attacks on IoT devices using ML, DL and RNN. The steps below explain the methodology in detail:

A. Dataset Reading

The RT-IoT2022 is acquired from the UC Irvine Machine Learning Repository [6] website which is a source for very reliable and detailed dataset that is widely used in IOT Security. Specifically, it is comprised of 83 features both quantitative and qualitative, with measurements obtained from normal IoT devices and under different attacks. The dataset contains all the range of frequency of twelve kinds of attack such as ARP Poisoning, DOS SYN Hping, NMAP scans.

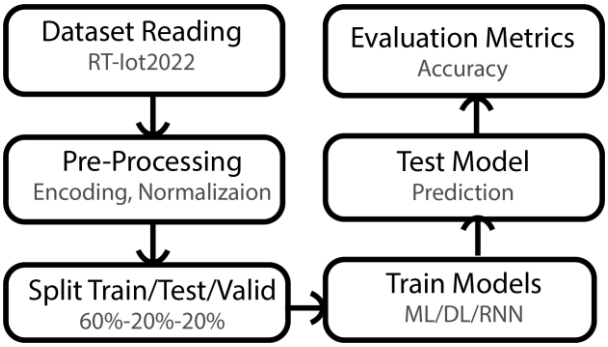


Fig. 3. Methodology

This dataset is particularly strong because of the real time attack simulation which makes it ideal for benchmarking of Intrusion Detection Systems (IDS). The RT-IoT2022 dataset has been employed in many recent studies such as Quantized Autoencoder (QAE) intrusion detection systems for anomaly detection in resource-scarce IoT devices [10]. Such studies prove the dataset useful in anomaly detection and resource-scarce scenarios, which proves its usefulness in the context of IoT security.

B. Pre-Processing

- Encoding: Out of all the prepared data, two categorical features were transformed with the one-hot encoding technique. This approach makes sure that data interpreted as numerical by the model is right without order relations which are not present.
- Normalization: To maintain uniform scaling across the features normalization was done. Since the scale of features differ greatly (some are as small as 0.31 while others as large as 20.06), we applied min-max normalization to get the values in the range of 0-1. This assists in enhancing model convergence if training through making certain that the gradient descent processes work uniformly with features.
- Data Quality Considerations: Numbers accuracy and credibility defines how well a machine learning algorithm works. About this subject, Sirisha et al (2023) points out that appropriate data pre-processing methods play a crucial role in improving data quality in IoT systems especially when dealing with datasets for multinational organizations. It is reaffirmed in this study that it is crucial to prescribe a proper set of preprocessing steps to achieve the required prediction accuracy [11].

C. Split Train/Test/Validation

The dataset is divided into training, validation, and testing sets in a 60%-20%-20% ratio. The training set is used to train the model, while the validation set is employed to tune hyper parameters and monitor the model's performance during training, helping to prevent overfitting. Finally, the testing set is reserved for evaluating the model's performance on unseen data, ensuring an unbiased assessment of its generalization capability.

D. Train Models

We employed a diverse approach by training multiple models to provide a comprehensive comparative analysis for IoT intrusion detection:

Decision Tree (ML): It is a well-known work for its good interpretability and high efficiency, which makes it one of the more suitable benchmarks. This provides it with its greatest strength – the ability to work on high-dimensional data and quickly identify attacks that are necessary given the ever-evolving IoT environment [12].

Deep Learning (DNN): Features of the presented sequential DNN include innovations such as adjusting learning rates depending on the achieved performance. High accuracy refers to precise calibration of solutions, and convergence in every cycle is faster, hence the benefit. Furthermore, incorporating state of the art dropout will reduce the issue of over fitting whereby

the model performs well on the training data but does poorly on the testing data, hence serves the purpose of addressing a typical problem in deep learning [13].

Recurrent Neural Networks (RNN): Based on best known Long Short-Term Memory (LSTM) architecture, our RNN is built to model sequential data which we believe is important in capturing temporal dependence crucial in microscope analysis of attack trajectory over time. In the context of IoT security, one of this architecture's unique features is its ability to retain information from previous inputs, which allows it to efficiently identify multi-stage attacks which differentiate our approach [14].

This diverse training approach not only provides significant information to the IoT security field but also strengthens the reliability and efficacy of the approaches used in the detection of intrusions. ML (Decision Tree), DL (Sequential), and RNN: Various learning architectures are utilized. Hyper parameter tuning in DL and RNN optimizes the configurations of these models, improving their learning capabilities.

E. Test Model

The trained models are first used to test on the validation set which allows tweaking of hyperparameters and ultimately avoid overfitting. This makes sure that before testing is carried out, every single model has been tuned to capture the best results. After tuning, models are evaluated for a separate, held-out benchmark to evaluate their effectiveness in real world. We use the confusion matrix to decompose the true positive, true negative, false positive and false negative findings of the model. This does not only reveal the strengths and limitations of a model but also the predictive attack performance against different attack classes.

Furthermore, the resulting model provides a classification report containing information referring to, among others precision, recall, F1-score. The assessment provided in this paper enables us to estimate the models based on their recommended approach in IoT security, in which it is crucial to minimize false negatives.

F. Comparative Evaluation

- Accuracy: An overall effectiveness of the model is calculated by determining the number or proportion of correct predictions with regard to the total. This gives a general idea of how often the model's predictions are correct.
- Precision: The metric describes the correctness of the model in identifying positive cases and can be measured as the proportion of predicted positives that are really true. This is of utmost relevance in situations where there is a high price for misclassification because of false positives.
- Recall (Sensitivity): Recall refers to the ability of the model to recognize the positive cases of correctness, meaning a share of actual positives that the model can catch with respect to those that have been missed out. It becomes especially important when a priority is put on minimizing false negatives.
- F1-Score: F1-score combines precision and recall by considering their harmonic mean. This makes F1 a balanced measure that is quite useful in those situations where class distribution is not even in a dataset, and it's useful when both precision and recall need to be balanced against each other.
- Confusion Matrix: A confusion matrix is a tabular representation of the various attributes used to explain the performance of a classification model, such as actual classes vs predicted classes. It gives a detailed report of model predictions. This matrix is the basic one for understanding what kind of mistakes the model is making and builds the foundation for other metrics like precision, recall, and F1-score.

G. Results and Findings

The RT-IoT2022 dataset comprises 83 features, including both numerical and categorical attributes, making it crucial for analyzing machine learning models in IoT intrusion detection. It includes 12 distinct attack types, such as MQTT Publish, ARP Poisoning, and various NMAP scans. What distinguishes this dataset is its comprehensive range of attacks, providing an excellent benchmark for developing and evaluating intrusion detection systems (IDS) in IoT environments. With 123,117 examples stored in int64, float64, and object formats, the dataset offers versatility in data processing.

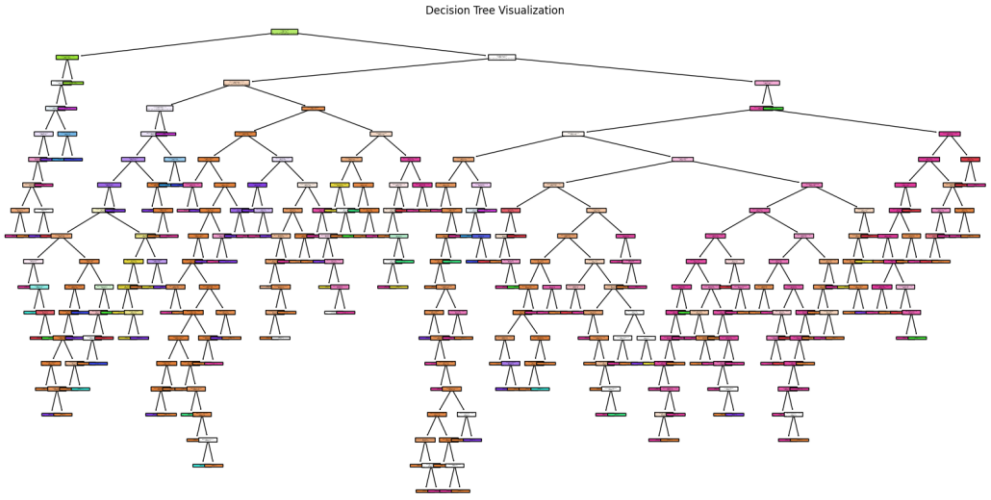


Fig. 4. Architecture of Decision Tree

Table 1
Evaluation Metrics for DT

Metric	Test	Validation
Accuracy	0.9971	0.9980
Precision	0.9971	0.9980
Recall	0.9971	0.9980
F1 Score	0.9971	0.9980

Table 2
Overall Evaluation of Decision Tree

Metric	Value
Training Time (s)	1.151318
Validation Accuracy	0.997969
Test Accuracy	0.997076

Table 3
Evaluation Metrics for DNN

Metric	Test	Validation
Accuracy	0.9956	0.9955
Precision	0.9956	0.9955
Recall	0.9956	0.9955
F1 Score	0.9955	0.9953

Table 4
Overall Evaluation of RNN

Metric	Value
Training Time (s)	3916.0432
Validation Accuracy	0. 9962
Test Accuracy	0. 9965

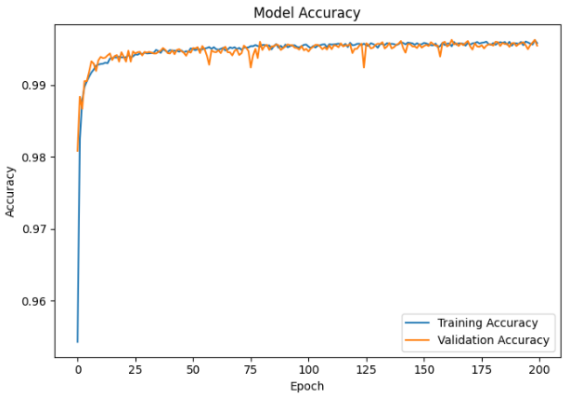


Fig. 5. Model Accuracy of DNN

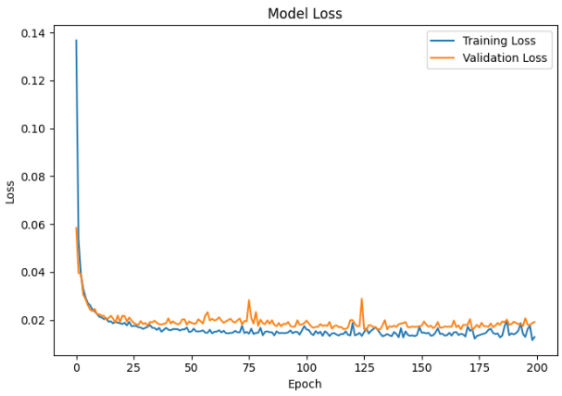


Fig. 6. Model Loss of DNN

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 128)	11,904
dropout (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 64)	8,256
dropout_1 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 12)	780

Total params: 62,822 (245.40 KB)
Trainable params: 20,940 (81.80 KB)
Non-trainable params: 0 (0.00 B)
Optimizer params: 41,882 (163.61 KB)

Fig. 7. Model Architecture of DNN

Layer (type)	Output Shape	Param
lstm (LSTM)	(None, 1, 128)	113,1
dropout_2 (Dropout)	(None, 1, 128)	
lstm_1 (LSTM)	(None, 64)	49,4
dropout_3 (Dropout)	(None, 64)	
dense_3 (Dense)	(None, 12)	

Total params: 490,022 (1.87 MB)
Trainable params: 163,340 (638.05 KB)
Non-trainable params: 0 (0.00 B)
Optimizer params: 326,682 (1.25 MB)

Fig. 8. Model Architecture of RNN

Table 5
Overall Evaluation of DNN

Metric	Value
Training Time (s)	1713.2822
Validation Accuracy	0. 9955
Test Accuracy	0. 9956

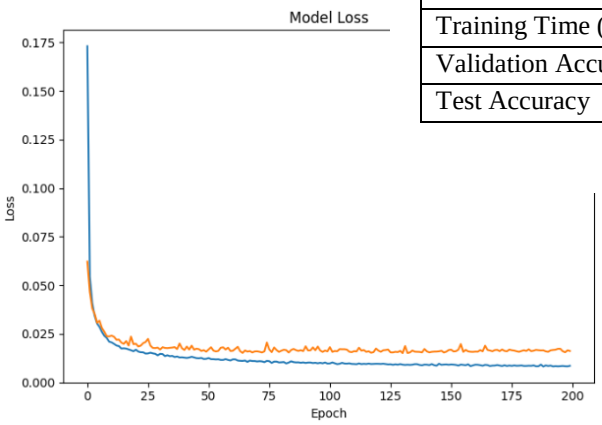


Fig. 9. Model Loss of RNN

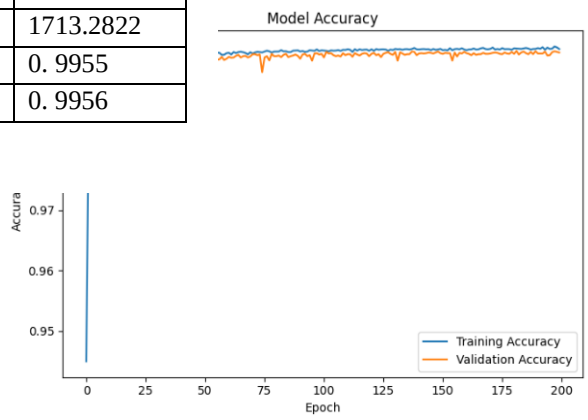


Fig. 10. Model Loss of RNN

IV. CONCLUSION AND FUTURE WORK

The proposed Decision Tree has outperformed the Deep Neural network (DNN) and Recurrent Neural Network (RNN) model for RT-IoT classification and prediction, having accuracy 99.80%, F1-score 99.80%, precision 99.80%, & recall of 99.80%. This model provides the highest figures for all the metrics, which sees it produce results in a 3.2 second time frame, making it extremely efficient.

It is important to remember that despite these advantages, the Decision Tree struggles with quite significant issues at scale. It can however be seen that as the RT-IoT dataset grows and becomes even larger, such basic models as the Decision Tree may start to falter. While, on one hand, DNN and RNN are effective to learn from large datasets since they can learn representation at different levels of abstraction and sequential order magnitude respectively.

Analyzing future perspectives, it can be concluded that studying other models – for example, the intermediate between DNN and RNN – is even more effective in terms of both, accuracy and training time. These innovations are necessary in managing the advanced issue complexity in IoT security and in enhancing the future possibilities of IDS. Finally, this research points to the fact that more efforts should be made to develop new methods of machine learning to cope with constantly expanding and diversifying networks of IoT systems.

Table 6
Comparative Evaluation of Models

Metric	Decision Tree	DNN	RNN
Test Accuracy	0.997076	0.9956	0.9965
Validation Accuracy	0.997969	0.9955	0.9962
Training Time (s)	1.151318	1713.2822	3916.0432

V. REFRENCES

[1] R. A. R. A. Mouha et al., "Internet of things (iot)," Journal of Data Analysis and Information Processing, vol. 9, no. 02, p. 77, 2021.

[2] A. A. Laghari, K. Wu, R. A. Laghari, M. Ali, and A. A. Khan, "Retracted article: A review and state of art of internet of things (iot)," Archives of Computational Methods in Engineering, vol. 29, no. 3, pp. 1395–1413, May 2022, retracted article. [Online]. Available: https://doi.org/10.1007/s11831-021-09622-6

[3] A. Giannaros, A. Karras, L. Theodorakopoulos, C. Karras, P. Kranias, N. Schizas, G. Kalogeratos, and D. Tsolis, "Autonomous vehicles: Sophisticated attacks, safety issues, challenges, open topics, blockchain, and future directions," Journal of Cybersecurity and Privacy, vol. 3, no. 3, pp. 493–543, 2023. [Online]. Available: https://www.mdpi.com/2624-800X/3/3/25

[4] Y. Abbasi, D. Kar, N. D. Sintov, M. Tambe, N. Ben-Asher, D. Morrison, and C. Gonzalez, "Know your adversary: Insights for a better adversarial behavioral model." in CogSci, 2016.

[5] M. M. Raikar and S. M. Meena, “Ssh brute force attack mitigation in internet of things (iot) network : An edge device security measure,” in 2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC), 2021, pp. 72–77.

[6] B. S. and R. Nagapadma, “RT-IoT2022 ,” UCI Machine Learning Repository, 2024, DOI: https://doi.org/10.24432/C5P338.

[7] M. Kuzlu, C. Fair, and O. Guler, “Role of artificial intelligence in the internet of things (iot) cybersecurity,” Discover Internet of Things, vol. 1, no. 1, p. 7, Feb. 2021. [Online]. Available: https://doi.org/10.1007/s43926-020-00001-4

[8] S. Zeadally, E. Adi, Z. Baig, and I. A. Khan, “Harnessing artificial intelligence capabilities to improve cybersecurity,” Ieee Access, vol. 8, pp. 23 817–23 837, 2020.

[9] Chen, Chien-Ying & Hasan, Monowar & Mohan, Sibin. (2017). Securing Real-Time Internet-of-Things. Sensors. 18. 10.3390/s18124356.

[10] B S, Sharmila & Nagapadma, Rohini. (2023). Quantized autoencoder (QAE) intrusion detection system for anomaly detection in resource-constrained IoT devices using RT-IoT2022 dataset. Cybersecurity. 6. 10.1186/s42400-023-00178-5.

[11] N. Sirisha, M. Gopikrishna, P. Ramadevi, Raveendranadh Bokka, K.V.B. Ganesh, M. Kalyan Chakravarthi, IoT-based data quality and data preprocessing of multinational corporations, The Journal of High Technology Management Research, Volume 34, Issue 2, 2023, 100477, ISSN 1047-8310, https://doi.org/10.1016/j.hitech.2023.100477. (https://www.sciencedirect.com/science/article/pii/S1047831023000275)

[12] Mittal, S., Mishra, A. K., & Tripathi, N. (2021). Integration of machine learning with IoT. Ilkogretim Online - Elementary Education Online, 20(3), 3726-3734. http://ilkogretim-online.org
https://doi.org/10.17051/ilkonline.2021.03.388

[13] Ehsan Tanghatari, Mehdi Kamal, Ali Afzali-Kusha, Massoud Pedram, Distributing DNN training over IoT edge devices based on transfer learning, Neurocomputing, Volume 467, 2022, Pages 56-65, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2021.09.045. (https://www.sciencedirect.com/science/article/pii/S0925231221014235)

[14] Muder Almiani, Alia AbuGhazleh, Amer Al-Rahayfeh, Saleh Atiewi, Abdul Razaque, Deep recurrent neural network for IoT intrusion detection system, Simulation Modelling Practice and Theory, Volume 101, 2020, 102031, ISSN 1569-190X, https://doi.org/10.1016/j.simpat.2019.102031. (https://www.sciencedirect.com/science/article/pii/S1569190X19301625)

A net-metering billing model for economically stability of the utility grid

Osama Majeed Butt
Institute of Electrical Electronics & Computer Engineering
University of the Punjab
Lahore, Pakistan
osama.ee@pu.edu.pk

Usama Arif
Institute of Electrical Electronics & Computer Engineering
University of the Punjab
Lahore, Pakistan
usamaarifbhatti594@gmail.com

Ali Tahir
Institute of Electrical Electronics & Computer Engineering
University of the Punjab
Lahore, Pakistan
alit91940@gmail.com

***Abstract*—Increasing trends towards installation of solar PV as a renewable source has become popular with injection of energy into grid. Beside its advantage it raised many issues including technical as well as economic problems. Different models of billing this 'import' and 'export' of energy units are in use to calculate charges of electricity. Grids are becoming economically unstable due to absence of limitations or putting conditions on injection of energy by distribution generators. This didn't helpful for grid network in terms of earning and hence becoming difficult for distribution companies to maintain their grid. This proposed study will certainly help the grid and its distribution companies in sustaining economically. Also it will encourage the DGs for self-consumption of energy and discourage for installing over capacity PV system.**

Keywords—net-metering, billing, grid, distribution generator (DG), solar

I. INTRODUCTION

The increasing global demand for sustainable energy has sparked widespread interest in photovoltaic (PV) systems, positioning them as a vital alternative to traditional energy sources [1]. As more countries embrace solar energy, there is a growing need to develop efficient billing models that not only incentivize renewable energy adoption but also ensure long-term financial sustainability for both energy providers and consumers [2]. In countries like Pakistan, the energy sector is grappling with severe challenges, including high electricity tariffs, frequent energy shortages, and an unreliable power grid. These issues make the transition to renewable energy sources, particularly solar power, essential for fostering economic growth and stability [3].

Net metering and net billing are two significant policy frameworks that encourage the installation of grid-connected PV systems owned by electricity consumers. These policies play a crucial role in promoting the use of renewable energy by offering consumers incentives to generate their own electricity while remaining connected to the public grid. Under the net metering model, the electricity generated and injected into the grid by a consumer is valued at the same rate as the electricity they consume from the grid (retail price). This allows consumers to offset their energy costs more effectively. In contrast, net billing values the electricity fed back into the grid at a lower rate, often referred to as the wholesale or "avoided cost" price, which is less than the rate at which energy is consumed from the grid [3, 4].

Net metering was first introduced in the United States during the 1980s and has since gained widespread adoption. Nearly all U.S. states now have net metering policies in place, albeit with various configurations. Some states offer simple net metering, while others include options for buy-back, rolling credit, or a combination of buy-back and rolling credit [3]. Over time, many other countries have adopted similar net metering and net billing policies to encourage the use of solar energy [5]. These frameworks have become instrumental in the global push towards renewable energy as they offer clear financial incentives for consumers to invest in small-scale PV systems.

In recent years, Net Energy Metering (NEM) has garnered substantial interest from a range of energy stakeholders, including regulatory authorities, utility companies, and policymakers. NEM represents an attractive investment opportunity, particularly for small-scale PV generation in the residential sector [6]. This system is specifically designed for prosumers — individuals who simultaneously produce and consume energy — and enables them to offset the energy they consume with the energy produced by their own PV systems, such as rooftop solar panels. A key feature of NEM is the ability to consume and produce energy at different times, which distinguishes it from other frameworks like self-consumption or tax reductions for self-generated electricity.

Under NEM, the financial charge a prosumer incurs is based on the "net" amount of energy consumed and produced over a specified netting period, as outlined in the contract. If the amount of energy generated by the prosumer is less than the amount consumed (referred to as negative netting), the consumer may be charged either for the net energy used or for additional charges to cover grid-related services such as transmission and distribution. In cases where the prosumer generates more energy than they consume (positive netting), NEM policies may allow for compensation of the excess energy at varying rates, or in some cases, no compensation at all [1].

As many countries approach grid parity — where the cost of producing solar energy becomes competitive with conventional energy sources [4] — the concept of self-consumption of PV-generated electricity is gaining considerable traction [6]. Self-consumption refers to the immediate use of the electricity generated by a consumer's PV system, typically within a 15-minute timeframe [7]. This has traditionally been the standard definition of self-consumption, but it poses challenges for residential prosumers, whose energy production and consumption patterns often do not align.

Due to this mismatch between production and consumption, compensating prosumers for surplus energy that is fed into the grid can be a complex process. Various support mechanisms have been introduced to address this challenge, with net metering and net billing emerging as the most prominent policy solutions . These policies offer prosumers a way to earn credits or financial compensation for the excess energy they supply to the grid, helping to reduce their overall energy costs.

At the global level, prosumers are anticipated to play a significant role in the expansion of PV installed capacity [8]. As this trend continues, many countries are shifting away from traditional Feed-in Tariff (FIT) policies, which guarantee a fixed payment for electricity produced by renewable energy sources, in favor of net metering and net billing schemes [9, 10]. These newer frameworks are seen as more sustainable and flexible, as they allow for the integration of renewable energy into the grid without placing an undue financial burden on energy providers.

In regulated net metering and net billing systems, the surplus electricity that prosumers feed into the grid is often rewarded with credits. These credits can be applied to offset future energy consumption, providing a tangible financial incentive for individuals to invest in renewable energy systems [9]. This shift from FIT to net metering and net billing represents a broader global movement towards more efficient and consumer-friendly renewable energy policies. The success of these policies will be crucial in ensuring that the growing adoption of PV systems continues to contribute meaningfully to both environmental sustainability and economic growth.

In this study a billing model is being proposed by keeping in view the objective of encouragement of self-consumption of energy by DGs, while discouraging of installation of the overcapacity or beyond need system in order to avoid and minimize the injection of energy into utility grid by DGs. Also it will help utility grid to have a sustainable economic situation while minimum liability on distribution generators (Solar Installations).

II. METHODOLGY

A forecasted model is formulated to determine and analysis the economical reflection of the power grid. This prediction is done to ease the financial pressure that could be face by the distribution companies and utility grid due to injection of power for domestic Distributed Generators (solar power producer residents). This model is termed as '90/10' model. As per model distributed generators (DGs) will retain 90% of exported units for their own utilization while power distribution companies will retain 10% of the total units exported by each residence in each month. Furthermore, in deemed months (Nov, Dec, Jan) when utilization is less as compared to other months of the year, share of the distribution companies for the exported units will increase to 30% of the total exported units. Additionally, a fix charge per month will be paid by each resident DG at the rate of Rs. 0.2/- per watt PV installed capacity. The tariff of 'Import' and 'Export' units will be determined on the rates of that will be revised as per government/competent authority. Credits units (if any) for previous months will be converted and shown as credited amount after three month of cycle. Credited amount (if any) can only be used for adjustment of bill of extra imported/used units by consumer and cannot be cashed at any time.

This model doesn't use Time-of-use (TOU) methodology as 'peak' and 'off-peak' tariff is not account for during calculations for trading import and export units. For current study, utility grid was considered of University of the Punjab and DGs were selected that were connected with this power grid system by using bi-directional meters.

In this study, it was observed from data collection that maximum exported units that were injected into grid by a DG after self-consumption was 60% while minimum injection was 40%. Therefore, this model was run for three scenarios of injection of units based on data collection i.e. 40%, 50% and 60%. Also the tariff was determine to be Rs.60/- for import units while Rs.21.42/- for export unit for a DG. This tariff was already set based on current tariff rate approved by competent authorities. Equation (i) is used for calculation of profitability of the normal months while equation (ii) is used for calculation for the deemed months i.e. Nov, Dec, Jan.

$$\sum C_{fix} + (U_{export} * 0.10 * 60) \quad \text{----------} \quad \text{(i)}$$

$$\sum C_{fix} + (U_{export} * 0.30 * 60) \quad \text{----------} \quad \text{(ii)}$$

Similarly, reflection of liability for the consumer can be predicted using equation (iii) for normal months while equation (iv) for deemed months

$$\sum C_{fix} + (U_{export} * 0.10 * 21.42) \quad \text{----------} \quad \text{(iii)}$$

$$\sum C_{fix} + (U_{export} * 0.30 * 21.42) \quad \text{----------} \quad \text{(iv)}$$

Here C_{fix} = fix charges, U_{export} = exported units by respective DG

III. Results & Discussion

Targeted sample distribution network (Punjab University) was classified into three categories based on the capacity of the install solar system i.e. 6kW, 8kW and 10kW. Furthermore, after analyzing the behavior of the DGs these cite were studied on three different patterns of the total exported units to network after being self-consumption.

A. DG installed with 6kW solar System

As per the behavior and data collected from the user of 6kW DG site, results of the proposed billing model is shown in figure 1. According the model, fix charges will be implemented Rs.1200/- per month. With a 40% of exported units an average of Rs.1600/- per regular month is earned by the grid while in deemed months this average is increased to 3200/- per month. A total of Rs.40000/- (approx.) annually is earned by grid network with 40% exported units of total produced. This profitability is in terms of electricity unit deduction and fix charges. While consumer have a reflection of approximately Rs.23000/- liability annually. This profitability and liability increase to Rs.45000/- and Rs.25000/- respectively when exported units were 50% of the total produced units after being self-consumed by DG.

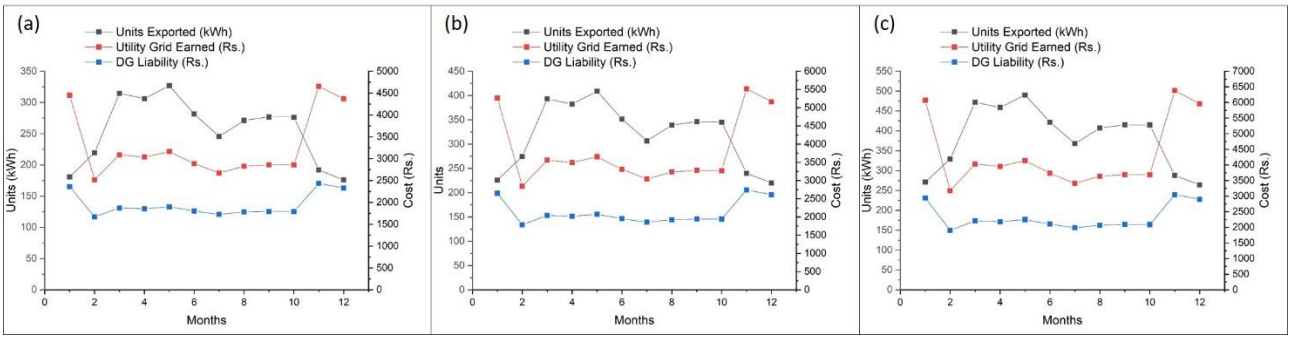


Fig. 1. Data for units production, profitibality of the distribution network and DG's liability of 6kW installed solar system with (a) 40% exported units; (b) 50% exported units; (c) 60% exported units

B. DG installed with 8kW solar System

Based on user behavior and data collected from the 8kW DG site, results are shown in Figure 2. Under this billing model, a fixed charge of Rs.1600 per month is applied. With 40% of the total generated energy being exported, the grid earns an average of Rs.3800 per month, increasing to Rs.6000 during peak months. Annually, the grid earns approximately Rs.52000 but for the consumer, this creates an annual liability of around Rs.30000. If the exported energy rises to 50% of total generation after self-consumption, the grid's earnings increase to Rs.60000 annually, while the consumer's liability grows to Rs.35000.

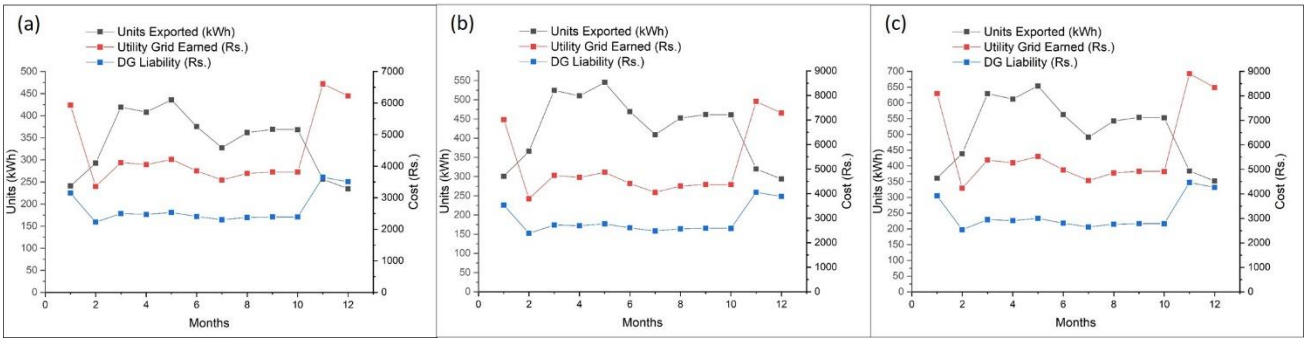


Fig. 2. Data for units production, profitibality of the distribution network and DG's liability of 8kW installed solar system with (a) 40% exported units; (b) 50% exported units; (c) 60% exported units

C. DG installed with 10kW solar System

User behavior and data from the 10kW DG site are summarized in Figure 3. For 10kW site fixed charge includes Rs.2000 per month. With 40% of the total generated energy from 10kW system is being exported to grid which earns an average of Rs.4800 per month for grid, and that increasing to Rs.7500 during peak months. Annually, the grid earns approximately Rs.65000, while the consumer faces a liability of around Rs.38000. If the exported energy rises to 50% of total generation after self-consumption, the grid's annual earnings increase to Rs.76000, and the consumer's liability grows to Rs.42000. Similarly if DG export approximately 60% of its total produced energy than grid will have a profit of Rs.85000 while DG will face a liability of Rs.46000 annually.

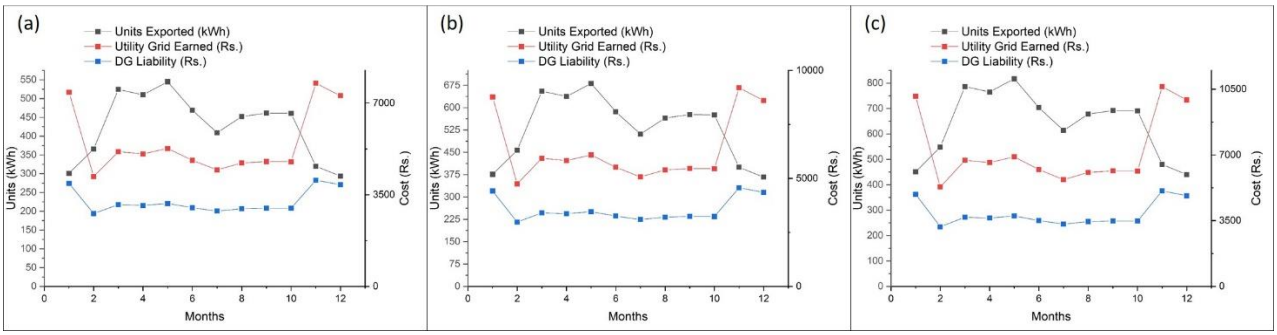


Fig. 3. Data for units production, profitibality of the distribution network and DG's liability of 10kW installed solar system with (a) 40% exported units; (b) 50% exported units; (c) 60% exported units

IV. CONCLUSION

This billing model could be beneficial for having an insight economic situation of net metering billing mechanism and will help in predicting the financial outcomes. Currently studies shows that this billing model will help in earning the grid approximately 44% more than the financial liabilities to distribution generators while no extra costing will have to bear by the DGs. Directly DGs will not suffer most of the cost as deduction is done in terms of energy unit. This is difference profitability and liability of utility grid and DGs is due to difference in the unit exchange rate. This will eventually help is sustaining the utility grid economically and encourage the solar user to self-consume the energy produce from solar by minimizing the export to grid.

REFERENCES

[1] C. E. Casillas and D. M. Kammen, "The energy-poverty-climate nexus," *Science*, vol. 330, pp. 1181-1182, 2010.

[2] M. Moezzi, A. Ingle, L. Lutzenhiser, and B. O. Sigrin, "A non-modeling exploration of residential solar photovoltaic (PV) adoption and non-adoption," National Renewable Energy Lab.(NREL), Golden, CO (United States)2017.

[3] M. Irfan, Z.-Y. Zhao, M. Ahmad, and M. C. Mukeshimana, "Solar energy development in Pakistan: Barriers and policy recommendations," *Sustainability*, vol. 11, p. 1206, 2019.

[4] M. Asim, A. Qamar, A. Kanwal, G. M. Uddin, M. M. Abbas, M. Farooq, M. Kalam, M. Mousa, and K. Shahapurkar, "Opportunities and challenges for renewable energy utilization in Pakistan," *Sustainability*, vol. 14, p. 10947, 2022.

[5] J. Khan and M. H. Arsalan, "Solar power technologies for sustainable electricity generation–A review," *Renewable and Sustainable Energy Reviews*, vol. 55, pp. 414-425, 2016.

[6] M. Alipour, H. Salim, R. A. Stewart, and O. Sahin, "Residential solar photovoltaic adoption behaviour: End-to-end review of theories, methods and approaches," *Renewable Energy*, vol. 170, pp. 471-486, 2021.

[7] M. F. Aziz and N. Abdulaziz, "Prospects and challenges of renewable energy in Pakistan," in *2010 IEEE International Energy Conference*, 2010, pp. 161-165.

[8] R. Zafar, A. Mahmood, S. Razzaq, W. Ali, U. Naeem, and K. Shehzad, "Prosumer based energy management and sharing in smart grid," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 1675-1684, 2018.

[9] B. Knopf, P. Nahmmacher, and E. Schmid, "The European renewable energy target for 2030–An impact assessment of the electricity sector," *Energy policy*, vol. 85, pp. 50-60, 2015.

[10] A. S. Alahmed and L. Tong, "Integrating distributed energy resources: Optimal prosumer decisions and impacts of net metering tariffs," *ACM SIGENERGY Energy Informatics Review*, vol. 2, pp. 13-31, 2022.

A Load Classification Strategy using NILM and DNN for Potential Demand-Side Management

Arsh Alam Bhatti
Institute of Electrical, Electronics and Computer Engineering, University of the Punjab
Lahore, Pakistan
arshalambhatti@gmail.com

Fazeela Arshad
Institute of Electrical, Electronics and Computer Engineering, University of the Punjab
Lahore, Pakistan
fazeelairshad07@gmail.com

Abdullah Al Sahaibi
Institute of Electrical, Electronics and Computer Engineering, University of the Punjab
Lahore, Pakistan
abdulelahalsuhaibi45@gmail.com

Akbar Ali Khan
Institute of Electrical, Electronics and Computer Engineering, University of the Punjab
Lahore, Pakistan
akbar.ee@pu.edu.pk

Kamal Shahid
Institute of Electrical, Electronics and Computer Engineering, University of the Punjab
Lahore, Pakistan
kamal.ee@pu.edu.pk

***Abstract*— Uncoordinated and unplanned increase in electricity demand is a critical concern in recent times owing to increasing population and appliance utilization. Significant focus has been given on optimizing load patterns for appliances and capitalizing the potential for savings in domestic energy management. This paper develops a low-cost demand-side management system for residential application through smart energy meters, combined with non-Intrusive load monitoring (NILM) and machine learning for accurate load disaggregation. This paper presents a real-time consumption-based dynamic pricing algorithm, exploiting the use of deep neural networks (DNN) for the classification of essential and non-essential loads with the help of real-time collected datasets. The system provides live energy monitoring through Modbus RTU, RS485 protocols, and a Postgre SQL database, which provides data visualization on a Power BI dashboard highlighting real-time advice on optimization of the consumed energy. The proposed approach demonstrates an effective demand response (DR) mechanism, shifting electricity consumption to off-peak hours throughout out the day without reducing overall energy use, hence optimizing the overall load curve metrics and enhancing energy efficiency.**

Keywords*— *NILM, Demand side Management, Deep Neural Networks, Smart Energy Meters, Energy Efficiency, Real Time Energy Monitoring

I. INTRODUCTION

Due to the increased population and growing consumption of electrical appliances, the requirement for electricity has dramatically escalated, especially in the countries like Pakistan. Some reports have shown a notable difference in generation capacity and demand leading to blackouts and load-shedding. This is posing formidable challenges both for power grids as well as for energy suppliers. On the other hand, skewed load utilization patterns among consumers causes significant economic challenge due to the difference of cumulative energy production and installed capacity. Classic energy management systems are normally unable to cope up with the intertemporal variance in demand, which leads to peak load problems and higher energy costs for consumers. Optimization of load patterns and appliance utilization in domestic environments is essential for smooth maintenance and cost savings of the grid. Thus, optimization of these factors forms the prime focus of researchers and industry professionals. The process of improving the load usage patterns and improving load curves is commonly known as Demand Side Management (DSM).

DSM makes consumers, through real-time load monitoring, load control and incentives/penalties from utility, shift their levels of energy consumption at the right time. The major goal is shifting the load from peak times to off-peak times, while fulfilling the overall energy demand and suitably ensuring the consumers' comfort. However, the effectiveness of DSM relies on efficient load patterns detection and proper mechanism of identifying the nature of loads. So that shifting of only those loads can be strategized which does not disrupt consumers' necessities. In this regard, advances in smart metering, and corresponding learning mechanism of load natures using machine learning tools can potentially play a pivotal role.

This work develops a low-cost, real-time load identification mechanism for residential consumers using smart energy meters. The meters use the exchange of data communication protocols such as Modbus RTU and RS485. Corresponding data is visualized by an interactive Power BI dashboard, and stored in a PostgreSQL database. The database and visualization is demonstrated to assist the consumer realize the energy consumption and gives hints to optimize this load.

The proposed mechanism classifies the appliances as essential and non-essential, using Nonintrusive Load Monitoring (NILM) algorithm and Deep neural network (DNN). The classification can be used to define dynamic pricing algorithms to encourage users to shift non-essential loads during off-peaking hours, hence flattening the load curve.

The rest of this paper is organized as follows: Section II, which presents a detailed literature review comparing the proposed work with existing solution in area. Section III describes the methodology and provides a step-by-step account towards how the solution was developed. The results section is presented in Section IV to describe what was obtained. Section V concludes the paper with highlights of main takeaways from this work.

II. LITERATURE REVIEW

DSM methods has been well explored in reducing energy consumptions and changing demands from peak to off-peak hours. [1]. Latest developments also depict those intelligent systems with real-time data inputs for load shifting do indeed result in a successful automation process. [2] DSM combined with real time feedback and dynamic pricing yields significant energy savings. The fact, however, is that traditional DSM still requires human intervention at times, which lowers the efficiency.

NILM breaks up household electrical consumption into individual appliances without requiring a sensor for every device, which reduces costs and complexity. Hart 1992 [3] was the pioneer of NILM in terms of signal processing techniques applied to determine appliances by their power signatures. NILM has been implemented in a large number of smart energy management systems in order to gain detailed insights into residential energy usage. [4] A survey has shown that NILM algorithms, indicating that advanced machine learning models have become considerably more accurate compared to early ones in the domain of load disaggregation. More recent highlights has shown deep learning algorithms in the form of convolutional neural networks (CNN) and Long Short-Term Memory (LSTM) networks are to enhance the accuracy of NILM [5]. This combines NILM with Machine Learning techniques in real-time classification of the other loads that would improve feedforward accuracy and optimize the usage of energy.

Machine learning is utilized in the optimization of energy systems, including load disaggregation and dynamic pricing. A study took into account the role of machine learning in the residential energy forecast, where ANNs and SVMs may help in the enhancement of energy consumption prediction [6]. DNNs are currently trending in load classification and has been applied for residential appliances classification, with high accuracy in load disaggregation [7].

Dynamic pricing forms part of the DSM systems which comprise incentives and rewards for the customers to adjust their energy consumption in direct accordance to price signals at any time. Dynamic pricing is acclaimed to reduce peak loads by as much as 30% depending on pricing model and consumer involvement [8]. In addition, Demand Response management (DRM) is needed for DSM to have improved efficiencies [9]. This work utilizes dynamic pricing in the form of a real-time consumption-based algorithm, hence prompting users to shift some of their non-essential loads to off-peak periods. The proposed system integrates DNN based load classification with dynamic pricing, encouraging users to better optimize the energy consumption levels without losing the overall consumption.

Real time consumption monitoring is quite vital for a good DSM technique. Smart meters and monitoring systems are handled in such a manner that they can contribute to helpful data towards developing the consumption pattern and adjusting appliance use. [10] Real time data is of paramount importance for optimizing consumption, most specifically in the residential sector, where consumer behavior varies greatly. [11] Integration of real time visualization tools along with dynamic pricing have a considerable effect on the behavior of the consumers, saving 20% of the consumed energy. Based on these, this paper extends previous works in the literature on demand side management (DSM), non-intrusive load monitoring (NILM), and load disaggregation using machine learning with various contributions that distinguish it. Rather than classic DSM schemes that need human help for solutions, an intense DNN system employed for the real time differentiation between essential and non-essential loads to carry out proficient load disaggregation. Moreover, the real time dynamic pricing algorithm, which is integrated into a Power BI dashboard to offer instantaneous feedback through the trained model, convinces users to move non critical loads towards off-peak times. This real time approach enhances energy efficiency and demand response capabilities that optimize the load curves without compromising the user comfort.

III. METODOLOGY

The flow of steps in this work comprises of systematized installation of smart meters in residential facilities, their data collection, and corresponding energy consumption analysis. Installed Smart meters installation communicated the load parameters in the MODBUS RTU protocol. This data is collected, processed and fed into machine learning models for disaggregation of load use. The learning outcome is then used to detect the number and nature of appliances in runtime data and its display on dashboard. The detailed steps are as follows:

1. The SME 104D Smart Meter was installed to measure voltage, current, power and energy levels inside the residence.
2. USB to RS-485 converter was used to transmit real-time data using the MODBUS RTU protocol. The parameters for reliable data communication were also configured.
3. The captured data was stored in the HMI memory and then in the PostgreSQL database. The pre-processing Steps are
    - Data Cleaning: This includes correcting the erroneous entries.
    - Normalization: Scaling the numerical data to improve the performance of the model.
    - Feature Extraction: Meaningful features to be extracted, such as mean power, peak load time.

4. Classification of the usage data, using a wide variety of machine learning algorithms with a focus on Deep Neural Networks Training Process
    - Dataset Split: Training 70% and validation/testing 15% each
    - Loss Function and Optimizer: Categorical cross-entropy and Adam optimizer.
    - Performance Metrics: Accuracy, predicted class probabilities, confusion matrix.

5. Real-time dynamic pricing model based on live data from PostgreSQL.
    - DNN to distinguish between critical and non-critical loads.
    - System recommendation to shut down the loads that are not critical during peak pricing.
    - Possible savings with the loads' reduction

6. Provides an energy consumption Power BI dashboard in real-time and costs. It reflects current usage and gives an estimate of what the future costs will be.

IV. RESULTS AND DISCUSSIONS

A. Smart Meter Installation and Communication

The SME 104D smart energy meter was installed as shown in **Fig 1** and used to measure the key electrical parameters, namely, voltage, current, power, and energy, in a residential setting in **Fig 3**. Various home appliances, such as an air conditioner (AC), personal computer (PC), resistive bulb, and washing machine, were used as test loads. Data communication between the meter and a PC was achieved via a USB to RS-485 converter, which transmitted data in real time through the MODBUS RTU protocol. The configuration of the communications parameters, including the device address, baud rate, stop bits, and parity, on the meter helps establish reliable and efficient data transfer.

Fig 1: SME 104 Smart meter collecting data

Data Collected for One Month (August)
Using SME 104 D Smart Meter and RS485 to USB Converter

Appliance	Power (W)	Weekly Usage (Hours)
Air Conditioner (AC)	1300	42
Washing Machine	3500	3
Bulb	100	119
PC	400	21

Fig 2: Real Time Data collection Duration

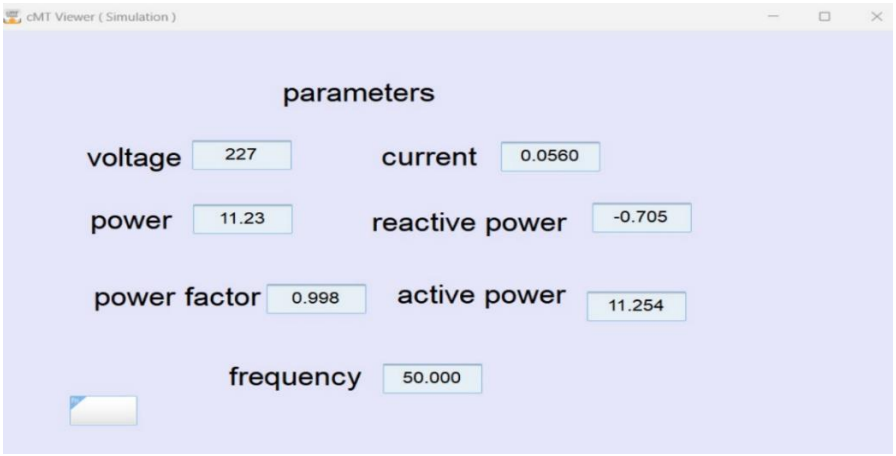


Fig 3: Collected features of appliances

B. Machine Learning for Load Disaggregation

The energy usage data acquired from the SME 104D smart energy meter was classified and analyzed using a variety of machine learning algorithms. Since DNNs are adept at capturing complex data patterns and yielding high classification accuracy, most attention was given to these.

The DNN model was trained on a labeled dataset; each entry of data was tagged with either being an essential load or not. Training was primarily divided to

• **Dataset Splitting:** the dataset was divided into training, validation, and testing subsets to test the performance of the model. In normal cases, 70 percent of the data would be used for training, 15 percent for validation, and 15 percent for testing.

• **Epochs and Batch Size:** Training the model was done by running it on several epochs, an epoch being simply one run through the training dataset. The batch size was decided to keep track of samples processed before the model updated its weights. The DNN was tested on the test dataset after it has been trained to evaluate its performance given in **Fig 4.**

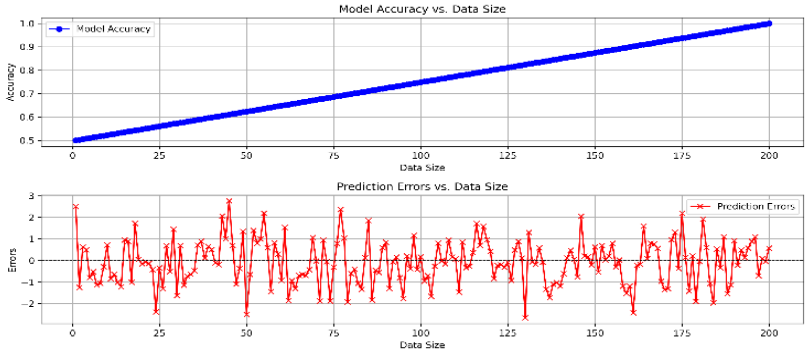


Fig 4 : Model Error Losses while handling dataset

This is the accuracy rate of number of correct predictions made by the model out of the total number of predictions. Similarly, the appropriate loss function which is categorical cross entropy is used for the training of the model. That has been deployed for evaluation of model performance. In **Fig 5** The Adam optimizer was used to update the weights of the model according to the computed gradients during back-propagation.

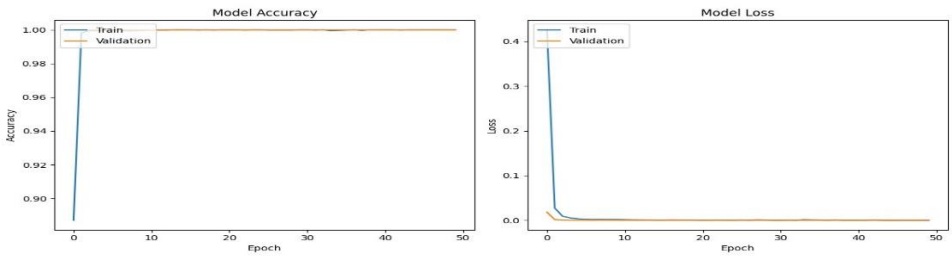


Fig 5: Model loss and Accuracy

The model's predicted class probabilities for every class illustrate the confidence of classifying essential and non-essential loads. The confusion matrix provides detailed classification result for each class **Fig 7**, and it depicts the true positive and false positive rate for every class.

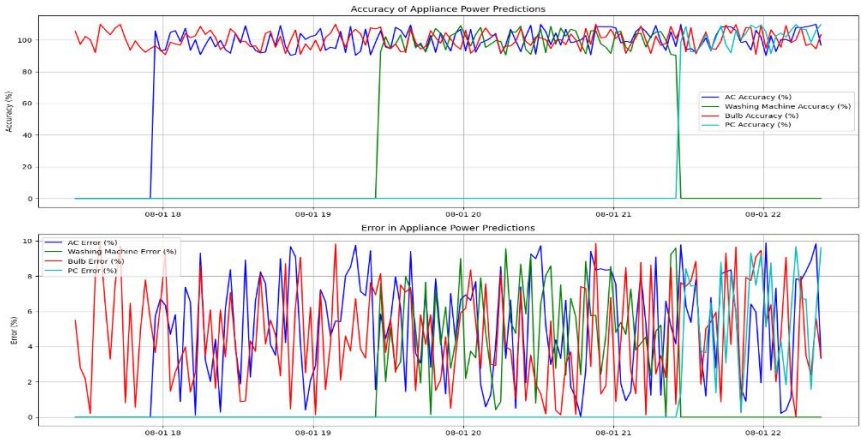


Fig 6: Probability and acuuracy for predicting appliances

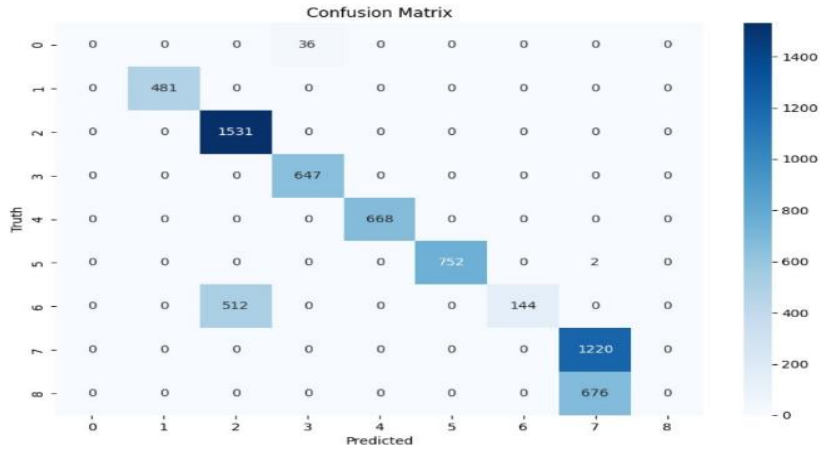


Fig 7: Confusion Matrix

First, the DNN was trained on the pattern of signatures of individual appliances. The different appliances had different characteristics with regard to energy consumption, thus allowing for learning and accurate recognition of their specific operating patterns **Fig 8** shows the individual appliance disaggregation among total power consumed**.** In this training, a labeled set had to be made to indicate when each appliance in a dataset was in use.

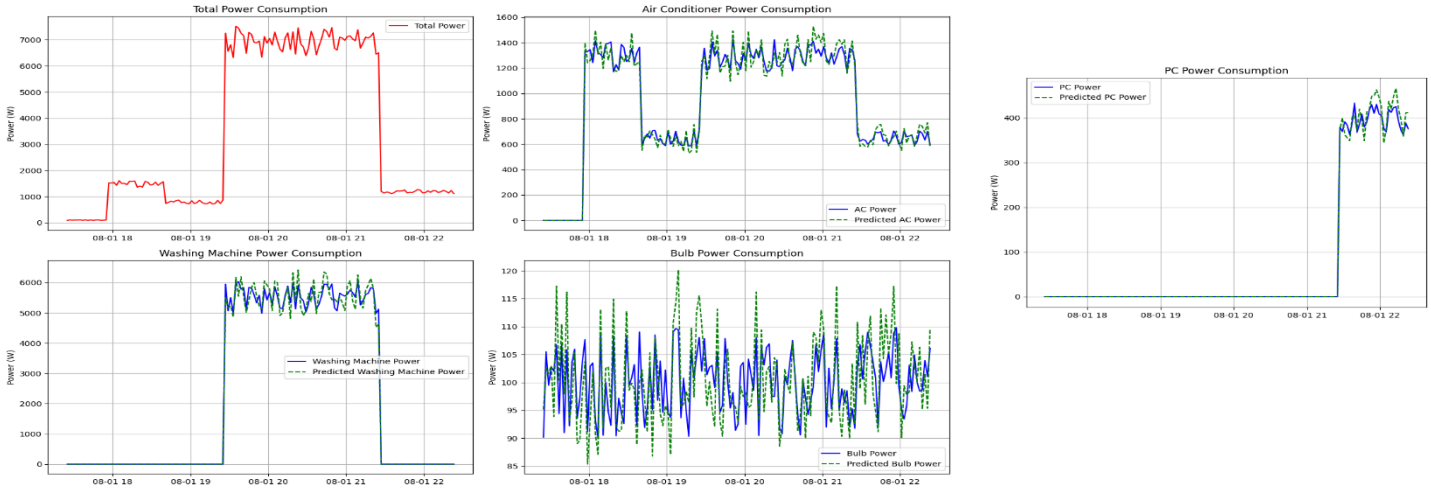


Fig 8: Individual Appliance contributionof power predicted vs true power

The trained DNN was tested under test scenarios where all appliances were running in sequence. This is necessary as the model would thus estimate the total energy usage in terms of aggregate usage of multiple appliances. Leverage on the knowledge sourced individually from training, DNN ensured a clear distinction of each appliance contribution even when running in tandem, showing the efficiency in load disaggregation. The application of DNNs is of extreme importance to provide a detailed insight into the usage of energy in residential settings.

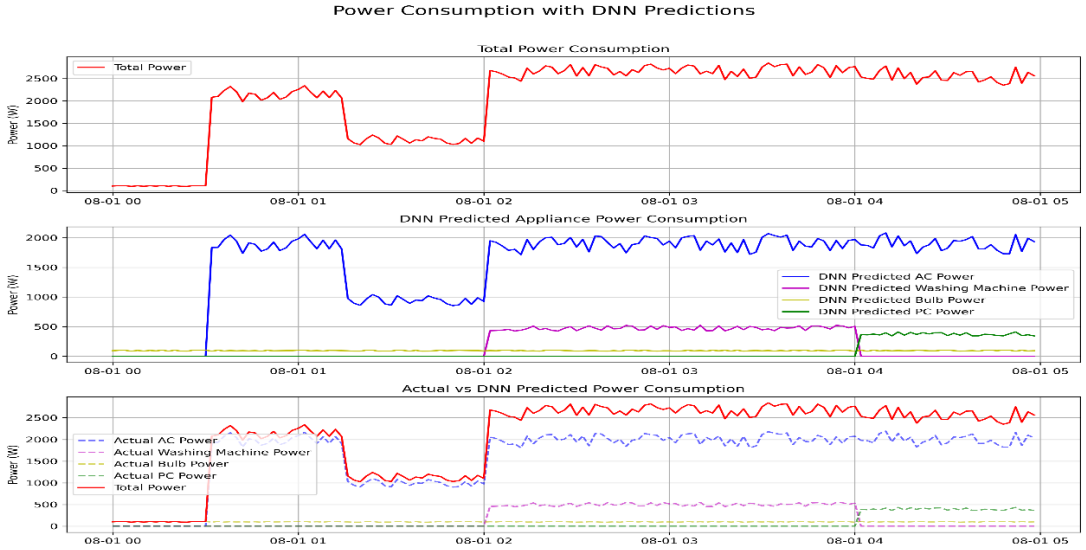


Fig 9: DNN predicted each appliance contribution from total power

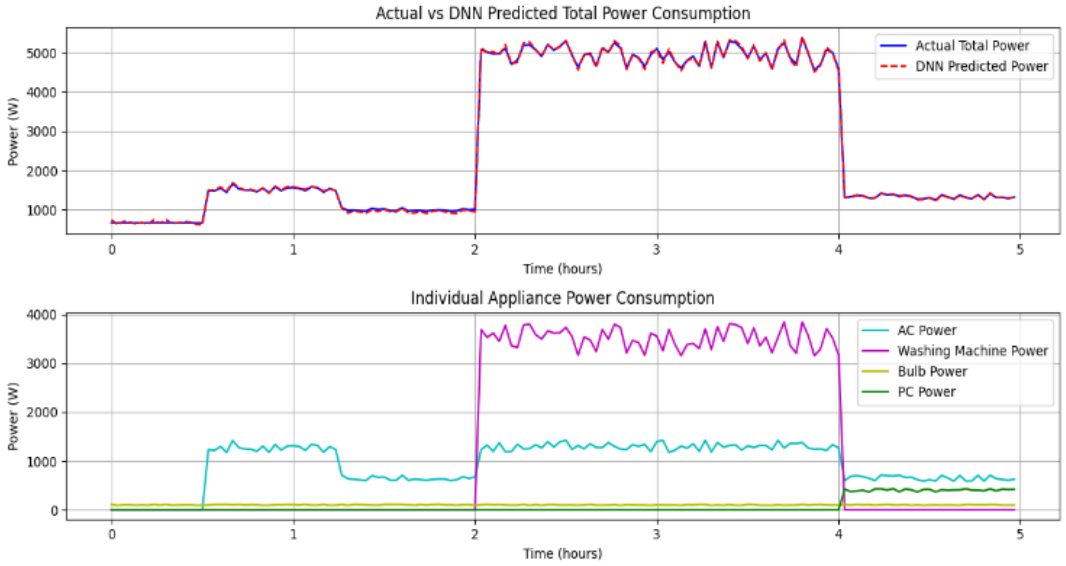


Fig 10: DNN predictions based on total power

The architecture of the DNN is a kind of arrangement in the form of a layer of interconnected neurons, functions similarly to that of biological neurons. It processes incoming inputs with the weighted sum and transmits it to an activation function. While features are extracted from energy consumption data, forming an input layer, more complex relationships embedded in data form hidden layers, with multiples neurons that can learn. The output layer finally generates predictions telling how likely each appliance will be active at any given time.

C. Dashboard for User Engagement

This Power BI dashboard gives real-time data visualization
through the monitoring of energy consumption. It fetches its data in real time from PostgreSQL, reflecting the current voltage and power consumption and the total energy usage shown in **Fig 12**. In addition, it reflects the current bill and also indicates the estimated charge for electricity for the rest of half an hour based on the consumption pattern as observed at this point in time. Based on this, users can predict their costs and hence avoid unnecessary expenses and ensure efficient management of electricity consumption.

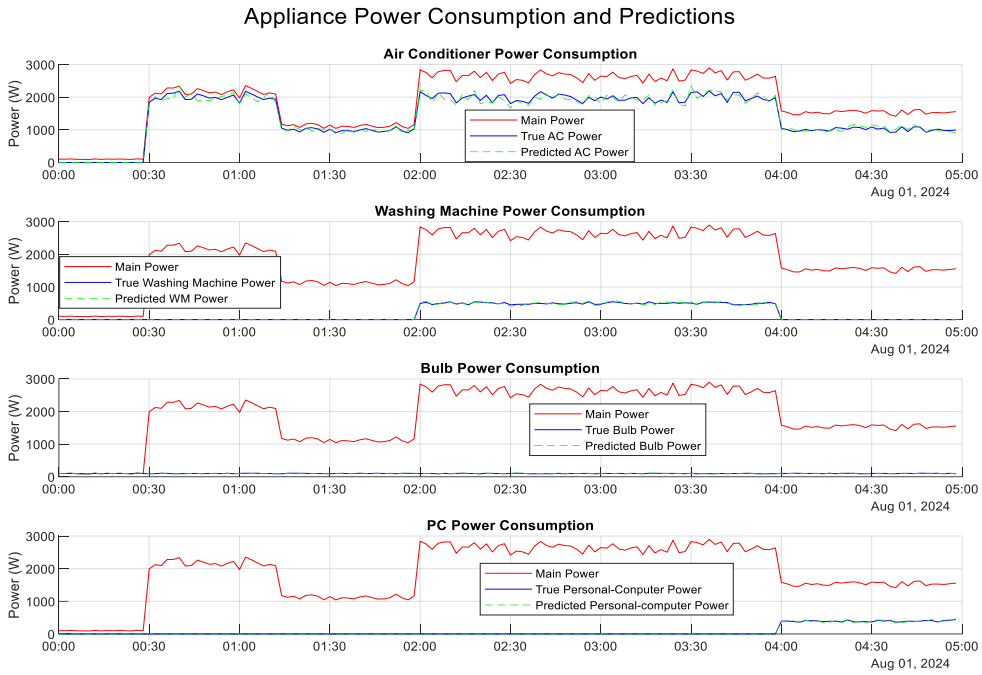


Fig 11: DNN predictions based on total power for each appliance

The solution shows these recommendations generated by the DNN model on the dashboard to the users depending on their type: essential and non-essential loads. Hence, the Dashboard prompts users to turn off those devices that they considered non-essential to save energy, thus minimize energy costs. More so, the dashboard provides a list of non-essential loads in operation to allow users to identify which non-essential devices are in operation at a given time shown in **Fig 13**. The combination of insights generated from machine learning along with database-driven metrics allows the consumer to be equipped with information to make intelligent use of energy.

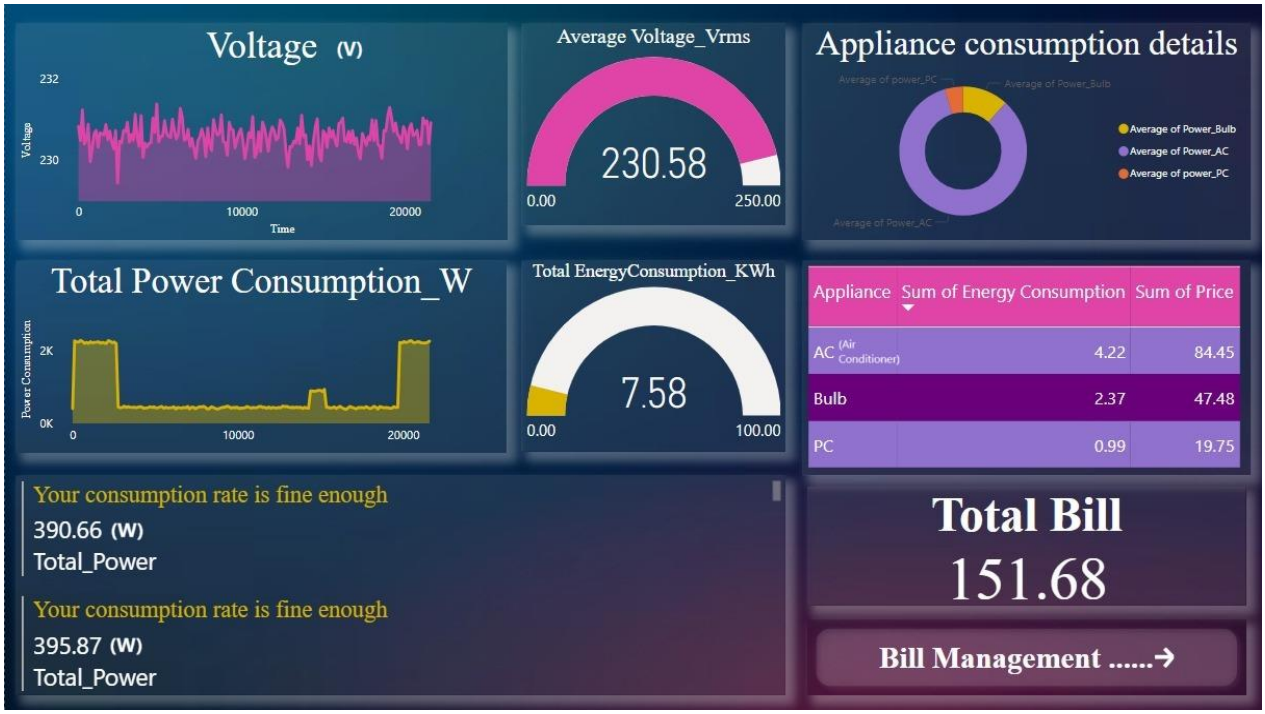


Fig 12: Real time power consumption visualization

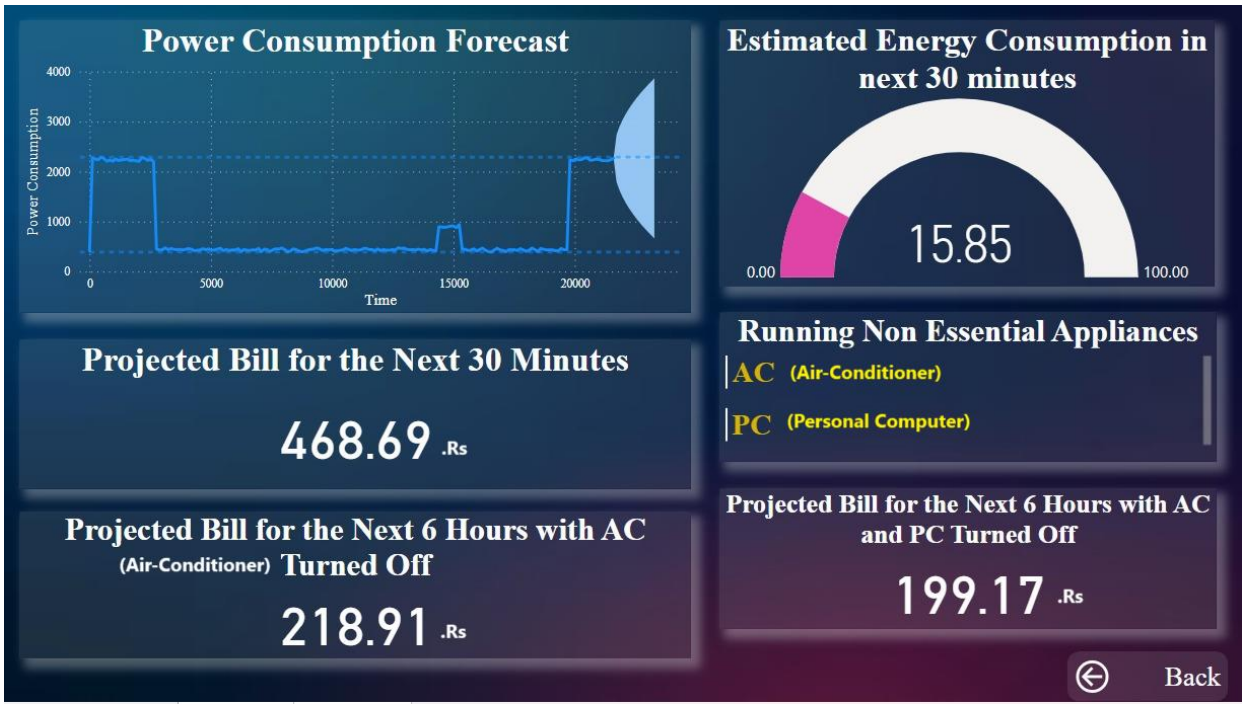


Fig 13: Forecasted Bill and list of non-essential loads

V. CONCLUSION

The framework proposed a DSM mechanism for residential loads by integrating smart energy meters, NILM and DNN for identification of shiftable loads in terms of appliances for better energy consumption. The work successfully displayed the monitoring and real-time classification of energy usage. This work also introduces the concept of categorization of essential and non-essential appliances using machine learning so that dynamic pricing strategies are made possible, thus enhancing the potential strategy making for energy efficiency.

The methodological approach demonstrates the strength of the DNN in potentially disaggregating load signatures from individual appliances, even if several appliances are running at the same time, and makes an accurate prediction about energy usage. A user-friendly dashboard is also demonstrated to provide real-time insights to the consumers about their energy consumption decisions. This can eventually help reducing electricity costs while maintaining comfort. The framework will encourage proactive practices in energy consumption and ensure stability for the power grid while facilitating the move toward sustainable energy practices.

REFERENCES

[1] Palensky, P., & Dietrich, D. (2011). *Demand-side management: Demand response, intelligent energy systems, and smart loads*. IEEE Transactions on Industrial Informatics, 7(3), 381-388.

[2] Gellings, C. W. (2020). *The Smart Grid: Enabling Energy Efficiency and Demand Response*. CRC Press.

[3] Hart, G. W. (1992). *Nonintrusive appliance load monitoring*. Proceedings of the IEEE, 80(12), 1870-1891.

[4] Zoha, A., Gluhak, A., Imran, M. A., & Rajasegarar, S. (2012). *Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey*. Sensors, 12(12), 16838-16866.

[5] Kelly, J., & Knottenbelt, W. (2015). *Neural NILM: Deep neural networks applied to energy disaggregation*. Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, 55-64.

[6] Alam, M. M., Sulaiman, S. A., & Rahman, M. A. (2017). *Machine learning techniques for energy consumption prediction: A review*. In *Advances in Energy Research* (pp. 65-78). Springer.

[7] Li, H., Yang, M., Wang, J., & Zhang, Y. (2018). *A deep learning approach for non-intrusive load monitoring: A review*. *Applied Energy*, 230, 818-827.

[8] Faruqui, A., Leslie, J., & Pate, A. (2010**).** *Dynamic pricing and its discontents: The impact of time-based pricing on residential energy consumption*. *Energy Economics*, 32(6), 1217-1233.

[9] Albadi, M. H., & El-Saadany, E. F. (2008). *Demand response in electricity markets: An overview*. *Energy*, 33(7), 1049-1060.

[10] Alahakoon, D., & Yu, Y. (2016**).** *Smart home energy management systems: A review of the state of the art and future challenges*. *Renewable and Sustainable Energy Reviews*, 60, 22-30.

[11] Wang, Y., Liu, Y., & Wang, J. (2017). "Impact of real-time feedback on household energy consumption: A study based on a dynamic pricing scheme." *Energy Reports*, 3, 21-28.

1st International Conference on Advancement in Engineering Sciences 2024

28- 29 OCTOBER 2024, Lahore , Pakistan

COMMITTEE MEMBERS

Committee Members

- Engr. Dr. Osama Majeed Butt (Conference Secretary)
- Engr. Dr. Kamal Shahid
- Mr. Zakaullah
- Mr. Usama Arif
- Mr. Sheikh Abubakar Kareem
- Mr. Shoaib Kaleem
- Ms. Nayab Shahid
- Ms. Jannat khurram
- Mr. Abdul Waheed
- Mr. Ali Tahir
- Ms. Mahnoor Sharafat