Journal of Nonlinear Analysis and Optimization: Theory & Applications

(UGC CARE Journals)
ISSN : 1906-9685
Vol-15 No-1 Special Issue-03 (April 2024)

Editors-in-Chief:

Sompong Dhompongsa
Somyot Plubtieng

Publisher
Naresuan University

Naipunnya
TO REACH THE UNREACHED

NAIPUNNYA

ACCREDITED WITH GRADE A NAAC

SCHOOL OF MANAGEMENT

(Affiliated to the University of Kerala)

NATIONAL RESEARCH CONFERENCE

"VIJNAN 2K24"

17 FEBRUARY 2024

ON

"COMPUTER SCIENCE AND INFORMATION SYSTEMS"

Organized By

DEPARTMENT OF COMPUTER SCIENCE AND APPLICATIONS

In Association With

NAIPUNNYA RESEARCH FORUM

NAIPUNNYA SCHOOL OF MANAGEMENT

(Affiliated to the University of Kerala)

Cherthala, Alappuzha, Kerala – 688 529

www.naipunnyacollege.ac.in

info@naipunnyacollege.ac.in

04782817476 || 04782817478 || +91 8606802255

CHIEF PATRON

Fr. Baiju George Ponthempilly

(Executive Director & Principal)

PATRONS

Fr. Chacko Kilukkan

(Assistant Executive Director)

Ms. Pushpa John

(Vice Principal)

Fr. Paul (Anil) Kiliyelikkudy

(Academic Co-Ordinator)

CONFERENCE COMMITTEE

Mr. Vinod Chandran

(Head of the Department, Department of Computer Science and Applications)

Ms. Amulia PM

Conference Coordinator and NRF Member

(Assistant Professor, Department of Computer Science and Applications)

Dr. Prasanth Kumar M

Director, Naipunnya Research Forum

(Head of the Department, PG Department of Commerce & Management)

JOURNAL OF NONLINEAR ANALYSIS AND OPTIMIZATION: THEORY & APPLICATIONS

Vol-15 No-1 Special Issue-03 (April 2024) UGC CARE Journals

ISSN-1906-9685

- **T. Suzuki,** Kyushu Institute of Technology, Japan
- **W. Takahashi,** Tokyo Institute of Technology, Japan
- **M. Thera,** Universite de Limoges, France
- **R. Wangkeeree,** Naresuan University, Thailand
- **H. K. Xu,** National Sun Yat-sen University, Taiwan
- **W. Anakkamatee,** Naresuan University, Thailand
- **P. Boriwan,** Khon Kaen University, Thailand
- **N. Nimana,** Khon Kaen University, Thailand
- **P. Promsinchai,** KMUTT, Thailand
- **K. Ungchittrakool,**Naresuan University, Thailand
- **N. Petrot**, Naresuan University, Thailand

JOURNAL OF NONLINEAR ANALYSIS AND OPTIMIZATION: THEORY & APPLICATIONS

Vol-15 No-1 Special Issue-03 (April 2024) UGC CARE Journals

ISSN-1906-9685

INDEX

AN EXTENSIVE STUDY OF WIRELESS SENSOR NETWORK (WSN) RULES

Mr. N. Karthik

Research Scholar, Department of Computer Science, Park's College, Tirupur.

Email: shreekarthy@gmail.com

Dr. V. Kathiresan

Principal, A.V.P. College of Arts and Science, Tirupur.

Email: vkathirmca@gmail.com

Abstract: *The Wireless Sensor Network (WSN) has been widely used in other technologies, such as the Internet of Things and artificial intelligence. Numerous relevant uses have been made of it. The WSN is becoming one of the network's more dependable and popular apps. WSN is thought to be the most significant technological advancement of the twenty-first century and is necessary for bridging the gap between the physical world and the logical information world. A Wireless Sensor Network (WSN) comprises many sensor nodes with varying power, storage, and processing capacities. By using clustering, these systems' performance can be improved. The clustering techniques used in Wireless Sensor Networks (WSN) mostly focused on power consumption, which could result in higher overhead. Cluster head re-election results from the cluster head sending data on a regular basis, which uses more energy. In cluster-based WSNs, there is a greater imbalance in the energy consumption of nodes because of the nonuniform distribution of nodes. Scheduling node activities for data transmission while preserving energy usage is the fundamental problem in WSN. The latest problem is detected and explored in this article, which also provides background information on WSN and its recent research endeavors.*

Keywords: Energy consumption, sensor node, power consumption, clustering, secure, and network lifetime.

INTRODUCTION

Cluster head re-election results from the cluster head sending data on a regular basis, which uses more energy. In cluster-based WSNs, there is a greater imbalance in the energy consumption of nodes because of the nonuniform distribution of nodes. Scheduling node activities for data transmission while preserving energy usage is the fundamental problem in WSN. The latest problem is detected and explored in this article, which also provides background information on WSN and its recent research endeavors.

A wireless sensor network (WSN) is made up of multiple nodes that are connected to each other. These nodes are first placed inside a network's coverage area and used for targets or event tracking. By carrying out quick calculations, these sensor nodes are used in the environmental domain extremely efficiently to perceive the data [2]. These gathered data are sent to the Base Station (BS), a sink node. Through the gateway nodes, the internet links the other nodes with the sink nodes.

Applications for wireless sensor networks (WSNs) can be categorized as either tracking or monitoring. When using monitoring applications, the sensor nodes keep an eye on the surroundings and periodically report any changes to the sink or when a predetermined event takes place. The event is generated by the sensor node itself or in response to the user's query when the measurement of the node surpasses a predefined threshold value. Programs for tracking data are updated in real-time with the measurements. WSN is extensively utilized in habitat, military surveillance, environmental, health, and structural monitoring. Vehicle tracking, person tracking, military tracking of adversaries, and animal tracking are among the most widely used tracking applications.

Nodes in the field with a distinct set of energy values remain in the same place. The system is divided into clusters, each of which consists of strong super nodes, sophisticated nodes, and regular nodes. A base station's location is fixed. In M-SEEC, MAN is the percentage of all nodes that are ready with α times more energy than normal nodes (NN), also known as advance nodes (AN). MSN is the subset of all nodes that have β times more energy than regular nodes (NN), also known as super nodes (SN). Research is now being focused on an effective technique that significantly increases the data transmission rate across wireless networks. By selecting the Cluster Head (CH), which is surrounded by a greater number of nodes next to the energy, the clustering is made more effective. It increases the system's scalability and lowers nodes' energy consumption [3].

WSNs are composed of low-cost, widely dispersed, multipurpose sensors that use little energy and communicate with one another to collect data for temporal and spatial measurements of several parameters, including sound and temperature. WSN was utilized in a wide range of applications, including home automation, machine health monitoring, manufacturing process control, congestion control, and surveillance of the surrounding area and environment. In a sensor system, sensors are dispersed randomly over a region without any infrastructure beforehand. Every sensor has the ability to scan the environment, collecting and sending information back to the sink. Because most WSN sensors are battery-operated and have limited capabilities, energy consumption becomes a major issue because the network must function for a predetermined amount of time.

Naturally, the two primary functions of sensors that use the most energy are identifying and compiling local data and transmitting data to the base station. Because the main factor affecting energy consumption during gathering is sampling speed, it is relatively constant. However, compared to sensing, the situation of power consumption during information uploading is far more complicated. The process of downloading data from sensors to wireless networks uses a significant amount of electricity, and energy consumption varies widely throughout sensors. Furthermore, the kind of network topology, the location of the data sink, and the swarm intelligence method all affect how much energy is utilized. Because these sensors must send more data packets than others, their batteries run out faster in the vicinity of sink nodes [4]. Energy consumption during data collection is therefore a major and challenging problem in WSNs since it mostly determines the network life cycle. Due to its great practical importance, a great deal of study has been done in the past few centuries on effective data collecting in WSNs as well as other creative ideas.

In order to create an efficient and scalable network, this research attempts to reduce power consumption and extend network life through energy consumption and routing. The primary goal is to resolve the optimization conundrum by creating the optimal cluster configuration and data transfer. Lastly, the goal of this research is to find a hybrid routing method that can assist in creating a more effective and superior network for WSNs[5].

Challenges in the Design of Wireless Sensor Networks

The following issues need to be taken into account when creating and implementing WSN algorithms. In order to achieve efficiency in WSN operations, these are the main prerequisites[6].

- Fault tolerance: In a wireless sensor network (WSN), node failure might happen at any time due to physical damage, interference, or low battery life. Fault tolerance is the capacity of a network to function even in the event that a particular sensor node fails. The architecture of the sensor network should be such that the loss of a few sensor nodes does not affect the operation of the network as a whole.
- Topology: When sensor nodes malfunction, the network topology is altered.

Topological changes in mobile WSNs are also caused by node migration.

- Scalability: Depending on the application, WSNs can have a large number of different sensor nodes. It is possible for certain applications to have more than one million sensor nodes. Under these conditions, scalability becomes an important consideration while designing the system. Scalability is another word for a sensor network's capacity to maintain performance standards as the number of sensor nodes grows.
- Hardware limitations: The main parts of a sensor node should be able to fit into a module the size of a matchbox. Because of their tiny size, the sensor nodes have restricted energy supply and storage capacity. Furthermore, the processing and transmission capabilities of the sensor node are constrained. Unlike nodes in wired networks, sensor nodes in wireless sensor networks are subject to bandwidth limitations.
- Power consumption: To detect, process, and send data to the sink, the sensor nodes need electricity. It will be difficult to communicate during these three jobs.
- Transmission channel: A shaky wireless connection is used by the sensor nodes to communicate. For wireless communication channels, the radio, infrared, and optical frequency ranges are utilized.

Clustering inside Wireless Sensor Networks

In order to improve stability and lower the energy consumption of the system, clustering is necessary. During clustering, the network is organized into clusters. The cluster head, or CH for short, is the leader of each cluster [8]. The cluster head's main duties include gathering data from nodes, aggregating it, and connecting it to the sink or base station. Between base stations and sensor nodes, cluster heads act as a conduit. The quantity of energy used is greatly reduced through clustering. The direction of energy savings and clustering was influenced by other concepts such as mobile sinks and rendezvous nodes. Every other node in a cluster is a member, and the central node is referred to as the cluster head. Members of the cluster submit data to the CH, which then aggregates the data, eliminates duplicated data, and performs additional tasks. Once filtered, CH transmits it to the base station.

Objectives of Clustering:

- Extending the lifetime of the network: - Energy management is not as important as it is in mobile systems, because portable devices (like phones) may be recharged when their batteries run out. The battery life of WSN is limited. As a result, extending the lifespan of networks is becoming increasingly important and one of the primary considerations when creating WSNs. In order to mitigate this issue, clustering helps to extend the lifetime of WSNs and reduce energy consumption.

- Tolerating errors: - Numerous approaches are taken to address the issue of node failure. Proxy cluster heads handle the issue when a node fails or has less transmission energy; in other cases, CH rotation can be the solution. Fault tolerance is one of the main objectives while developing a clustering technique.

LITERATURE REVIEW

P. Sanjeevi and colleagues (2020) [9] Recent advances in the Internet of Things (IoT) have made farming and agriculture part of a precision sensor network. Built on wireless sensor networks (WSNs) and cloud computing, the Internet of Things (IoT) wide-area network has potential applications for the agricultural and distant farming industries. A scalable wireless sensor network architecture was presented by the study's authors for internet-of-things-based remote farming and agricultural monitoring and control. Effective water resource management and efficient water use (PAF) are critical components of precision agriculture and farming. Materials. Combining WSN and IoT technologies can result in effective water irrigation control. IoT is used in agriculture to improve farmer productivity and enable the efficient connection of several wireless sensors. In terms of throughput maximization, delay reduction, high signal-to-noise ratio (SNR), low mean square error, and enlarged coverage area, writers have examined the WSN architecture. The studies' findings demonstrate that the recommended methodology outperforms more traditional IoT-based farming and agriculture. Precision agriculture based on IoT and the farming system both confirm their high worth. for farmers, as agriculture benefits from both much and little irrigation. The parameter values of sensor conditions, such as temperature and wetness, may be fixed based on the condition of the agricultural field. In addition to addressing the problem of insufficient irrigation, the proposed system will make efficient use of available resources. Graphics retrieval and display capabilities from earlier technologies serve as a proxy for WSN performance.

Vandana Bhasin et al. (2020) [10] Sensor networks are an efficient way to combine sensing, processing, and communication. These networks are constructed from small, inexpensive sensor nodes with constrained radio ranges and processing capabilities. They focus on different applications at the same time that they encounter severe energy constraints and low memory resources. The complex personalities of sensor networks also have a direct bearing on these nodes' hardware architecture. Several hardware platforms have been created to test research-community theories and implement applications in all fields of science and technology, such as Crossbow, Intel, and Inmate. Security concerns are crucial because of how many apps are developed in this industry. The two layers of the network stack—the link layer and the network layer—are covered in this article's overview of security architectures.

Subramanian, Balaji, and others (2019) [11] Low-power miniature sensors used in scientific research are supported by a wireless sensor network, a wide-area monitoring tool. Requirements for WSNs include memory, computational power, bandwidth, and energy, which are all limited. The highest degrees of energy economy are provided by the Cluster Routing protocol used in wireless sensor networks. When selecting a cluster head, Cluster Routing Protocols are used to establish the cluster (CH). Finally, data packets are being transmitted between CHs and subsequently routed to the base station. CH is selected in the setup phase. This system examined a multichip transmission in which each hop delivers a data packet. Subsequently, the finalized data packets are transmitted to the base station. The packets are transferred by the cluster head from the source sensor to the base station of the wireless sensor network. Three factors are employed for fuzzy logic type 1, including the trust and distance factors. The nodes with high trust that are closest to the base station are predicted by fuzzy logic. The optimal forwarder will be chosen by Type 1 fuzzy logic to be CH. In addition to immediately extending the network lifetime, it will reduce network overhead.

Nitin Goyal and colleagues (2019) [12] Despite the fact that the oceans and rivers are still unexplored areas, specialists are particularly interested in underwater monitoring because misfortunes and disasters often occur there. Underwater Wireless Sensor Networks (UWSN) are intended to be used in aquatic environments for a range of purposes, such as gathering data on the ocean, controlling or averting calamities, aiding in navigation, repelling assaults, and keeping an eye on pollution. Unmanned Wireless Sensor Networks (UWSN) are made up of sensor nodes that collect data and send it to sinks, just like terrestrial Wireless Sensor Networks (WSN). water-based media Among these challenges include mobile sensor nodes, long propagation delays, limited network capacity, and multiple message receptions. An extensive overview of the difficulties associated with underwater sensor networks is provided in this book. The authors include descriptions of the available test beds, routing protocols, experimental projects, simulation platforms, tools, and analysis. Approximately 96% of the water on Earth is found in oceans, seas, and other bodies of water.

For academics, the creation of networks or protocols to analyze data from a wide range of underwater environments has never been easy. Nonetheless, UWSN has been designed with special capabilities to collect, process, and store the enormous volumes of data that are found underwater. The battle to function like terrestrials in the difficult aquatic medium has made UWSN an intriguing field of study. Unlike the electromagnetic or radio waves—which have limitations—used in land-based WSNs, the UWSN operates on distinct principles. There are a number of unanswered questions in UWSN that require further research, such as developing energy-efficient routing techniques, limited battery depletion, and available bandwidth. For data to be transported effectively and reliably in a variety of applications—including disaster relief, mine detection, pollution monitoring, aided navigation, offshore exploration, and tactical surveillance—these challenges must be addressed. To demonstrate how the network operates, the authors of this paper have supplied the UWSN's communication architecture. The difficulties of fault tolerance, energy efficiency, and quality of service are also covered. Table 1 provides a detailed examination of WSN.

Table 1. Analysis of WSN Techniques

Routing Protocol	Classification	Energy Efficiency	Location Awareness	Data Aggregation	Scalability	Quality of service	Fault Tolerance	Load Balance	Disadvantages
HCEH-UC [13]	Classical	Moderate	No	No	Yes	Yes	No	No	The performance of the approach is moderate, which necessitates further enhancement
Fuzzy-based Energy Efficient Routing Protocol (E-	Classical	Moderate	No	No	Yes	Yes	No	No	The clustering process met the convergence at the early stage,

FEERP)[14]									where an efficient technique is needed for CH selection.
BSR [15]	Classical	Moderate	Yes	No	Yes	Yes	No	No	The network is prone to attack.
RRS [15]	Classical	Moderate	Yes	No	Yes	Yes	No	No	The topology needs improvement
ANS [15]	Classical	Moderate	Yes	No	Yes	Yes	No	No	The proposed work needs to be enhanced in handling the network overhead.
LEACH-SWDN [16]	Classical	Good	No	Yes	Limited	No	No	Yes	Network overload is not considered due to the data transmission overhead
ASLPR [17]	Classical	Good	No	Yes	Moderate	No	Yes	Yes	When the network is

									large, and the data transmission necessitates multi-hop routing
Q-LEACH [18]	Classical	Good	Yes	No	Limited	No	No	Yes	The protocol is not
									effective in handling mobility and large-scale network.
ERP [19]	Classical	Good	No	Yes	Limited	No	No	Yes	Fitness is not effective, and the protocol is not stable.
C-RPL[20]	Classical	Good	Yes	Yes	Moderate	Yes	No	Yes	This is not effective for distributed context.
OZEEP [21]	Classical	Very Good	Yes	Yes	Very Good	Yes	Yes	Yes	Re-clustering and clustering are not effective.

MTPCR [22]	Classical	Very Good	No	No	Very Good	No	Yes	Yes	Awareness of quality of service is absent
EAODV [23]	Classical	Good	No	Yes	Limited	Yes	No	Yes	The protocol is not strong enough to handle large-scale data transmission.
PHASeR [24]	Classical	Good	Yes	Yes	Very Good	No	Yes	Yes	The channel fading influences the performance of the proposed approach
HEEDML [25]	Classical	Very Good	Yes	Yes	Limited	No	No	Yes	Not effective in handling dynamic network

Energy Efficiency as a Crucial Concern

Given the limited battery life of sensor nodes, energy is considered one of the prime problems in WSN. Although earlier methods for solving this issue have been explored, they have not shown to be successful. Possibly the best approach to handle this issue is clustering. The network has been divided into clusters. The Cluster Head (CH) is the head of each cluster. Data collection, aggregation, and transmission from nodes to the BS are the responsibilities of the CH. The quantity of energy used is greatly reduced through clustering.

In the WSN, a sensor uses more energy when it detects, processes, transmits, or receives data in order to meet the application's requirements. Data acquisition is the only focus of the detecting subsystem. Very few sensors will use less power if there is less data generated. WSNs' inherent redundancy will provide identical reports that the network is in charge of sending to the sink. Results from experiments confirm that the communication subsystem is a major source of power loss [30]. Regarding the transmission, a significant quantity of power is also lost in states that are pointless from the perspective of the application, like:

- Collision: When multiple packets are received by a node simultaneously, a packet collision takes place. Each packet that causes a collision needs to be discarded and retransmitted.
- Overhearing: When a source transmits a packet, it is received by all nodes in its communication range, even if they are not the intended recipient. Thus, when a node accepts packets meant for other nodes, energy is wasted.
- Control packet overhead: In order to initiate data communication, a minimum quantity of control packets must be used.
- Interference: A packet that cannot be decoded is sent to each node that is positioned between the communication interval and the interference. Numerous ways have been developed to reduce power consumption and improve network lifespan, as network longevity has emerged as a critical aspect for evaluating WSNs. Energy efficiency strategies may be broadly classified into two categories: topology management and efficient energy path.
- Topology control: By balancing transmitting energy and preserving network connectivity, topology management reduces power consumption. Based on local knowledge, a novel reduced topology is created.
- Energy-efficient routing: Routing protocols must be designed to minimize power consumption and optimize system duration by avoiding nodes with little energy

remaining and end-to-end transmission. A number of protocols make use of wireless communication transmission characteristics or node mobility to reduce the amount of power used at the sink node. On the other hand, some create a path to the goal using the geographical coordinates of the nodes. Concurrently, certain other protocols establish a hierarchical structure among nodes in order to reduce overhead and streamline routing

CONCLUSION

An analysis of the energy-efficient protocols of Wireless Sensor Networks (WSNs) has revealed how important these protocols are to the resource efficiency, node lifetime, and sustainability of the network. By analyzing the procedures, advantages, and disadvantages of these energy-efficient protocols, the study provided insight into the potential contributions of these protocols to the field of WSNs. The investigation uncovered a number of tactics employed by energy-efficient protocols, including data aggregation, low-power listening, sleep scheduling, and enhanced routing algorithms. These tactics work together to cut down on energy use during talking and idle periods. Many scalable, flexible, and energy-efficient protocols can be used to accommodate dynamic network conditions and evolving application requirements. Because of their adaptability, WSNs can function successfully in a range of scenarios without compromising energy efficiency. The study found a number of problems and uncharted territory for the development of WSN protocols in energy efficiency. These include finding synergies between data quality and energy efficiency, cutting down on control message overhead, and fixing security vulnerabilities. They also include heterogeneous network protocol optimization. The extensive study emphasizes how vital energy-efficient protocols will be in shaping the future of wireless sensor networks. The findings demonstrate how innovative techniques and energy-efficient design principles combine to provide WSN installations that are robust and long-lasting.

REFERENCE

[1] Behera, T. M., Samal, U. C., Mohapatra, S. K., Khan, M. S., Appasani, B., Bizon, N., & Thounthong, P. (2022). Energy-efficient routing protocols for wireless sensor networks: Architectures, strategies, and performance. *Electronics*, *11*(15), 2282.

[2] Dhabliya, D., Soundararajan, R., Selvarasu, P., Balasubramaniam, M. S., Rajawat, A. S., Goyal, S. B., ...& Suciu, G. (2022). Energy-efficient network protocols and resilient data transmission schemes for wireless sensor Networks—An experimental survey. *Energies*, *15*(23), 8883.

[3] Gupta, S. K., & Singh, S. (2022). Survey on energy efficient dynamic sink optimum

routing for wireless sensor network and communication technologies. *International Journal of Communication Systems*, *35*(11), e5194.

[4] Khan, Z. U., Gang, Q., Muhammad, A., Muzzammil, M., Khan, S. U., Affendi, M. E., ... & Khan, J. (2022). A comprehensive survey of energy-efficient MAC and routing protocols for underwater wirelesssensor networks. *Electronics*, *11*(19), 3015.

[5] Alomari, M. F., Mahmoud, M. A., & Ramli, R. (2022). A Systematic Review on the Energy Efficiency of Dynamic Clustering in a Heterogeneous Environment of Wireless Sensor Networks (WSNs). *Electronics*, *11*(18), 2837.

[6] Ullah, F., Khan, M. Z., Mehmood, G., Qureshi, M. S., & Fayaz, M. (2022). Energy efficiency and reliability considerations in wireless body area networks: a survey. *Computational and Mathematical Methods in Medicine*, *2022*.

[7] Jonnalagadda, S., Shyamala, K., & Roja, G. (2022). Energy-efficient routing in WSN: a review. *ECS Transactions*, *107*(1), 1111.

[8] Sahu, S., &Silakari, S. (2022). Energy efficiency and fault tolerance in wireless sensor networks: Analysis and review. *Soft Computing: Theories and Applications: Proceedings of SoCTA 2021*, 389-402.

[9] Sanjeevi, P., S. Prasanna, B. Siva Kumar, G. Gunasekaran, I. Alagiri, and R. Vijay Anand. "Precision agriculture and farming using Internet of Things based on wireless sensor network." Transactions on Emerging Telecommunications Technologies 31, no. 12 (2020): e3978.

[10] Bhasin, Vandana, Sushil Kumar, P. C. Saxena, and C. P. Katti. "Security architectures in wireless sensor network." International Journal of Information Technology 12, no. 1 (2020): 261-272

APPLICATIONS OF NEUROIMAGING USING MACHINE LEARNING AND DEEP LEARNING TECHNIQUES

Dr. Ambily Merlin Kuruvilla
Associate Professor
PG Department of Computer Applications and Artificial Intelligence.
Saintgits College of Applied Sciences
Ambilykuruvilla@gmail.com
Phone Mo: 9447661714

Mr. Kevin Justees
Third Year BCA Student
PG Department of Computer Applications and Artificial Intelligence
Saintgits College of Applied Sciences
kevinjustees.11@gmail.com
Phone No: 8136945584

Mr.Jerrin John
Third Year BCA Student
PG Department of Computer Applications
and Artificial Intelligence
Saintgits College of Applied Sciences
Jerrinjohn2002@gmail.com
Phone No: 8129395175

Mr. Motti Zachariah Varghese
Third Year BCA Student
PG Department of Computer Applications and Artificial Intelligence
Saintgits College of Applied Sciences
Mottizacharia122@gmail.com
Phone No: 8078776180

Ms. Nandana P
Third Year BCA Student
PG Department of Computer Applications and Artificial Intelligence
Saintgits College of Applied Sciences.
nandanavinodh@gmail.com
Phone No: 9778529140

Abstract

Neuroimaging allows us to study the brain in both health and disease, neuroimaging is essential to clinical care and research. The morphological structure, physiological architecture, and corresponding imaging features of the brain are intricately linked.

A person's brain changes in shape, function, and relationships between different parts as they develop, grow, get sick, and recover. Neuroimaging is one of the few fields that fully utilizes modern analysis tools to investigate the brain and its function from imaging data. ML has begun to play a role in the following areas recently tracking changes in imaging over time, quickly identifying acute disorders like stroke and anatomical measurements, detection, segmentation, and quantification of lesions and disease patterns. As our capacity to visualize and study the brain.

Keywords: ML, DL, SVM, CNN, MRI, fMRI, ANN, PET, EEG: Machine Learning, Deep Learning, Support Vector Machines, Convolutional Neural Networks, Magnetic Resonance Imaging, functional Magnetic Resonance Imaging, Artificial Neural Network, Positron Emission Tomography, Electroencephalography.

1. INTRODUCTION

Neuroimaging, the process of visualizing the structure and function of the brain, has greatly benefited from the integration of ML and DL techniques. These methodologies offer powerful tools to analyse complex patterns within neuroimaging data, facilitating both diagnosis and understanding of neurological disorders. ML algorithms, such as SVM and Random Forest, have been employed for tasks like classification [12] of brain images based on specific features or biomarkers. These methods excel at identifying subtle differences in brain structure or activity that might elude human observers, thus aiding in the early detection and differentiation of conditions like Alzheimer's disease or schizophrenia.

DL, a subset of ML that utilizes ANN with multiple layers, has revolutionized neuroimaging analysis. CNN, in particular, have shown remarkable success in tasks like image segmentation, where regions of interest within brain scans are delineated automatically. This capability is invaluable for neuroscientists and clinicians seeking to pinpoint areas of abnormality or track changes over time in conditions such as brain tumours or multiple sclerosis. Moreover, DL techniques enable the extraction of hierarchical representations from raw neuroimaging data, uncovering intricate patterns that may hold crucial insights into brain function and dysfunction.

The integration of ML and DL into neuroimaging workflows also extends to functional imaging modalities like fMRI and electroencephalography (EEG). These techniques capture brain activity in real-time, generating vast amounts of data that require sophisticated analysis. ML algorithms can decode complex neural signals, enabling tasks such as predicting cognitive states or characterizing brain connectivity networks. DL architectures[13], such as Recurrent Neural Networks (RNN) and Graph Neural Networks (GNN), are well-suited for modelling temporal dynamics and capturing complex interactions[14] between brain regions, advancing our understanding of brain networks and their role in cognition and disease.

Overall, the synergy between neuroimaging and ML/DL techniques holds tremendous promise for advancing our understanding of the brain and improving clinical outcomes for neurological and psychiatric disorders. By harnessing the power of artificial intelligence, researchers and clinicians can unlock new insights into brain structure, function, and pathology, ultimately leading to more accurate diagnoses, personalized treatments, and enhanced patient care.

2. LITERATURE REVIEW

- Karl-Heinz Nenning et al.[1] Because it allows us to study the brain in both health and disease, neuroimaging is essential to clinical care and research. The morphological structure, physiological architecture, and corresponding imaging features of the brain are intricately linked. A person's brain changes in shape, function, and relationships between different parts as they grow, age, get sick, and recover. Neuroimaging is one of the few fields that fully utilizes modern analysis tools to investigate the brain and its function from imaging data. ML has begun to play a role in the following areas recently: (a) tracking changes in imaging over time; (b) quickly identifying acute disorders like stroke; and (c) anatomical measurements, detection, segmentation, and quantification of lesions and disease patterns. Our knowledge of the complex relationships inside the brain and how they affect therapeutic decision-making is expanding along with our capacity to visualize and analyse the brain
- Mahsa Dadar et al.[2] In this research, we present a comprehensive overview of ML methods used to extract clinical classifiers from structural MRI data. To help researchers better apply these methodologies in future works, we carefully address practical issues that are frequently seen in the literature. Furthermore, in order to present a thorough picture of the state of the art in several domains, we examine how these algorithms are used to treat a variety of illnesses and conditions (such as

Alzheimer's disease (AD), Parkinson's disease (PD), autism, multiple sclerosis, traumatic brain injury, etc.).

- Daniel Ranti et al.[3] Understanding the vast amount of intricate electronic data that hospital systems have accumulated over the years could transform modern medicine, but it also poses many difficulties. DL is particularly well-suited to tackle these problems, and the field of medical ML is set for revolutionary expansion because to recent developments in methods and hardware. Because neurologic diseases often manifest with mild symptoms, the clinical neurosciences stand to gain most from these developments. Here, we review the various domains where deep learning algorithms have already sparked change: medical image segmentation for the quantitative assessment of neuroanatomy and vasculature; connectome mapping for the diagnosis of Alzheimer's, autism spectrum disorder, and attention deficit hyperactivity disorder; mining of microscopic EEG signals and granular genetic signatures; and medical image analysis for the improved diagnosis of Alzheimer's disease and the early detection of acute neurologic events. We also address the obstacles to addressing the current problems and highlight significant obstacles in the incorporation of DL techniques in the therapeutic setting.
- Ritu Gautam et al.[4] This study conducts a comprehensive analysis of DL techniques for prognosticating eight neuropsychiatric and neurological disorders—stroke, Alzheimer's, Parkinson's, epilepsy, autism, migraine, cerebral palsy, and multiple sclerosis. These disorders pose significant health risks and can lead to further complications. Utilizing insights from 136 relevant publications, the research explores the methodologies and frameworks employed by various DL algorithms in diagnosing these conditions. It examines morbidity and mortality rates, analyzes the performance and publication trends of deep learning methods, and scrutinizes key performance metrics. The study underscores the need for further investigation into DL models for diagnosing stroke, cerebral palsy, and migraine, while also identifying opportunities to leverage Deep Boltzmann Machine, Restricted Boltzmann Machine, and Deep Belief Network for diagnosing neurological and neuropsychiatric disorders.
- Bin Jiang et al.[5] Numerous DL based clinical applications related to radiology have been proposed and investigated. These applications include risk assessment, segmentation tasks, diagnosis, prognosis, and even therapeutic response prediction. Other cutting-edge uses of AI in medical imaging include the removal of image

artifacts, normalization and harmonisation of images, enhancement of image quality, reduction of radiation and contrast dose, and abbreviation of imaging study duration. These applications are specifically focused on the technical aspects of medical imaging and are particularly relevant to image acquisition. This paper will discuss this subject and aim to give a general overview of DL methods used in neuroimaging.

- Peter A. Bandettini et al.[6] Neuroscience research has changed as a result of the quick development of neuroimaging techniques and their expanding accessibility. The quality of the data we are able to gather regarding the locations, dynamics, fluctuations, magnitudes, and types of brain activity and structural changes will determine the answers to many of the questions we have about how the brain is structured. An attempt is made to capture the state-of-the-art in a small portion of the quickly developing field of neuroimaging in this review. A brief background is given for each topic discussed, along with an overview of some of the most recent advancements and problems. The next section describes a few notable articles that were published within the last year or two, giving an idea of the paths that each field is heading in. Among the topics discussed are PET, EEG, magnetoencephalography (MEG), diffusion tensor imaging (DTI), voxel-based morphometry (VBM), optical imaging, and fMRI.
- Soonmee Cha et al.[7] The field of neuroimaging brain tumors has changed over time, moving from being primarily focused on morphology to also including function, physiology, and anatomy. In addition to summarizing the most recent developments in physiology-based imaging techniques that support established brain tumor imaging procedures, this review describes the current imaging standard for patients with brain tumors. A summary of various modern imaging techniques, such as diffusion-weighted MRI, perfusion MRI, and proton magnetic resonance spectroscopic (MRS) imaging, is included along with an emphasis on the advantages and disadvantages of the current imaging standards. Each imaging technique's fundamental physical concepts are briefly explained, followed by a more thorough examination of its therapeutic applications and its drawbacks.
- Manan Binth Taj Noor et al.[8]Over the past few decades, neuroimaging—more specifically, MRI—has been more crucial in our understanding of brain functioning and illnesses. Innovative ML approaches and high-performance computing tools have made it possible to identify neurological illnesses with previously unheard-of

precision thanks to these state-of-the-art MRI scans. However, it is exceedingly challenging to reliably identify such illnesses from the acquired neuroimaging data due to similarities in disease characteristics. In order to identify neurological disorders—with a particular focus on Parkinson's disease, schizophrenia, and Alzheimer's disease—this article critically evaluates and contrasts the capabilities of the current DL - based techniques using MRI data obtained using various modalities, such as functional and structural MRI

- Grega Repovš et al.[9] The assumption of linearity in neural processes by classic statistical methods limits their applicability to the analysis of neuroimaging data, as this work explores. It presents DL as a viable strategy to get beyond these restrictions, outlining its fundamental ideas and typical uses in neuroimaging research, such as data collection, segmentation, internal representation interpretation, and outcome prediction. The study discusses and suggests possible solutions for deep learning's problems with multidimensionality, multimodality, overfitting, and computing expense. The current application of DL in neuroimaging analysis is assessed, with a focus on the potential benefits of multimodality, raw data processing, and sophisticated visualization techniques. Research gaps and avenues for future investigation are noted, including the use of RDoC, transfer learning, and synthetic data generation frameworks.
- Li Zhang et al.[10] Recently, DL has been applied to the analysis of neuroimages, including PET, structural MRI, and fMRI. Compared to traditional ML, DL has significantly improved performance in computer-aided diagnosis of brain disorders. The uses of DL techniques for the analysis of brain disorders based on neuroimaging are reviewed in this work. By introducing several kinds of deep neural networks and recent advancements, we first give a thorough overview of DL techniques and widely used network structures. Next, we examine DL techniques for computer-assisted examination of four common brain disorders: schizophrenia, Parkinson's disease, Alzheimer's disease, and autism spectrum disorder. The latter two disorders are psychiatric and neurodevelopmental, respectively, and the first two are neurodegenerative.
- Jyoti Islam et al.[11] Alzheimer's disease is a degenerative neurological brain ailment that is incurable. Prompt detection of Alzheimer's disease can aid in appropriate management and avert harm to brain tissue. Researchers have used a number of

statistical and ML models to diagnose Alzheimer's disease. The exacting nature of Alzheimer's Disease detection stems from the similarities between standard healthy older people's MRI data and MRI data associated with Alzheimer's Disease. Modern DL methods have recently shown themselves to be as effective as humans in a variety of domains, including the processing of medical images. We suggest utilizing brain MRI data analysis to diagnose Alzheimer's disease using a deep CNN. We have performed extensive tests to show that, on the Open Access Series of Imaging Studies (OASIS) dataset, our suggested model performs better than comparing baselines.

3. RESULTS AND DISCUSSIONS

In this paper on neuroimaging using ML and DL techniques, several key findings and discussions emerge. Firstly, the integration of ML and DL methodologies into neuroimaging has significantly advanced the field, enabling more accurate diagnosis, prognosis, and understanding of neurological disorders. ML algorithms such as support vector machines and random forests have demonstrated efficacy in classifying brain images and identifying biomarkers for conditions like Alzheimer's disease and epilepsy. DL techniques, particularly CNN, have shown remarkable success in tasks like image segmentation and feature extraction, providing insights into brain structure and function. The results highlight the importance of these techniques in handling complex multivariate patterns in neuroimaging data. Additionally, discussions focus on the challenges faced, including data preprocessing, model interpretability, and generalization to new datasets.

Despite these challenges, the survey underscores the immense potential of ML and DL in advancing neuroimaging research and clinical applications. Further research directions are proposed, emphasizing the need for standardization, robust validation methods, and interdisciplinary collaborations to maximize the impact of ML and DL in neuroimaging. Overall, the survey provides a comprehensive overview of the current state, challenges, and future prospects of neuroimaging using ML and DL techniques, offering valuable insights for researchers and practitioners in the field.

4. CONCLUSION

In summary, the integration of ML and DL techniques into neuroimaging has revolutionized our understanding and diagnosis of neurological disorders. ML algorithms, such as SVM and random forests, excel in classifying brain images for early detection and differentiation of conditions like Alzheimer's disease and epilepsy, while DL methods, notably CNN, offer unparalleled accuracy in tasks like image segmentation, enabling precise delineation of brain regions. Despite challenges like data multidimensionality and overfitting, innovative solutions like transfer learning and synthetic data generation hold promise. Standardized methodologies for architecture and hyperparameter selection are crucial moving forward. With these advancements, neuroimaging promises to unveil new insights into brain function and pathology, leading to enhanced diagnostics, personalized treatments, and improved patient outcomes.

5. REFERENCES

[1] Nenning, KH., Langs, G. "Machine learning in neuroimaging: from research to clinical practice". *Radiologie* **62** (Suppl 1), 1–10 (2022), doi: 10.1007/s00117-022-01051-1

[2] José María Mateos-Pérez, Mahsa Dadar, María Lacalle-Aurioles, Yasser Iturria-Medina, Yashar Zeighami, Alan C. Evans, "Structural neuroimaging as clinical predictor: A review of machine learning applications", NeuroImage: Clinical, Volume 20, 2018, Pages 506-522, ISSN 2213-1582, doi: 10.1016/j.nicl.2018.08.019.

[3] Valliani, A.AA., Ranti, D. & Oermann, E.K. "Deep Learning and Neurology: A Systematic Review". *Neurol Ther* **8**, 351–365 (2019), doi: 10.1007/s40120-019-00153-8

[4] Gautam, R., Sharma, M. "Prevalence and Diagnosis of Neurological Disorders Using Different Deep Learning Techniques: A Meta-Analysis". *J Med Syst* **44**, 49 (2020), doi: 10.1007/s10916-019-1519-7

[5] Zhu G, Jiang B, Tong L, Xie Y, Zaharchuk G and Wintermark M(2019) "Applications of Deep Learning to Neuro-Imaging Techniques". Front. Neurol. 10:869. doi:10.3389/fneur.2019.00869

[6] Bandettini P. A. (2009). "What's new in neuroimaging methods?". Annals of the New York Academy of Sciences, *1156*, 260–293, doi: 10.1111/j.1749-6632.2009.04420.x

[7] Soonmee Cha (2009). "Neuroimaging in Neuro-Oncology". Neurotherapeutics: The Journal of the American Society for Experimental NeuroTherapeutics Vol. 6, 465–477, July 2009 The American Society for Experimental NeuroTherapeutics, Inc.

[8] Noor, M.B.T., Zenia, N.Z., Kaiser, M.S. et al. "Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia". Brain Inf. 7, 11 (2020), doi: 10.1186/s40708-020-00112-2

[9] Avberšek LK and Repovš G (2022) "Deep learning in neuroimaging data analysis: Applications, challenges, and solutions". Front. Neuroimaging 1:981642, doi:

10.3389/fnimg.2022.981642

[10] Zhang L, Wang M, Liu M and Zhang D (2020) "A Survey on Deep Learning for Neuroimaging-Based Brain Disorder Analysis". Front. Neurosci. 14:779, doi: 10.3389/fnins.2020.00779.

[11] J. Islam and Y. Zhang, "Early Diagnosis of Alzheimer's Disease: A Neuroimaging Study with Deep Learning Architectures", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 2018, pp. 1962-19622, doi: 10.1109/CVPRW.2018.00247.

[12] Manjima Sree, 2023, Estimation of Learners' Levels of Adaptability in Online Education Using Imbalanced Dataset, 2023 IEEE International Conference on Recent Advances in Systems Science and Engineering (RASSE) DOI: 10.1109/RASSE60029.

[13] Ambily Merlin Kuruvilla,2021, Improved Artificial Neural Network Through Metaheuristic Methods And Rough Set Theory For Modern Medical Diagnosis, Indian Journal of Computer Science and Engineering.

[14] NV Balaji,2021, Heart disease prediction system using Correlation Based Feature Selection with Multilayer Perceptron approach, IOP conference series: materials science and engineering

" A COMPARATIVE STUDY ON CLOUD COMPUTING IN HEALTH CARE INDUSTRY"

Ms.Rima K Ouseph

Assistant Professor

Department of Computer Science and Application

Naipunnya School of Management, Cherthala, Kerala

rima897@gmail.com

Mob-8921058236

Mr.Vivek S Varma

Assistant Professor

Department of Computer Science and Application

Naipunnya School of Management, Cherthala, Kerala

viveksvarmacpd@gmail.com

Mob-9746616621

Ms.Vidya S Mony

Assistant Professor

Department of Computer Science and Application

Naipunnya School of Management,Cherthala, Kerala

vidhyasmony@gmail.com

Mob-8089357187

Abstract

Cloud computing has emerged as a transformative technology with vast potential to revolutionize various industries, including healthcare. This study conducts a comparative analysis of cloud computing adoption and implementation within the healthcare sector, aiming to explore its benefits, challenges, and future prospects.

The research methodology involves a comprehensive review of existing literature, case studies, and empirical data to assess the impact of cloud computing on healthcare delivery, patient outcomes, data management, and cost-effectiveness. Through comparative analysis, the study examines different models of cloud deployment, such as public, private, and hybrid clouds, and evaluates their suitability for healthcare organizations of varying sizes and complexities.

Key factors influencing the adoption of cloud computing in healthcare, including security, privacy, regulatory compliance, interoperability, and data sovereignty, are critically evaluated. Additionally, the study investigates the role of emerging technologies, such as artificial intelligence and blockchain, in enhancing the capabilities and functionalities of cloud-based healthcare systems.

The comparative study delves into the experiences of healthcare organizations that have successfully implemented cloud solutions, highlighting best practices, lessons learned, and recommendations for future implementations. The analysis also considers geographical variations and regulatory frameworks shaping cloud adoption in different regions.

The findings of this comparative study contribute to a deeper understanding of the opportunities and challenges associated with cloud computing in the healthcare industry. By synthesizing insights from diverse perspectives, the research aims to inform strategic decision-making and policy formulation for healthcare stakeholders, IT professionals, policymakers, and researchers striving to leverage cloud technologies for improved healthcare delivery and patient outcomes.

Keywords— cloud computing, healthcare

I. Introduction

The healthcare industry is undergoing a significant transformation driven by technological advancements, with cloud computing emerging as a pivotal enabler of innovation and efficiency. [1] Cloud computing offers healthcare organizations scalable and flexible IT infrastructure, data storage, and computing resources on-demand, facilitating enhanced collaboration, data management, and service delivery. As a result, an increasing number of healthcare providers, payers, and other stakeholders are embracing cloud technologies to streamline operations, improve patient care, and drive cost savings.[1]

This comparative study aims to examine the adoption and utilization of cloud computing within the healthcare industry, focusing on its implications for healthcare delivery, data management, security, and regulatory compliance. By analyzing the experiences of healthcare organizations across different regions and settings, the study seeks to identify common trends, challenges, and best practices associated with cloud adoption in healthcare.

The introduction of cloud computing in healthcare represents a paradigm shift in how healthcare data is stored, accessed, and utilized. Traditional on-premises IT infrastructure often faces limitations in terms of scalability, agility, and cost-effectiveness, hindering the seamless exchange of information and the implementation of innovative healthcare solutions.

[2] Cloud computing offers a compelling alternative by providing healthcare organizations with the ability to scale resources dynamically, optimize IT spending, and leverage advanced analytics and machine learning algorithms for actionable insights.

However, the adoption of cloud computing in healthcare is not without challenges. Security concerns, privacy regulations, interoperability issues, and data governance remain significant barriers to widespread adoption. Healthcare organizations must navigate complex regulatory landscapes, such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in Europe, to ensure compliance and protect sensitive patient information in the cloud.

Furthermore, the comparative study acknowledges the diversity of cloud deployment models, including public, private, and hybrid clouds, each with its own advantages and considerations for healthcare organizations. [3] While public clouds offer scalability and cost-effectiveness, private clouds provide greater control and security, making them suitable for organizations with strict regulatory requirements. Hybrid cloud environments, combining on-premises infrastructure with cloud services, offer a balance between flexibility and control, allowing healthcare organizations to leverage cloud benefits while maintaining critical data on-premises. Cloud computing in the healthcare industry holds immense potential to drive innovation, improve patient outcomes, and enhance operational efficiency. [5] However, realizing these benefits requires careful consideration of various factors, including security, compliance, interoperability, and the selection of appropriate cloud deployment models. Through this comparative study, we aim to provide valuable insights and recommendations to support informed decision-making and successful cloud adoption in healthcare.

II. CLOUD COMPUTING

Cloud computing refers to the delivery of computing services over the internet, providing users with access to a wide range of resources, including data storage, processing power, and software applications, without the need for local infrastructure or hardware.[12] In essence, cloud computing allows individuals and organizations to leverage shared pools of resources hosted remotely by service providers.

Key characteristics of cloud computing include:

On-Demand Self-Service: Users can provision and access computing resources, such as storage and applications, on-demand without requiring human intervention from the service provider.

Broad Network Access: Cloud services are accessible over the internet from a variety of devices, including desktop computers, laptops, tablets, and smartphones.

Resource Pooling: Computing resources are pooled together to serve multiple users, enabling efficient resource utilization and scalability.

Rapid Elasticity: Cloud resources can be rapidly scaled up or down in response to changing demand, allowing users to accommodate fluctuations in workload without disruption.

Measured Service: Cloud usage is typically metered and billed based on consumption, allowing users to pay only for the resources they use.

Cloud computing can be categorized into several deployment models:

Public Cloud: Services are provided over the internet and shared among multiple users by a third-party cloud service provider. Examples include Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP).

Private Cloud: Computing resources are dedicated to a single organization and hosted either on-premises or by a third-party provider. Private clouds offer greater control, security, and customization but may require higher initial investment.

Hybrid Cloud: Combines elements of both public and private clouds, allowing organizations to integrate on-premises infrastructure with cloud services to achieve greater flexibility, scalability, and data sovereignty.

Cloud computing offers numerous benefits to users and organizations, including:

Scalability: Cloud resources can be scaled up or down dynamically to meet changing demands, enabling organizations to efficiently manage peak workloads and optimize resource utilization.

Cost Savings: Cloud computing eliminates the need for upfront capital investment in hardware and infrastructure, reducing operational costs and enabling organizations to pay only for the resources they use.

Flexibility and Agility: Cloud services provide users with the flexibility to deploy and access computing resources from anywhere with an internet connection, enabling remote work and collaboration.

Reliability and Redundancy: Cloud providers typically offer high levels of uptime and redundancy through distributed data centers and advanced infrastructure, minimizing the risk of downtime and data loss.

Innovation and Collaboration: Cloud computing enables organizations to leverage advanced technologies, such as artificial intelligence, machine learning, and big data analytics, to drive innovation and gain competitive advantage.

Cloud computing has become an indispensable tool for individuals and organizations seeking to harness the power of technology to drive business growth, innovation, and digital transformation.

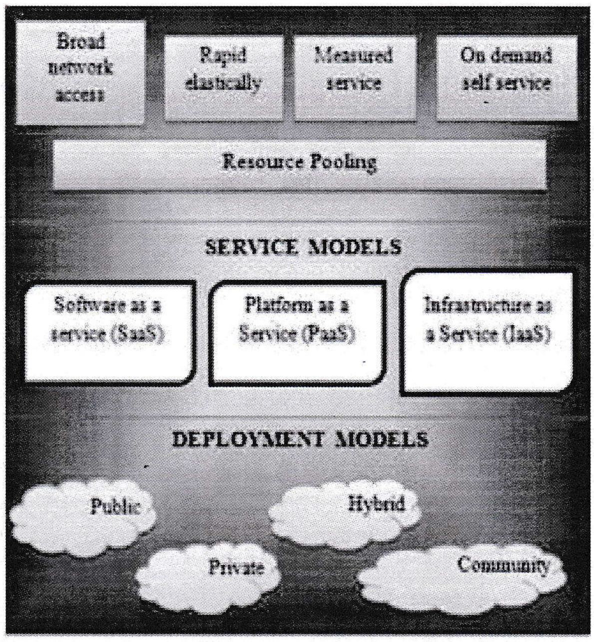


Fig 1: Definition of cloud computing.

III. CLOUD COMPUTING BASED HEALTHCARE SERVICES

Cloud computing-based healthcare services refer to the use of cloud computing technologies to deliver various healthcare-related solutions and applications. [14] These services leverage the scalability, flexibility, and cost-effectiveness of cloud infrastructure to streamline healthcare delivery, improve patient care, and enhance operational efficiency. Here are some examples of cloud computing-based healthcare services:

Electronic Health Records (EHR) Systems: Cloud-based EHR systems allow healthcare providers to securely store, access, and manage patient health information over the internet. These systems enable real-time access to patient records, facilitate seamless information exchange between healthcare providers, and support interoperability across different healthcare settings.

Telemedicine and Telehealth Platforms: Cloud-based telemedicine platforms enable remote consultations, diagnosis, and treatment delivery via video conferencing and other communication technologies. These platforms allow patients to access healthcare services from the comfort of their homes, improve access to care in remote or underserved areas, and reduce healthcare costs.

Health Information Exchange (HIE) Platforms: Cloud-based HIE platforms facilitate the secure exchange of patient health information between healthcare organizations, such as hospitals, clinics, and pharmacies. These platforms promote care coordination, enhance patient safety, and support population health management initiatives by providing a comprehensive view of patient health data across different care settings.

Medical Imaging Solutions: Cloud-based medical imaging solutions allow healthcare providers to store, analyze, and share diagnostic images, such as X-rays, MRIs, and CT scans, using cloud infrastructure. These solutions enable remote access to imaging studies, facilitate collaboration between radiologists and other healthcare professionals, and support advanced image analysis techniques, such as machine learning-based image interpretation.

Healthcare Analytics and Business Intelligence: Cloud-based healthcare analytics platforms leverage big data technologies to analyze large volumes of healthcare data and generate actionable insights for decision-making. These platforms help healthcare organizations identify trends, patterns, and opportunities for improvement in patient care, population health, and operational efficiency.

Clinical Decision Support Systems (CDSS): Cloud-based CDSS provide healthcare providers with evidence-based clinical guidelines, alerts, and recommendations at the point of care. These systems help improve diagnostic accuracy, prevent medical errors, and enhance patient outcomes by assisting clinicians in making informed decisions based on the latest medical knowledge and best practices.

Healthcare IoT Solutions: Cloud computing enables the integration of Internet of Things (IoT) devices, such as wearable sensors, remote monitoring devices, and smart medical devices, into healthcare systems. These IoT solutions collect real-time patient data, transmit it to the cloud for analysis, and generate personalized insights and recommendations for patients and healthcare providers.

Therefore, Cloud computing-based healthcare services offer numerous benefits, including improved accessibility, scalability, interoperability, and cost-efficiency, driving innovation and transformation in the healthcare industry. However, it is essential to address security, privacy, and regulatory compliance considerations to ensure the safe and effective use of cloud-based solutions in healthcare.

IV. ADVANTAGES OF ADOPTING CLOUD HEALTHCARE ORGANIZATION

Adopting cloud computing offers several advantages for healthcare organizations, including:[14]

Scalability: Cloud computing allows healthcare organizations to scale their IT infrastructure and resources up or down based on demand. This scalability enables organizations to accommodate fluctuations in data storage needs, processing power, and user access, ensuring optimal performance without the need for costly hardware investments or over-provisioning.

Cost Efficiency: Cloud computing eliminates the need for healthcare organizations to invest in and maintain on-premises hardware and infrastructure, reducing upfront capital expenses and ongoing operational costs. Cloud services typically operate on a pay-as-you-go model, allowing organizations to pay only for the resources they use, which can result in significant cost savings over time.

Accessibility and Flexibility: Cloud-based healthcare solutions enable remote access to data, applications, and services from any location with an internet connection. This accessibility and flexibility support remote work, telemedicine, and mobile healthcare initiatives, empowering healthcare professionals to deliver care more efficiently and effectively while improving patient access to services.

Enhanced Collaboration and Communication: Cloud computing facilitates collaboration and communication among healthcare providers, patients, and other stakeholders by providing centralized access to shared data and applications. Cloud-based collaboration tools, such as secure messaging, video conferencing, and document sharing, enable real-time communication and information exchange, improving care coordination and decision-making.

Improved Data Management and Security: Cloud computing offers robust data management capabilities, including automated backups, data replication, and disaster recovery, to ensure the integrity, availability, and reliability of healthcare data. Cloud providers implement advanced security measures, such as encryption, identity and access management, and threat detection, to protect sensitive patient information from unauthorized access, breaches, and cyber threats.

Innovation and Agility: Cloud computing enables healthcare organizations to rapidly deploy and integrate new technologies, applications, and services to support innovation and digital transformation initiatives. Cloud-based platforms provide access to advanced technologies, such as artificial intelligence, machine learning, and big data analytics, which can drive insights, improve patient outcomes, and optimize operational processes.

Compliance and Regulatory Support: Cloud providers often offer compliance certifications and adherence to industry-specific regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in Europe. Cloud computing can help healthcare organizations meet regulatory requirements, maintain compliance, and mitigate legal and financial risks associated with data privacy and security.

Overall, adopting cloud computing offers numerous benefits for healthcare organizations, including scalability, cost efficiency, accessibility, enhanced collaboration, improved data security, innovation, and regulatory compliance. By leveraging cloud-based solutions, healthcare organizations can modernize their IT infrastructure, streamline operations, and deliver higher-quality care to patients while driving efficiency and cost savings.

V. Current State Of Cloud Computing In Healthcare

The current state of cloud computing in healthcare is characterized by increasing adoption and recognition of its transformative potential across the industry. [15] Here's an overview of the key aspects defining the current landscape:

Adoption Trends: Healthcare organizations, including hospitals, clinics, insurers, and pharmaceutical companies, are increasingly adopting cloud computing solutions to modernize their IT infrastructure, streamline operations, and improve patient care. According to various reports and surveys, the adoption of cloud computing in healthcare is steadily growing, with a significant portion of healthcare organizations either already using cloud services or planning to do so in the near future.

Use Cases and Applications: Cloud computing is being utilized across a wide range of healthcare applications, including electronic health records (EHR) systems, telemedicine platforms, health information exchange (HIE) networks, medical imaging solutions, healthcare analytics, clinical decision support systems (CDSS), and Internet of Things (IoT) devices. These cloud-based solutions enable healthcare organizations to enhance accessibility, scalability, collaboration, and innovation in care delivery.

Benefits Realization: Healthcare organizations are experiencing tangible benefits from adopting cloud computing, including cost savings, improved scalability and flexibility, enhanced data security and privacy, streamlined workflows, better interoperability, and access to advanced technologies such as artificial intelligence (AI) and machine learning (ML). Cloud-based solutions are enabling healthcare organizations to achieve operational efficiencies, optimize resource utilization, and deliver higher-quality care to patients.

Security and Compliance Considerations: Despite the benefits, security and compliance remain top concerns for healthcare organizations adopting cloud computing. Data security breaches, compliance with regulations such as the Health Insurance Portability and Accountability Act (HIPAA), and data residency requirements are critical considerations when migrating sensitive patient health information to the cloud. Cloud service providers are responding to these concerns by implementing robust security measures, encryption technologies, and compliance certifications to protect healthcare data and ensure regulatory compliance.

Emerging Trends and Technologies: The current state of cloud computing in healthcare is also characterized by emerging trends and technologies shaping the future of the industry. These include the adoption of hybrid cloud environments to balance the needs for flexibility and control, the integration of AI and ML algorithms for predictive analytics and personalized medicine, the use of blockchain for secure data exchange and identity management, and the expansion of cloud-based telemedicine and remote patient monitoring solutions in response to the COVID-19 pandemic.

The current state of cloud computing in healthcare is marked by increasing adoption, expanding use cases, realization of benefits, and ongoing efforts to address security and compliance challenges. As healthcare organizations continue to leverage cloud technologies to innovate and transform care delivery, collaboration between stakeholders, investment in cybersecurity, and adherence to regulatory requirements will be crucial for realizing the full potential of cloud computing in healthcare.

VI. CURRENT STATE OF CLOUD COMPUTING IN HEALTHCARE

Data Collection:

Possible disease indications derive from wearable sensors. Heart rate, respiration percentage, physical behavior, body temperature, oxygen saturation amount, cough problems, and tension are all measured using a variety of sensors. The wearable gadgets that may be used to identify and track patients, including accelerometer, thermometer, Global positioning system (GPS), electrocardiogram (ECG), and oxygen saturation. Sensors transmit the biological readings to the cloud, where further computation and research forecast the patients' conditions

Data preprocessing using normalization:

Preprocessing is the process of transforming raw data into a form that computers can interpret and analyze as part of the data extraction and research process. Data from the real world, including text, images, videos, and other forms, are mixed. In addition to being difficult, it lacks a logical framework and is filled with errors and contradictions.

Cloud database:

The suggested scheme used a variety of healthcare data types while determining recommendations. Meanwhile, IoT systems gather health information for numerous patients from several faraway locations for evaluation. For evaluation, this information should be kept on a server. It is difficult to preserve and manage such a big amount of information. The cloud infrastructure offers enough area for preserving this significant amount of information. Hadoop can scale and preserve this huge amount of information.

In this part, the effectiveness of the suggested illness detection in remote healthcare is examined. Accuracy, sensitivity, specificity, and computation time are among the performance criteria used for evaluation. Convolutional neural network (CNN), Deep-learning Diabetic Retinopathy (DeepDR), Extended Kalman Filter with Support Vector Machine (EKF-SVM), and Bagging Ensemble with K-Nearest Neighbor (BKNN) are the methods used for comparison.

Accuracy:

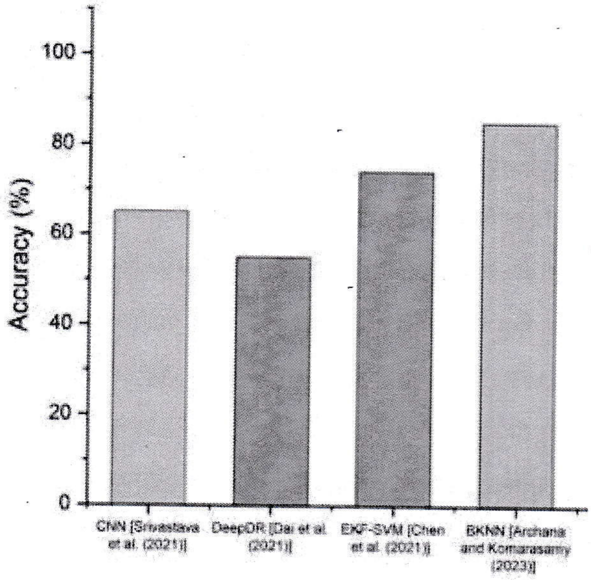


Table 1. Findings of accuracy

Methods	Accuracy (%)
CNN [Srivastava et al. (2021)]	65
DeepDR [Dai et al. (2021)]	55
EKF-SVM [Chen et al. (2021)]	74
BKNN [Archana and Komarasamy (2023)]	85

Sensitivity:

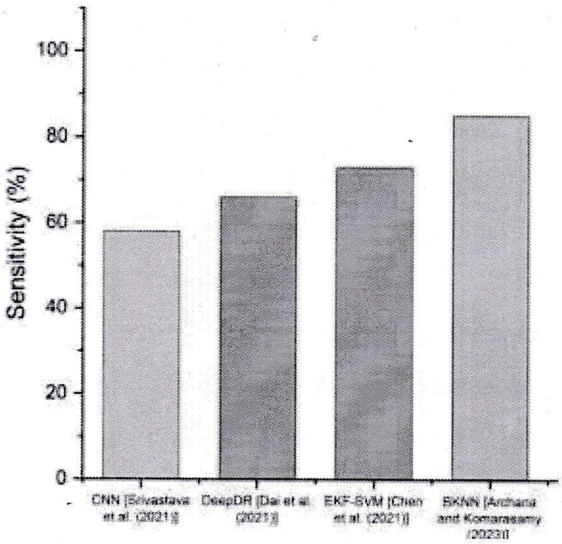


Table 2. Findings of sensitivity

Methods	Sensitivity (%)
CNN [Srivastava et al. (2021)]	58
DeepDR [Dai et al. (2021)]	66
EKF-SVM [Chen et al. (2021)]	73
BKNN [Archana and Komarasamy (2023)]	85

Specificity:

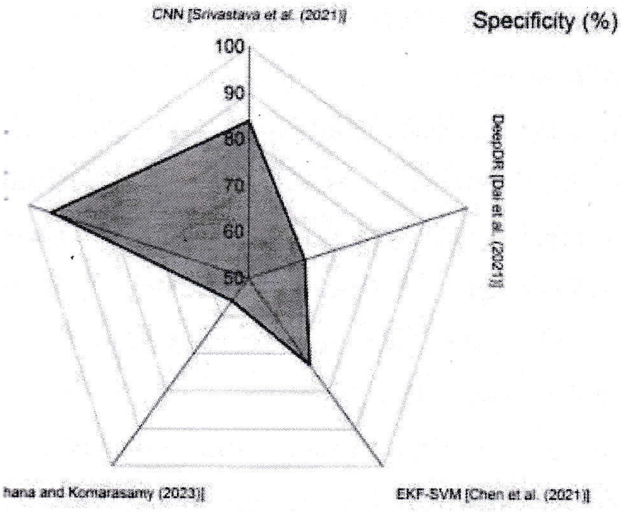


Table 3. Findings of specificity

Methods	Specificity (%)
CNN [Srivastava et al. (2021)]	84
DeepDR [Dai et al. (2021)]	63
EKF-SVM [Chen et al. (2021)]	73
BKNN [Archana and Komarasamy (2023)]	56

VII. CONCLUSIONS AND FUTURE SCOPE

From the table 1,

- Convolutional neural network (CNN), - 65%
- Deep-learning Diabetic Retinopathy (DeepDR)- 55%
- Extended Kalman Filter with Support Vector Machine (EKF-SVM)- 74%
- Bagging Ensemble with K-Nearest Neighbor (BKNN) - 85%
- Therefore, BKNN is provide the good accuracy result.

From the table 2,

- Convolutional neural network (CNN), - 58%
- Deep-learning Diabetic Retinopathy (DeepDR)- 66%
- Extended Kalman Filter with Support Vector Machine (EKF-SVM)- 73%
- Bagging Ensemble with K-Nearest Neighbor (BKNN) - 85%
- Therefore, BKNN is provide the good sensitivity result.

From the table 3,

- Convolutional neural network (CNN), - 84%
- Deep-learning Diabetic Retinopathy (DeepDR)- 63%
- Extended Kalman Filter with Support Vector Machine (EKF-SVM)- 73%
- Bagging Ensemble with K-Nearest Neighbor (BKNN) - 56%
- Therefore, BKNN is provide the good specificity result.

This comparative study on cloud computing in the healthcare industry has shed light on the transformative potential, benefits, challenges, and future prospects of adopting cloud technologies in healthcare organizations. Through an analysis of existing literature, case studies, and empirical data, key findings and insights have been synthesized to inform strategic decision-making and policy formulation for stakeholders in the healthcare sector.

The study has highlighted several important conclusions:

Benefits of Cloud Computing: Cloud computing offers numerous advantages for healthcare organizations, including scalability, cost efficiency, accessibility, enhanced collaboration, improved data management, security, innovation, and regulatory compliance.

Challenges and Considerations: Despite the benefits, the adoption of cloud computing in healthcare is not without challenges. Security concerns, privacy regulations, interoperability issues, data governance, and regulatory compliance remain significant barriers to widespread adoption. Healthcare organizations must carefully consider these factors and implement appropriate measures to ensure the safe and effective use of cloud-based solutions.

Diversity of Cloud Deployment Models: Hybrid cloud environments offer a balance between flexibility and control, allowing organizations to leverage cloud benefits while maintaining critical data on-premises.

Role of Emerging Technologies: The study has highlighted the role of emerging technologies, such as artificial intelligence, machine learning, big data analytics, and the Internet of Things, in enhancing the capabilities and functionalities of cloud-based healthcare systems.

Recommendations for Future Research and Implementation: Moving forward, there is a need for further research and collaboration to address the remaining challenges and unlock the full potential of cloud computing in healthcare. Future studies should focus on evaluating the long-term impact of cloud adoption on patient care, clinical outcomes, cost-effectiveness, and healthcare delivery models. Additionally, healthcare organizations should prioritize investments in cybersecurity, data privacy, interoperability, and talent development to ensure successful cloud implementations.

Overall, this comparative study contributes to a deeper understanding of the opportunities and challenges associated with cloud computing in the healthcare industry. By synthesizing insights from diverse perspectives, the study aims to inform strategic decision-making and support the successful adoption and implementation of cloud-based solutions to drive innovation, improve patient care, and transform healthcare delivery.

REFERENCES

[1] World Health Organization, A Glossary of Terms for Community Health Care and Services for Older Persons. Who Centre For Health Development Ageing and Health Technical Report, 5, World Health Organization, Japan.

[2] Invest healthcare Iskandar Malaysia, Asia's Healthcare Destination of Choice. Iskandar Regional Development Authority (IRDA).

[3] Inside Malaysia, Focus on healthy growth in the medical and biotech sector, Healthcare, 129-132

[4] Bamiah, M., Brohi, S., Chuprat, S., & Ab Manan, J. L. A study on significance of adopting cloud computing paradigm in healthcare sector. In Cloud Computing

Technologies, Applications and Management (ICCCTAM). International Conference on IEEE.

[5] Peter, M., & Timothy, G. Recommendations of the National Institute of Standards and Technology. National Institute of Standards and Technology Special Publication 800-145, 1-3

[6] Tamil Ilakkiya N. S. Role of Cloud in Improving Patient Care, International Journal of Advanced Research in Computer Science and Software Engineering, 5(3), pp. 171-175.

[7] Cloud Standards Customer Council, Impact of Cloud Computing on Healthcare, White Paper. Retrieved from http://www.cloud- council.org/deliverables/CSCC-Impact-of-Cloud- Computing-on-Healthcare.pdf

[8] Al Masud, & Rashid, S. M. A Novel Approach to Introduce Cloud Services in Healthcare Sectors for the Medically Underserved Populations in South Asia. International Journal of Engineering Research and Applications, 2(3), pp. 1337-1346.

[9] De la Torre-Díez, I., Díaz-Pernas, F. J., Fernández, G., Antón-Rodríguez, M., Martínez-Zarzuela, M., González-Ortega, D., & Boto-Giralda, D. . Analysis of the benefits and constraints for the implementation of Cloud Computing over a EHRs system. In Proceedings of the 6th Euro American Conference on Telematics and Information Systems, pp. 151-156.

[10] Chen, T. S., Liu, C. H., Chen, T. L., Chen, C. S., Bau, J. G., & Lin, T. C. Secure Dynamic access control scheme of PHR in cloud computing. Journal of medical systems,36(6),pp.4005-4020.

[11] Mell, P., & Grance, T. The NIST definition of cloud computing. National Institute of Standards and Technology, 53(6), p. 50.

[12] Sanjay, P. A., Sindhu M., & Jesus Z. A Survey of the State of Cloud Computing in Healthcare, Network and Communication Technologies, 1(2).

[13] Wang, X., Application of Cloud Computing in the Health Information System. Computer Application and System Modeling (ICCASM).

[14] Tamil Ilakkiya N. S. Role of Cloud in Improving Patient Care, International Journal of Advanced Research in Computer Science and Software Engineering, 5(3), pp. 171-175.

[15] Rolim, C. O., Koch, F. L., Westphall, C. B., Werner, J., Fracalossi, A., & Salvador, G. S. A cloud computing solution for patient's data collection in healthcare institutions. In e-Health, Telemedicine, and Social Medicine, ETELEMED'10, Second International Conference on IEEE, pp. 95-99.

[16] Maslin, M., & Ailar, R., Cloud Computing Adoption in the Healthcare Sector: A SWOT Anaylsis, Asian Social Science, 11(10), pp. 12-17.

[17] Linthicum, D. The true rate of cloud adoption in healthcare,Logicworks.

[18] Shetty, S. Gartner Says Cloud Computing Will Become the Bulk of New IT Spend.

[19] Saeid, A., Zohreh, S., Ali, T., Steven, R., Abdullah, G., Samee, U. K.,. Cloud Adoption in Malaysia: Trends, Opportunities, and Challenges, IEEE Cloud Computing, IEEE Computer Society.

[20] Nurbaini, Z., Nurazean, M., Wardah, Z. A., NurulHuda, F. A., Ganthan N. S, Cloud Computing Adoption Conceptual Model of Malaysia Hospitals, Open International Journal of Infromatics, 3(1), pp. 1-10.

[21] H. Marwaha, & Singh, R.,, "Deploying cloud applications in educational organizations," International journal of advanced research in computer science and software engineering, vol. 3, pp. 227-29, 2013.

[22] S. Mathew, "Implementation of cloud computing in education- a revolution," International journal of computer theory and engineering, vol. 4, pp. 473-75, 2012.

[23] P. Mell, & Grance, T.,, "NIST definition of cloud computing," presented at National Institute of Standards and Technology, US department of Commerce, 2011.

[24] E. Mitchell, "Using cloud services for library IT infrastructure," 2010.

[25] K. Siau, "Health care informatics," IEEE transactions on information technology in biomedicine, vol. 7, 2003.

A SYSTEMATIC ANALYSIS OF DIFFERENT MACHINE LEARNING & DEEP LEARNING TECHNIQUES USED FOR PREDICTING CYCLONES

Mr.Abhishek A K

Saintgits College Of Applied Sciences

Pathamuttom,Kottayam, Kerala – 686532

cptabhishekak47@gmail.com

Mr.Bijesh P B

Saintgits College Of Applied Sciences

Pathamuttom,Kottayam, Kerala – 686532

bijeshpbinu@gmail.com

Mr. Alwin

Saintgits College Of Applied Sciences

Pathamuttom,Kottayam, Kerala – 686532

alwin6000@gmail.com

Mr.Dhananjay Suresh

Saintgits College of Applied Sciences

Pathamuttom,Kottayam,Kerala – 686532

dhananajays75@gmail.com

Ms.Johnsymol Joy

Assistant Professor

Saintgits College of Applied Sciences

Pathamuttom,Kottayam,Kerala – 686532

johnsymol.joy@saintgits.org

Abstract

Cyclones, with their catastrophic potential, pose enormous hazards to coastline communities around the world. Promising outcomes have been observed in the prediction of cyclone behavior through the incorporation of machine learning techniques in meteorological research in recent years. This research paper presents a systematic analysis of various machine learning *methodologies employed for cyclone prediction. This research deals with the collection and preprocessing of extensive meteorological data, spanning historical cyclone occurrences and associated atmospheric conditions. This research paper conducts a comprehensive assessment of different ML methods for cyclone prediction. The main goal is to assess how well these methods work for predicting important cyclone characteristics including intensity, track, and landfall locations. The objective is to shed light on how various algorithms—including SVM, decision trees, neural networks, etc.—perform.*

Keywords— Machine Learning, supervised learning, multi-class classification algorithms, Tropical Cyclones, Deep Learning, Support vector machines, Decision tree

I. Introduction

The rise in the intensity and frequency of cyclonic occurrences in recent years have highlighted the urgent need for sophisticated and precise prediction techniques. This research paper conducts a thorough analysis of several ML strategies used for cyclone prediction, acknowledging the critical role that ML plays in improving forecasting abilities. As coastal regions grapple with the escalating threats posed by these natural catastrophes, understanding the efficacy of various algorithms in predicting cyclone trajectories, intensities, and landfall locations becomes paramount. This systematic analysis delves into the nuances of neural networks, SVM, decision trees, and ensemble methods, evaluating their performance against rigorous metrics. By shedding light on the strengths and limitations of each approach, this study aims to contribute valuable insights to meteorological research, guiding the refinement of predictive models and fortifying early warning systems to mitigate the impact of cyclones on vulnerable populations.

In the present situation, cyclone prediction is decisive for various reasons- first one is Mitigation of Losses. By enabling authorities to safeguard infrastructure, remove people from high-risk regions, and plan for emergency response, timely and accurate cyclone forecast helps to mitigate possible losses. This lessens the effect on property, life, and the economy. Second one is on humanitarian considerations. Cyclones can cause extensive destruction, uprooting populations, demolishing houses, and interrupting vital services. When human casualties are unlikely, authorities can more accurately plan and execute evacuation procedures. Another one is Infrastructure Planning and Preparedness. Using cyclone forecasts, governments and local authorities can organize and carry out infrastructure projects that are resilient to the effects of cyclones. This entails creating pre-disaster warning systems, enhancing drainage systems, and designing resilient structures. Last one is economic stability. Cyclones can have a serious negative impact on the economy, especially in coastline regions where they are more frequent. Businesses, farmers, and communities can protect their investments, crops, and livelihoods by making educated decisions based on accurate forecasting.

II. Literature Survey

A. Papers on Cyclone prediction

Chinmoy Kar et al.[1] This study highlights the need for computational support for effective forecasting as it examines the difficulty of precisely estimating the intensity of Tropical Cyclones in the NIO region. Using TCs' best track data from 2011 to 2020, the study analyzes the effectiveness of various ML classifiers. Naive Bayes, Logistic Regression, Multilayer Perceptron, Sequential Minimal Optimization, C4.5 Decision Trees, Random Trees, and Random Forests are among the classifiers that were assessed. For classification, five predictors are used: maximum sustained wind speed, latitude, longitude, central pressure, and pressure decrease. According to the results, these machine learning classifiers can achieve classification accuracy of 97% to 99%.

Yuqiao Wu et al.[2] Tropical cyclones have a substantial influence on human lives, needing precise and timely forecasting for catastrophe avoidance. Machine learning techniques have the potential to be more advantageous than traditional numerical forecasting methods, which call for substantial resources. However, current approaches frequently miss important details. With the aim to anticipate tropical storm intensity and route, a multitask ML framework is introduced in this letter. It is composed of an estimating module that makes use of two DNN and a prediction module that makes use of an improved GAN. The efficacy of our suggested approach in forecasting tropical cyclones is demonstrated by its 116-kilometer 24-hour path forecast error and 13.06-knot 24-hour intensity forecast error.

G. Vijayakumar et al It can be difficult to identify and predict weather changes, particularly given how frequently and dramatically the weather can change. The Indian subcontinent has seen more frequent and intense cyclones, making accurate detection vital. ML approaches, favored over traditional methods, rely heavily on manual feature engineering. Deep learning, an advanced ML technique, automatically selects features, removing this barrier. It has proven success in weather forecasting and is distinguished from traditional ML algorithms in this study. Three aspects - modeling inputs, methodology, and preprocessing techniques - are examined. Results highlight the performance of different ML algorithms in predicting rainfall based on meteorological data, improving weather awareness and informed decision-making.

Fan Meng et al.[4] This paper introduces a fresh strategy for accurately estimating cyclone size that uses deep CNN. This is a pioneering use of deep learning approaches in estimating TC size. The dataset used includes around 1,000 TC events and approximately 30,000 IR remote sensing photos. A comparison with proven best track archives demonstrates that the suggested model achieves a mean error of 24 nautical miles, beating the National Oceanic and Atmospheric Administration's (NOAA) Multiplatform Tropical Cyclone Surface Winds Analysis. These findings highlight deep learning approaches' significant potential for refining the veracity of TC size prediction.

Ming Xie et al.[5] With multiple techniques created employing cloud photography, wind field data, and sea level pressure, cyclone identification is entrenched yet constantly developing discipline. The data fusion method mentioned in this article integrates information from several remote sensors. An object detection technique based on DL is used to originate an accurate model globally. For training and testing the model, rainfall intensity data from global precipitation measurement is paired with wind field data obtained from the scatterometer measurements. The model comprises two modules: a feature extractor and a region proposal network based on the feature pyramid network (FPN) to detect potential cyclone areas, and a region of interest processor that refines cyclone locations using a fully-connected neural network and bounding box regression. An ablation experiment confirms the importance of data fusion, with wind field data showing more significant contribution to cyclone detection than precipitation data.

Nahruma Mehzabeen Pieu et al.[6] The study's nine ensembles of high-resolution pressure and wind data for 12 historical TCs in the Bay of Bengal (BoB) are supplied by MOUM. Unprecedented devastation and casualties were wreaked by these storms. A reliable coastal model may be created to anticipate the sites of tropical storm landfalls and their surge heights by precisely modelling these storms using high-resolution atmospheric forcing. This would enable the prompt and accurate broadcast of cyclone information. By examining cyclone courses, landfall locations, and forecast periods in the Bay of Bengal, the study streamlines the process of identifying effective ensembles for the creation of coastal models. Every 72 hours, each ensemble is started, the first 60 hours prior to the cyclone's impact, and the others three hours later to determine the forecast's accuracy.

Akshath Mahajan et al.[7] Every year, tropical storms that originate in the NIO basin pose a threat to India and cause extensive property and human damage Accurate prediction of these catastrophes is critical for implementing prompt preventative measures. Utilizing the

CyINSAT dataset spanning from 2014 to 2022, this study compares multiple techniques for wind speed forecasting. Employing recurrent networks and image feature extractors, the models predict future wind speeds from sequential images. Architectural variances focus on handling current wind speed data. The proposed architecture, which prioritizes current wind speed records, outperforms baseline models, achieving an RMSE of 6.31, MAE of 0.093, and MAPE of 4.53. An intercomparison of cyclone tracks among ensembles is conducted 2 to ascertain the most accurate forecasts.

Pavitha N et al.[8] Cyclones are a serious threat to people's lives and property, which makes it crucial to anticipate their severity with accuracy. A methodical technique has been devised to tackle this problem, which involves creating a database of annotated photos that holds crucial information about cyclones, like names, locations, timings, occurrence years, and matching images. Prior to analysis, the data undergoes preprocessing steps including standardization and data augmentation to ensure uniformity and balance representation. A CNN model is then utilized, trained meticulously over 100 epochs using the rmsprop optimizer. The training process yields pertinent metrics, revealing a mean absolute error of 10.83 knots and a root mean square error of 14.49 knots. The model's efficacy is visually assessed through graphical representations of errors. Subsequently, the trained model is applied to make predictions on test images. Additionally, an interactive web application is developed to visualize past cyclones. Users can input specific years and cyclone names to access relevant information. This all-inclusive solution combines rigorous preprocessing, cutting-edge ML methods, and an intuitive user interface to improve our comprehension and readiness for tropical storms.

A.T.R. Krishna Priya et al.[9] Weather forecasters have long studied cyclones, with scientists researching various aspects such as their structure, dynamics, and prediction methodologies. DL has emerged as a promising approach to address the challenges of cyclone prediction, either through purely data-driven systems or by enhancing traditional numerical methods. Intensity estimation, track forecasting, cyclone generation, and related severe weather occurrences are all explored in this study of the use of DL. Despite the potential of DL and the accessibility of vast multi-source data, current utilization for improving cyclone prediction accuracy remains limited. Cyclones' complexity and susceptibility to various factors pose challenges in leveraging DL effectively, which may impact the reliability and duration of cyclone forecasting.

Adam Agus Kurniawan et al.[10] Tropical cyclone events are on the rise as a result of weather patterns becoming more unpredictable due to global warming. A ML algorithm was created to help in the classification of tropical storm intensity. The system uses SVM for classification and GLCM for feature extraction. First, the color spaces of RGB, Ycbcr, and Grayscale are used to extract 14 GLCM features. Feature combinations of 3, 4, and 5 are then tested in the classification stage using SVM with OAO and OAA coding designs and Gaussian, Linear, and Polynomial kernels. Accuracy is assessed in relation to feature combinations. Using infrared photos, the method classifies tropical cyclone intensity with an accuracy rate of 88%, which is consistent with the Saffir-Simpson Hurricane Wind Scale.

Chong Wang et al.[11] In order to forecast the movement direction of cyclones, often called typhoons, in the Northwestern Pacific basin, a deep CNN specifically designed for this purpose was developed using satellite pictures from Himawari-8 (H-8The CNN model was trained on a dataset of 2250 infrared images that included 97 typhoon events that were documented between 2015 and 2018. The study's mean error in forecasting the typhoon movement angle was 27.8°, highlighting the considerable potential of DL approaches in improving the accuracy of tropical storm track prediction. This was achieved by integrating images from various channels as inputs into the CNN architecture.

A. Geetha et al In several domains, including business econometrics, predictive analytics, big data, and statistical studies, the Auto Regressive Integrated Moving Average (ARIMA) model is commonly used, especially for time series researchGiven the substantial effects Tropical Cyclones (TC) have on coastal communities and human life, forecasting TC tracks and intensities with precision is essential to efficient disaster managementIn this study, a statistical time series modeler (TSM) specifically designed for India's cyclonic storm forecasting is presented. The training and testing phases of the TSM of SPSS (Statistical Package for Social Studies) make use of a dataset consisting of 14 attributes that spans six years (2007-2012). The model, developed using training data from 2007 to 2011, is applied to the testing dataset from 2012. The model is built upon the ARIMA model within the TSM framework of SPSS 20.0.

S. Shakya et al.[13] Satellite images are crucial for weather prediction. DL requires diverse annotated data for effective training. Temporal resolution is enhanced using interpolation and data augmentation. Classical approaches are employed in preprocessing. Testing is done on three different optical flow technologies using various optimization strategies and error estimates. The enriched dataset trains a CNN, achieving over 90% accuracy in cyclone classification and over 84% accuracy in cyclone vortex location. Linear regression is explored for path prediction.

S. Gujral et al.[14] Tropical cyclones (TCs), known as typhoons or hurricanes, are significant weather phenomena across five oceans. Traditional monitoring techniques have been ineffective, making accurate TC intensity estimation crucial for reducing human suffering. Recent interest in innovative image processing methods by data scientists and meteorologists shows promise. A modified CNN architecture yields precise results on a benchmark dataset, demonstrating remarkable stability across scenarios and demonstrating data science's promise for meteorological research.

Tushar Paul et al.[15] With over 90 storms annually worldwide,TCs are highly damaging weather systems in tropical oceans. Swift TC detection and tracking are crucial for advanced warnings. Remote sensing is vital due to their origin far from continents. Our novel deep learning-based technique for TC detection from satellite pictures is composed of three phases: a CNN classifier, a wind velocity filter, and a Mask CNN detector. Bayesian optimization tunes hyperparameters for optimal performance. Results show high specificity (97.59%), precision (97.10%), and accuracy (86.55%) in test images.

III. Results and Discussions

Review papers used various ML, DL, and ensemble techniques for prediction and analysis. Data from several sources, including as satellite imaging, atmospheric measurements, and pacific data, can be integrated by ML and DL models. The ability to fuse data allows for a more thorough comprehension of the complex atmospheric conditions conducive to cyclone formation With the aim of anticipating cyclones, meteorological data must have the ability to capture both temporal and spatial relationships. The modeling of cyclonic systems can be improved by RNNs, Long Short-Term Memory Networks (LSTMs), and CNNs, which are well-suited for handling time-series data and spatial information. Without the requirement for manual modifications, DL & ML models are very well capable of reacting to shifting circumstances and developing patterns. Climate prediction relies heavily on this flexibility because the climate is subject to variations brought on by factors such as climate change and global warming.

These studies confront a number of difficulties, such as poor data quality and accessibility, imbalances in uncommon occurrences and data, unpredictability in both space and time, and data integration, among other issues. ML and DL model performance can be strongly impacted by the standard of meteorological data, including aspects like accuracy, resolution, and coverage. Obtaining extensive and superior quality datasets might be difficult, particularly in isolated or under-monitored areas, denying to train resilient models. Since cyclones are comparatively uncommon occurrences, the model's capacity to generalize to them may be impacted by imbalances in the dataset. Precise predictions depend on training data having a fair representation of cyclonic events. The temporal and spatial diversity exhibited by cyclones in datasets may be challenging to characterize. Variations in cyclone features across seasons, regions, and climate patterns must be accommodated by models. It is a difficult challenge to combine several data sources, such as satellite imaging, atmospheric measurements, and oceanic data, into a coherent dataset for model training. It is essential to create data assimilation techniques so as to successfully combine information from multiple sources.

IV. FUTURE ENHANCEMENTS

Develop models that not only predict cyclone paths but also assess the associated risks, such as storm surge and wind intensity. Improve communication strategies to convey predictions and risks effectively to communities, aiding in better preparedness and response. Foster collaboration between meteorologists, data scientists, and domain experts to enhance the understanding of meteorological phenomena and integrate it with existing ML/DL strategies.

V. CONCLUSION

Comprehensive modeling, ensemble forecasting, and integration of satellite data are the few advances in cyclone prediction that are highlighted in this literature review. AI and ML have the potential to increase accuracy. Challenges persist in long-term prediction and addressing climate change impacts. Continued research and collaboration are necessary to enhance preparedness and mitigate risks in cyclone-prone areas.

VI. REFERENCES

1) Y. Wu, X. Geng, Z. Liu and Z. Shi, "Tropical Cyclone Forecast Using Multitask Deep Learning Framework," in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 6503505, doi: 10.1109/LGRS.2021.3132395.
2) C. Kar and S. Banerjee, "Tropical Cyclone Intensity Prediction Using Best Track Data Over North Indian Ocean By Machine Learning Classifiers," 2021 IEEE International India Geoscience and Remote Sensing Symposium(InGARSS),Ahmedabad, India, 2021, pp.297300,doi:10.1109/InGARSS51564.2021.9792071.
3) G. V. Kumar, B. Vijay, G. S. Babu and B. Ramesh, "Detection of Cyclones and Extreme Weather Events Using Advanced Deep Learning Methods," 2022 Second International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE), Bangalore, India, 2022, pp. 1-6, doi: 10.1109/ICATIECE56365.2022.10046958.

4) F. Meng, P. Xie, Y. Li, H. Sun, D. Xu and T. Song, "Tropical Cyclone Size Estimation Using Deep Convolutional Neural Network," 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 2021, pp. 8472-8475,doi:10.1109/IGARSS47720.2021.9554759.
5) M. Xie, Y. Li and S. Dong, "A Deep-Learning-Based Fusion Approach for Global Cyclone Detection Using Multiple Remote Sensing Data," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 9613-9622, 2022, doi: 10.1109/JSTARS.2022.3219809.
6) N. M. Pieu, A. K. M. Saiful Islam, M. K. Das and F. Abdullah, "Assessing Tropical Cyclone Tracks in the Bay of Bengal," OCEANS 2022, Hampton Roads, Hampton Roads, VA, USA, 2022, pp. 1-5, doi: 10.1109/OCEANS47191.2022.9977334.
7) P. N, A. Mankawade, S. R. M and A. B. P, "Cyclone Intensity Estimation Using CNN," 2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), Erode, India, 2023, pp. 164-169, doi: 10.1109/ICSSAS57918.2023.10331821.
8) A. A. Kurniawan, K. Usman and R. Y. Nur Fuadah, "Classification of Tropical Cyclone Intensity on Satellite Infrared Imagery Using SVM Method," 2019 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob), Bali, Indonesia, 2019, pp. 69-73, doi: 10.1109/APWiMob48441.2019.8964184.
9) C. Wang, Q. Xu, X. Li and Y. Cheng, "CNN-Based Tropical Cyclone Track Forecasting from Satellite Infrared Images," IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 2020, pp. 58115814, doi:10.1109/IGARSS39084.2020.9324408.
10) A. Geetha and G. M. Nasira, "Time series modeling and forecasting: Tropical cyclone prediction using ARIMA model," 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2016, pp. 3080-3086.
11) A. T. R. K. Priya, R. Anuja, R. S. Devi, I. S. Manivannan and N. C. Ramya, "A Comprehensive Investigation on Cyclone Prediction Using Deep Learning Technique," 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT), Kollam, India, 2023, pp. 319-324, doi: 10.1109/ICCPCT58313.2023.10245150.
12) S. Shakya, S. Kumar and M. Goswami, "Deep Learning Algorithm for Satellite Imaging Based Cyclone Detection," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 827-839, 2020, doi: 10.1109/JSTARS.2020.2970253.

13) S. Gujral, D. Rao, A. Khanvilkar, U. Shah and T. Maktum, "TCI-Net: A Deep Learning Approach for Tropical Cyclone Intensity Prediction," 2023 International Conference on Network, Multimedia and Information Technology (NMITCON), Bengaluru, India, 2023, pp. 1-7, doi: 10.1109/NMITCON58196.2023.10276163.

14) T. Paul and D. R, "A Study Of Cyclone Tracking Mechanism Using Deep Learning Techniques," 2023 International Conference on System, Computation, Automation and Networking (ICSCAN), PUDUCHERRY, India, 2023, pp. 1-5, doi: 10.1109/ICSCAN58655.2023.10395440.

15) A. Mahajan, D. Daru, A. Thaker, M. Narvekar and D. Mukhopadhyay, "Forecasting North Indian Ocean Tropical Cyclone Intensity," 2022 Smart Technologies, Communication and Robotics (STCR), Sathyamangalam, India, 2022, pp. 1-4, doi: 10.1109/STCR55312.2022.10009275.

A COMPARATIVE STUDY ON CPU SCHEDULING ALGORITHM

Mr. Deepak M

Assistant Professor

Department of Computer Science and Applications

Naipunnya School of Management

Cherthala, Kerala

deepakpm1985@gmail.com

Ms. Nidhy S Ravi

Assistant Professor

Department of Computer Science and Applications

Naipunnya School of Management

Cherthala, Kerala

nidhy.ravi@gmail.com

Abstract

CPU scheduling algorithms play a critical role in optimizing system performance and resource utilization in modern operating systems. This study presents a comprehensive comparative analysis of various CPU scheduling algorithms to evaluate their efficiency under diverse workload scenarios. The examined algorithms include First Come First Serve (FCFS), Shortest Job Next (SJN), Round Robin (RR), Priority Scheduling, and Multilevel Queue Scheduling (MLQ).

The study employs simulation-based experiments using a custom-built CPU scheduler simulator, which accurately models the behavior of each algorithm under controlled conditions. Metrics such as average waiting time, turnaround time, response time, and throughput are measured and compared across different algorithms to assess their performance characteristics.

Results indicate that no single scheduling algorithm universally outperforms others across all workload types. FCFS exhibits simplicity but suffers from high average waiting times, particularly under heavy loads. SJN demonstrates improved turnaround time by prioritizing shorter jobs but may lead to starvation for longer processes. RR offers fair scheduling by allocating fixed time slices to each process, yet it may incur higher context switch overhead.

Priority Scheduling proves effective in scenarios where process priorities significantly impact system performance. MLQ presents a flexible approach by segregating processes into multiple queues based on priority or other criteria, though it requires careful tuning of parameters.

Furthermore, the study investigates the impact of varying workload characteristics, such as job length, arrival patterns, and priority distributions, on the performance of scheduling algorithms. Through sensitivity analysis, insights are derived regarding the strengths and weaknesses of each algorithm under different operating conditions.

This comparative study provides valuable insights into the suitability of various CPU scheduling algorithms for different computing environments and workload profiles. Understanding the trade-offs inherent in each algorithm facilitates informed decision-making for system designers and administrators in selecting the most appropriate scheduling strategy to meet specific performance objectives. Future research directions may explore hybrid or adaptive scheduling approaches to further optimize system responsiveness and resource utilization.

Keywords— CPU, Algorithms, RR, FCFS, SJF.

INTRODUCTION

In modern computer systems, the efficient allocation of CPU resources is essential for maximizing system throughput, minimizing response times, and ensuring fair access to computational resources among competing processes. [3] CPU scheduling algorithms play a pivotal role in achieving these objectives by determining the order in which processes are executed on the CPU. [4] The choice of a scheduling algorithm significantly influences system performance metrics such as average waiting time, turnaround time, response time, and overall throughput.

This comparative study aims to evaluate the performance of various CPU scheduling algorithms under different workload scenarios. The algorithms under scrutiny include First Come First Serve (FCFS), Shortest Job Next (SJN), Round Robin (RR), Priority Scheduling, and Multilevel Queue Scheduling (MLQ). [5] Each algorithm possesses unique characteristics and trade-offs, making it imperative to analyze their behavior comprehensively across a range of operating conditions.

FCFS, the simplest scheduling algorithm, schedules processes based on their arrival order.

While easy to implement, FCFS may lead to long waiting times, especially for processes with lengthy execution times, resulting in poor overall system performance. SJN, on the other hand, prioritizes shorter jobs, aiming to minimize average waiting time and turnaround time. However, it can potentially starve longer processes, leading to fairness issues. [8]

RR scheduling allocates fixed time slices to each process, ensuring fairness and preventing any single process from monopolizing the CPU for extended periods. [10] Nevertheless, RR introduces overhead due to frequent context switches, impacting system performance, particularly in scenarios with high process churn. Priority Scheduling assigns priorities to processes based on factors such as execution time, importance, or user-defined criteria. This allows critical processes to receive preferential treatment, enhancing system responsiveness in time-sensitive environments.

MLQ scheduling organizes processes into multiple queues based on priority levels or other attributes, allowing for different scheduling policies within each queue. This approach provides flexibility and allows administrators to tailor scheduling behavior to meet specific application requirements. [15] However, effective management of multiple queues and dynamic allocation of processes pose implementation challenges.

Through this comparative study, we aim to assess the strengths and weaknesses of each scheduling algorithm across a spectrum of workload characteristics, including job length, arrival patterns, and priority distributions. [9] By conducting simulation-based experiments and analyzing performance metrics, we seek to provide insights into the suitability of different algorithms for various computing environments.

I. LITERATURE SELECTION METHODOLOGY

Literature selection methodology is the methodology that was followed in literature selection as follow:

A keyword Filter phase

The research articles were searched in the Google Scholar database using the following keywords: (1) Operating System (2) CPU management (3) CPU scheduling algorithms (4) RR scheduling algorithm (5) SJF scheduling algorithm (6) FCFS scheduling algorithm (7) Priority scheduling algorithm.

Abstract Filter phase

At this stage, the abstract is read to determine the research articles relevant to the research. The selection resulted in 40 articles from CPU scheduling algorithms, RR scheduling algorithm, SJF scheduling algorithm, FCFS scheduling algorithm, and Priority scheduling algorithm.

Due to the diversity of the CPU scheduling algorithm, many studies and research have improved and developed this algorithm to increase system performance. The following sections review several papers related to CPU scheduling algorithms where the studies have been categorized into four CPU scheduling algorithm types: RR, SJF, FCFS, and Priority Algorithms.

II. CPU SCHEDULING ALGORITHMS

CPU scheduling algorithms are an integral component of operating systems responsible for determining the order in which processes are executed on the CPU. These algorithms aim to optimize system performance by efficiently allocating CPU resources among competing processes. There are various CPU scheduling algorithms, each with its own characteristics and trade-offs. Some of the commonly used CPU scheduling algorithms include:

First Come First Serve (FCFS):

FCFS is one of the simplest CPU scheduling algorithms, where processes are executed in the order of their arrival.

It is easy to implement but may lead to high average waiting times, especially for processes with long execution times (convoy effect). FCFS is non-preemptive, meaning once a process starts execution, it continues until it completes or blocks.

Shortest Job Next (SJN) or Shortest Job First (SJF):

SJN selects the process with the shortest burst time (execution time) next for execution.

It aims to minimize average waiting time and turnaround time, making it an optimal algorithm in terms of average waiting time.

However, it can lead to starvation for longer processes if shorter processes continuously arrive.

Round Robin (RR):

RR is a preemptive CPU scheduling algorithm where each process is assigned a fixed time slice or quantum.

Processes are executed in a cyclic manner, with each process receiving a time slice of CPU time.

RR ensures fairness among processes and prevents any single process from monopolizing the CPU.

However, it may incur overhead due to frequent context switches, affecting system performance.

Priority Scheduling:

Priority scheduling assigns priorities to processes, and the process with the highest priority is selected for execution.

It can be either preemptive or non-preemptive, depending on whether processes can be interrupted while executing.

Priority scheduling allows for the execution of critical processes first, but it may lead to starvation for lower- priority processes if higher-priority processes continuously arrive.

Multilevel Queue Scheduling (MLQ):

MLQ organizes processes into multiple queues based onpriority levels or other criteria.

Each queue may have its own scheduling algorithm, allowing for different scheduling policies within each queue.

MLQ provides flexibility in managing processes with varying characteristics and requirements.

These are just a few examples of CPU scheduling algorithms, and there are other variations and hybrid approaches designed to address specific system requirements and workload characteristics. The choice of a CPU scheduling algorithm depends on factors such as system goals, workload characteristics, and implementation considerations.

III. RR SCHEDULING ALGORITHM

Round Robin (RR) scheduling is a preemptive CPU scheduling algorithm commonly used in operating systems. It is designed to provide fair allocation of CPU time among multiple processes by dividing CPU time into equal time slices, also known as quanta or time quantum. Each process is assigned a time quantum, and they are executed in a cyclic manner. When a process's time quantum expires, it is preempted, and the CPU is given to the next process in the ready queue.

Key features of the Round Robin scheduling algorithm include:

Preemptive: RR is a preemptive scheduling algorithm, meaning that processes can be interrupted while executing if their time quantum expires. This allows for fair allocation of CPU time among competing processes.

Time Slicing: CPU time is divided into fixed-size time slices, and each process is allocated a time quantum. When a process starts executing, it is allowed to run for its time quantum or until it voluntarily relinquishes the CPU (e.g., by blocking or completing). If the process doesn't finish within its time quantum, it is preempted, and the CPU is given to the next process in the queue.

Cyclic Execution: Processes are executed in a cyclic manner, where each process gets a turn to execute for its allocated time quantum. Once a process's time quantum expires, it is placed back at the end of the ready queue, and the next process in line is scheduled for execution.

Fairness: RR scheduling aims to provide fairness among processes by ensuring that no single process monopolizes the CPU for an extended period. Each process gets an equal share of CPU time, as defined by the time quantum.

Frequent Context Switches: RR scheduling may incur overhead due to frequent context switches between processes, especially when the time quantum is relatively small. Context switches involve saving the state of the currently running process and loading the state of the next process to be executed, which introduces some overhead.

Response Time: RR scheduling typically offers good response time for interactive processes since each process gets a chance to execute within a short time quantum.

Round Robin scheduling strikes a balance between fairness and responsiveness, making it suitable for time-sharing systems where multiple users or processes require access to the

CPU. However, the choice of an appropriate time quantum is crucial to achieving optimal performance, balancing fairness and minimizing overhead due to context switches.

IV. SJF SCHEDULING ALGORITHM

The Shortest Job First (SJF) scheduling algorithm is a non- preemptive or preemptive CPU scheduling algorithm that selects the process with the shortest burst time (execution time) for execution. In SJF, the ready queue is ordered based on the burst time of processes, with the shortest job at the front. When a process completes its execution or blocks, the scheduler selects the next process with the shortest burst timefrom the ready queue.

Key features of the Shortest Job First scheduling algorithm include:

Non-Preemptive and Preemptive Variants: SJF can be implemented as either non-preemptive or preemptive. In the non-preemptive version, once a process starts executing, it continues until it completes its CPU burst or blocks. In the preemptive version, if a new process arrives with a shorter burst time than the currently executing process, the schedulermay preempt the running process and start executing the newone.

Optimality: SJF is provably optimal in terms of average waiting time among all scheduling algorithms. This means that SJF minimizes the average waiting time, leading to better overall system performance compared to otherscheduling algorithms.

Starvation: While SJF ensures optimal average waiting time, it may lead to starvation for longer processes. If shorter processes continuously arrive, longer processes may wait indefinitely, unable to execute until all shorter processescomplete.

Prediction of Burst Time: In real-world scenarios, predicting the burst time of processes accurately can be challenging. Implementations of SJF often use estimated burst times based on historical data or heuristics to make scheduling decisions.

Aging: To mitigate the issue of starvation, some variants of SJF incorporate aging mechanisms. Aging increases the priority of processes waiting in the ready queue for an extended period, ensuring that all processes eventually get a chance to execute.

Short Response Time for Short Jobs: SJF provides short response times for short jobs, making it suitable for interactive systems where quick response to user inputs is essential.

The Shortest Job First scheduling algorithm is highly efficient in terms of minimizing average waiting time, making it well-suited for environments where turnaround time is a critical metric. However, it requires accurate estimation or prediction of process burst times and may suffer from the potential issue of starvation for longer processes. Variants of SJF with preemptive features or agingmechanisms aim to address these challenges while retaining the algorithm's efficiency.

V. FCFS SCHEDULING ALGORITHM

The First Come First Serve (FCFS) scheduling algorithm is one of the simplest CPU scheduling algorithms used in operating systems. In FCFS, processes are executed in the order they arrive in the ready queue. The process that arrives first is the first to be executed, hence the name "First Come First Serve."

Key features of the First Come First Serve scheduling algorithm include:

Non-Preemptive: FCFS is a non-preemptive scheduling algorithm, meaning that once a process starts executing, it continues until it completes its CPU burst or blocks for I/O. Other processes cannot preempt it and take over the CPU until the currently executing process finishes.

FIFO Queue: Processes are placed in a FIFO (First-In-First- Out) queue as they arrive. The process at the front of the queue is the one selected for execution by the CPU scheduler.

Simple Implementation: FCFS is straightforward to implement, as it only requires a simple queue data structure to maintain the order of processes. However, its simplicity comes with potential drawbacks in terms of performance.

Convoy Effect: FCFS scheduling may suffer from the convoy effect, where long CPU-bound processes hold up short CPU-bound processes that arrive later. This can result in inefficient use of CPU time and increased average waitingtime for processes.

Low Overhead: FCFS scheduling has low overhead compared to some other scheduling algorithms because it does not require frequent context switches or complex priority calculations.

Fairness: FCFS is fair in the sense that it follows a strict order of execution based on arrival time. However, fairness does not always translate to optimal performance, especially if long processes arrive first and cause shorter processes to wait for an extended period.

Performance Limitations: FCFS scheduling may not be suitable for time-sharing systems or environments where responsiveness and turnaround time are critical metrics. Long-running processes can monopolize the CPU, leading topoor overall system performance.

Overall, while the First Come First Serve scheduling algorithm is easy to understand and implement, its performance may be suboptimal in many scenarios due to thepotential for the convoy effect and lack of prioritizationbased on process characteristics. It is commonly used in batch processing systems or environments where simplicity is preferred over performance optimization.

VI. PRIORITY SCHEDULING ALGORITHM

Priority scheduling is a CPU scheduling algorithm where each process is assigned a priority. The CPU scheduler selects the process with the highest priority for execution. In priority scheduling, processes with higher priority are given preference over processes with lower priority. The priority of a process can be determined based on various factors, such as the importance of the process, its execution time, its deadline, or user-defined criteria.

Key features of the priority scheduling algorithm include:

Priority Assignment: Each process is assigned a priority value, which determines its order of execution. Processes with higher priority values are executed before those with lower priority values.

Preemptive and Non-Preemptive Variants: Priority scheduling can be implemented as either preemptive or non- preemptive. In preemptive priority scheduling, a currently executing process may be preempted if a higher-priority process becomes ready to execute. In non-preemptive priority scheduling, once a process starts executing, it continues until it completes or blocks, regardless of the arrival of higher-priority processes.

Dynamic Priority Adjustment: Priority values may be static or dynamically adjusted during runtime based on changes in process characteristics or system conditions. For example, aging mechanisms may be employed to increase the priority of processes that have been waiting for an extended period toprevent starvation.

Fairness and Responsiveness: Priority scheduling allows for the execution of critical or time-sensitive processes withhigher priority, ensuring that important tasks are completed promptly. However, it may lead to starvation for lower- priority processes if higher-priority processes continuously arrive.

Starvation Mitigation: To mitigate the issue of starvation, priority scheduling algorithms may incorporate aging mechanisms or priority boosting strategies. Aging mechanisms gradually increase the priority of waiting processes over time, ensuring that all processes eventually get a chance to execute.

Implementation Complexity: Priority scheduling algorithms may be more complex to implement compared to some other scheduling algorithms, especially when considering dynamicpriority adjustments and handling priority inversion or priority inversion problems.

Performance Implications: The performance of priority scheduling depends on the accuracy of priority assignments and the effectiveness of priority adjustment mechanisms. Poorly assigned priorities or inadequate handling of priority inversion may degrade system performance.

Priority scheduling is suitable for environments where different processes have varying levels of importance or urgency. It allows for the efficient execution of critical tasks while ensuring fairness and responsiveness. However, careful consideration must be given to priority assignment strategies and the management of priority-related issues to achieve optimal performance.

VII.COMPARSIONS CHARTS FOR USAGES

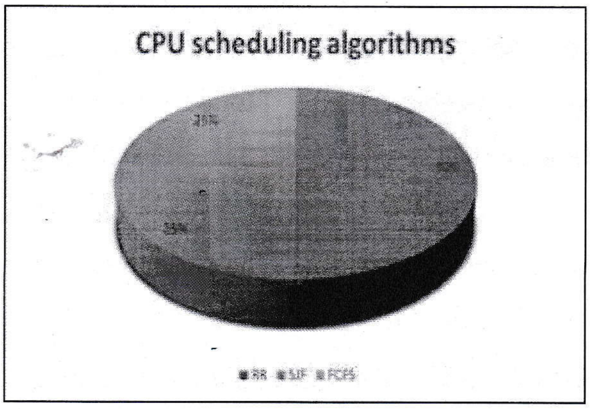
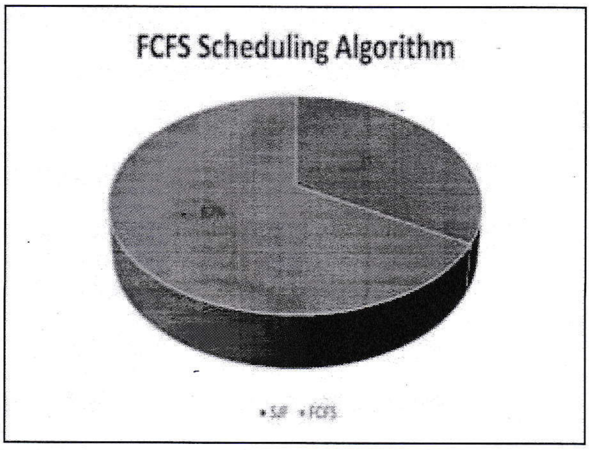


Fig. 1. The percentage of studies on the CPU schedulingalgorithm



The percentage of studies on the RR scheduling algorithm

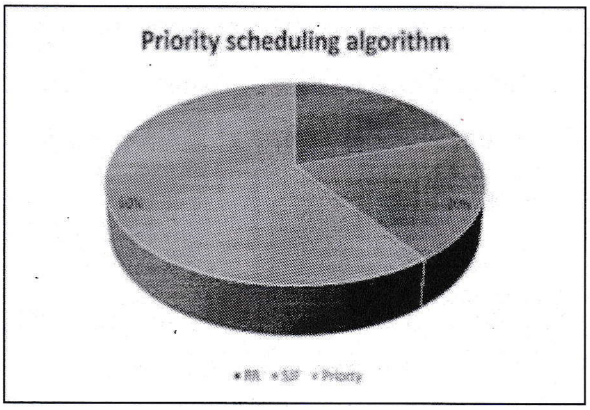


Fig. 4. The percentage of studies on the FCFSscheduling algorithm

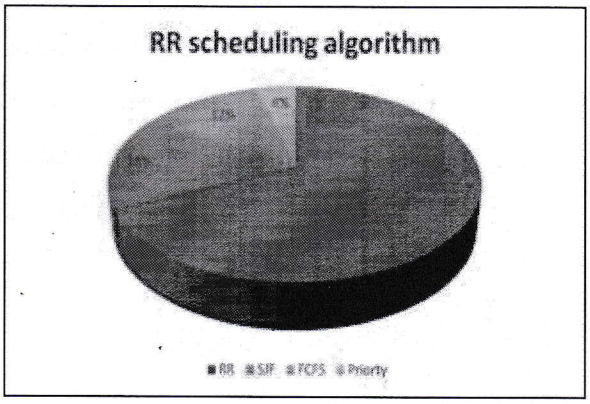


Fig. 5. The percentage of studies on the Priorityscheduling algorithm

VIII.CONCLUSIONS AND FUTURE SCOPE

In conclusion, this comparative study sheds light on the nuanced performance characteristics of various CPU scheduling algorithms across diverse workload scenarios. Through meticulous simulation-based experiments and analysis, it becomes evident that no single algorithm reigns supreme in all situations. Each algorithm exhibits its own set of strengths and weaknesses, necessitating careful consideration of system requirements and workload profiles when selecting a scheduling strategy.

First Come First Serve (FCFS) offers simplicity but suffers from prolonged waiting times, particularly under heavy loads. Shortest Job Next (SJN) prioritizes shorter jobs for quicker turnaround, yet may lead to starvation for longer processes. Round Robin (RR) ensures fairness with fixed time slices but incurs overhead due to frequent context switches. Priority Scheduling proves effective in scenarios where process priorities significantly impact performance. Multilevel Queue Scheduling (MLQ) presents a flexible approach but requires parameter tuning.

Furthermore, the study highlights the influence of workload characteristics on algorithm performance, emphasizing the importance of adaptability in scheduling strategies. Insights derived from sensitivity analysis aid in informed decision-making for system designers and administrators, guiding the selection of the most suitable scheduling approach to meet specific performance objectives.

Looking ahead, future research directions may explore hybrid or adaptive scheduling methodologies to further enhance system responsiveness and resource utilization. By continuing to refine scheduling algorithms and adapting them to evolving computing environments, we can unlock new avenues for optimizing system performance in the dynamic landscape of modern operating systems.

REFERENCES

[1]Galvin, P. B., Gagne, G., & Silberschatz, A. (2003). Operating system concepts. John Wiley & Sons.

[2]Abbas, A. (2015). Multiprocessor and Real-Time Scheduling Shortest-Job-First (SJF) Scheduling Algorithm. European Journal of Computer Science and Information Technology, 3(5), 8–20.

[3]Kamalam, G. K., & Bhaskaran, D. V. M. (2014). Resource Fitness Task Scheduling Algorithm for Scheduling Tasks on Heterogeneous Grid Environment. Australian Journal

of Basic and Applied Sciences, 8(18), 128-135.

[4]Siahaan, A. P. U. (2016). Comparison analysis of CPU scheduling: FCFS, SJF and Round Robin. International Journal of Engineering Development and Research, 4(3), 124-132.

[5]M Mostafa, S., & Amano, H. (2020). Dynamic Round Robin CPU Scheduling Algorithm Based on K-Means Clustering Technique. Applied Sciences, 10(15), 5134.

[6]Farooq, M. U., Shakoor, A., & Siddique, A. B. (2017, March). An efficient dynamic round robin algorithm for cpu scheduling. In 2017 International Conference on Communication, Computing and Digital Systems (C- CODE) (pp. 244-248). IEEE.

[7]Zouaoui, S., Boussaid, L., & Mtibaa, A. (2019). Priority based round robin (PBRR) CPU scheduling algorithm. International Journal of Electrical & Computer Engineering (2088-8708), 9(1).Technologies and Optimization (Trends and Future Directions)(ICRITO) (pp. 397- 400). IEEE.

[8]Gupta, A. K., Yadav, N. S., & Goyal, D. (2016). Design and Performance Evaluation of Smart Job First Dynamic Round Robin (SJFDRR) Scheduling Algorithm with Smart Time Quantum. American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS), 26(4), 66-78.

[9]Manuel, J. I., Baquirin, R. B., Guevara, K. S., & Tandingan, D. (2019, February). Fittest Job First Dynamic Round Robin (FJFDRR) scheduling algorithm using dual queue and arrival time factor: a comparison. In IOP Publishing Ltd, IOP Conf. Ser.: Mater. Sci. Eng (Vol. 482, p. 012046).

[10] Khatri, J. (2016). An enhanced Round Robin CPU scheduling algorithm. IOSR Journal of Computer Engineering (IOSR-JCE), 18(4), 20-24.

[11] Srilatha, N., Sravani, M., & Divya, Y. (2017). Optimal Round Robin CPU Scheduling Algorithm Using Manhattan Distance. International Journal of Electrical and Computer Engineering, 7(6), 3664

[12] Ali, K. F., Marikal, A., & Kumar, K. A. A Hybrid Round Robin Scheduling Mechanism for Process Management. International Journal of Computer Applications, 975, 8887.

[13] Pratap, R., & Zaidi, T. (2018, August). Comparative study of task scheduling algorithms through Cloudsim. In 2018 7th International Conference on Reliability, Infocom

[14] Srujana, R., Roopa, Y. M., & Mohan, M. D. S. K. (2019, April). Sorted Round Robin

Algorithm. In 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 968-971). IEEE.

[15] Biswas, D., & Samsuddoha, M. (2019). Determining Proficient Time Quantum to Improve the Performance of Round Robin Scheduling Algorithm. International Journal of Modern Education and Computer Science, 11(10), 33.

[16] Shafi, U., Shah, M. A., Wahid, A., Abbasi, K., Javaid, Q., Asghar, M., & Haider, M. (2020). A novel amended dynamic round robin scheduling algorithm for timeshared systems. Int. Arab J. Inf. Technol., 17(1), 90-98.

[17] Aijaz, M., Tariq, R., Ghori, M., Rizvi, S. W., & Qazi, E.F. (2019, March). Efficient Round Robin Algorithm (ERRA) using the Average Burst Time. In 2019 International Conference on Information Science and Communication Technology (ICISCT) (pp. 1-5). IEEE.

[18] Mody, S., & Mirkar, S. (2019, December). Smart Round Robin CPU Scheduling Algorithm For Operating Systems. In 2019 4th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT) (pp. 309-316). IEEE.

[19] Parekh, H. B., & Chaudhari, S. (2016, December). Improved Round Robin CPU scheduling algorithm: Round Robin, Shortest Job First and priority algorithm coupled to increase throughput and decrease waiting time and turnaround time. In 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC) (pp. 184-187). IEEE.

[20] Alworafi, M. A., Dhari, A., Al-Hashmi, A. A., & Darem, A. B. (2016, December). An improved SJF scheduling algorithm in cloud computing environment. In 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT) (pp. 208-212). IEEE.

[21] Joshi, A., & Gosswami, S. (2017). Modified Round Robin algorithm by using Priority Scheduling. Advances in Computational Sciences and technology, 10(6), 1543-1549.

[22] Zouaoui, S., Boussaid, L., & Mtibaa, A. (2019). Priority based round robin (PBRR) CPU scheduling algorithm. International Journal of Electrical & Computer Engineering (2088-8708), 9(1).

[23] Berhanu, Y., Alemu, A., & Mishra, M. K. (2017). Dynamic time quantum based round robin CPU scheduling algorithm (Doctoral dissertation).

"CREDIT CARD FRAUD ANALYSIS AND DETECTION METHOD"

Ms. Aksa Biju

BCA Student

PG Department of CA & AI

Saintgits College Of Applied Sciences Kottayam, Kerala

Mahatma Gandhi University - Kerala

aksabiju003@gmail.com

Ms. Clarin Joe

BCA Student

PG Department of CA & AI

Saintgits College Of Applied Sciences Kottayam, Kerala

Mahatma Gandhi University - Kerala

clarinamalajoe@gmail.com

Mr. P M Wajid Rehman

BCA Student

PG Department of CA & AI

Saintgits College Of Applied Sciences Kottayam, Kerala

Mahatma Gandhi University - Kerala

pmwr.bca2124@saintgits.org

Mr. Aswin A

BCA Student

PG Department of CA & AI

Saintgits College Of Applied Sciences Kottayam, Kerala

Mahatma Gandhi University - Kerala

aswin.bca2124@saintgits.org

Dr. Meenu Suresh

Assistant Professor

PG Department of CA & AI

Saintgits College Of Applied Sciences Kottayam, Kerala

Mahatma Gandhi University - Kerala

meenupillai1988@gmail.com

Abstract

Electronic payment (e-payment) has transformed financial transactions, offering speed and convenience. However, it has also brought about significant challenges, notably in credit card security. This paper explores the landscape of e-payment, focusing on credit card fraud detection—a crucial aspect of financial security systems. Advanced algorithms and machine learning techniques are employed to analyse transaction data for patterns indicative of fraudulent activity. Despite these advancements, credit cards face various security threats, including card skimming, phishing scams, and data breaches. Cybercriminals continuously adapt their tactics, necessitating ongoing advancements in credit card protection. This paper highlights the importance of evolving security measures to safeguard users' financial information in the ever-changing digital landscape.

I.Introduction

Electronic payment, or e-payment, refers to the digital exchange of money between parties for goods or services, conducted through electronic devices and online platforms. This method of financial transaction eliminates the need for physical currency or traditional paper-based checks, offering a convenient and efficient means of conducting business in the modern digital era[2]. E-payment systems encompass various forms, including credit/debit card transactions, mobile payments, online banking transfers, and digital wallets[4,7,8]. The widespread adoption of e-payment solutions has revolutionised the way individuals and businesses manage financial transactions, offering speed, security, and accessibility.

Credit card fraud detection is a critical aspect of financial security systems aimed at identifying and preventing unauthorised or fraudulent transactions. Advanced algorithms and machine learning techniques play a pivotal role in analysing vast amounts of transaction data to identify patterns and anomalies indicative of fraudulent activity[9]. These systems assess various factors, including transaction frequency, location, purchase amounts, and unusual spending patterns.

Credit cards face various security and fraud challenges, ranging from unauthorised transactions and identity theft to sophisticated cyberattacks[10-12]. One prevalent issue is card skimming, where criminals install devices on ATMs or point-of-sale terminals to capture card information[13]. Phishing scams also target cardholders through deceptive emails or websites, aiming to obtain sensitive details. Additionally, data breaches at retailers or financial institutions can compromise large volumes of credit card information. Cybercriminals constantly evolve their tactics, employing malware and other advanced techniques to exploit vulnerabilities in online transactions[14,15]. The ongoing battle between security measures and fraudulent activities underscores the need for continuous advancements in credit card protection to safeguard users' financial information.

II.Related Work

Sadgali et al[6] focuses on addressing the increasing challenge of credit card fraud in the digital age. The authors use a consistent dataset to evaluate different methods, aiming to select the best technique for implementation in future work. The paper underscores the potential of machine learning in enhancing the accuracy and efficiency of fraud detection systems, thereby providing a significant contribution to the security of financial transactions in the digital era. Zhang et.al [5] explores the application of the Xgboost model for detecting fraud in customer transactions. The study uses a dataset from the IEEE-CIS Fraud Detection Competition on Kaggle, involving data mining techniques like feature engineering, visualization, and the use of SMOTE (Synthetic Minority Oversampling Technique) for addressing class imbalance. The authors compare Xgboost with other machine learning methods like Support Vector Machine, Random Forest, and Logistic Regression, demonstrating that the Xgboost-based model outperforms these in terms of ROC AUC score and accuracy. They also highlight the importance of feature selection in improving model performance. Liu et.al [16] presents a comprehensive study on the application of machine learning techniques in detecting online transaction fraud. The paper introduces two fraud detection algorithms based on Fully Connected Neural Networks and XGBoost, respectively, and highlights the design of an interactive online transaction fraud detection system that utilizes the XGBoost model. The study includes extensive experiments and comparisons and also discusses the system's capacity to analyze transaction data automatically and provide users with fraud detection results, contributing significantly to the field of online financial security. Sailusha et.al [1] explores the application of machine learning algorithms for detecting credit card fraud. The study focuses on using Random Forest and Adaboost algorithms to analyze credit card transaction data. The effectiveness of these algorithms is evaluated based on accuracy, precision, recall, and F1-score. The researchers also employ a Receiver Operating Characteristic (ROC) curve based on the confusion matrix for further analysis. The results show that while both algorithms perform well, the Random Forest algorithm exhibits a higher performance. GPT

The author Jain et.al [3] examines the effectiveness of various machine learning algorithms in detecting credit card fraud. The study specifically focuses on three algorithms: Decision Tree, Random Forest, and XGBoost. It compares their performance in terms of prediction accuracy and uses a dataset of over one lakh credit card transactions for testing. The results reveal that XGBoost has the highest prediction accuracy (99.962%), followed by Random Forest (99.957%) and Decision Tree (99.923%).

III.Proposed Method

The research methodology comprises three main sections: data pre-processing, addressing imbalanced classifiers, and providing descriptions of the models used in the study.

3.1 Dataset Description

The dataset utilized in this study to evaluate students' adaptability to online learning was sourced from Kaggle, a well-known machine learning repository widely used for sharing and assessing datasets. Kaggle serves as a platform where individuals, organizations, and scholars contribute datasets spanning various sectors. Ensuring alignment with the goals and parameters of our investigation was crucial in selecting an appropriate dataset from Kaggle's diverse offerings. Researchers conducted a thorough evaluation of the dataset to ascertain its dependability and quality. This evaluation included a meticulous assessment for completeness, correctness, and relevance to ensure that the dataset adequately met the requirements of our study.

A. Data Preprocessing

In this research, the data underwent thorough cleaning, transformation, and organization to meet quality standards for analysis. Subsequently, normalization was performed, applying rules to categorise data into low, moderate, and high adaptability levels. The Interquartile Range [71,72] was utilized to eliminate any inaccurate data points, ensuring data accuracy.

B. Balancing the Dataset using SMOTE

Addressing the inherent data imbalance, the dataset comprises 625 instances of Moderate values, 480 instances of Low values, and 100 instances of High values. To rectify this imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) [73,74] is employed through oversampling and is depicted in Fig.1. This technique involves generating synthetic instances to augment the minority class, ensuring a more balanced representation in the dataset. Subsequently, the balanced dataset is fed into each classifier for effective classification.

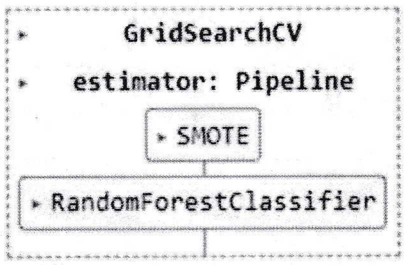


Figure 1. Balancing the Dataset using SMOTE

C. Description of Models

Numerous machine learning algorithms have been employed to forecast potential outcomes regarding student adaptability levels in this study. The dataset underwent training and testing with a diverse set of ML algorithm - RandomForest. Given the dataset's multi-labeled nature, a One Vs. Rest approach was integrated into binary class classification algorithms to tackle the challenges of multi-class classification. In this methodology, one class is designated as the positive class, while all other classes are collectively treated as the negative class. Each class undergoes individual testing, and results are generated by evaluating one class against all the other classes using this approach.

1) Random Forest

The Random Forest Classifier is a machine learning method specifically designed for addressing classification problems. This ensemble learning technique combines multiple decision trees to generate predictions. The key feature of a random forest lies in training each tree with a different subset of the training data and features, resulting in a diverse set of decision trees. As incoming data is fed into the ensemble, each tree independently categorises it, and the final prediction is determined by a majority vote among the trees. Renowned for its ability to handle high-dimensional data, mitigate overfitting, and deliver robust and accurate classification results, the Random Forest Classifier finds widespread application in various industries such as banking, healthcare, and image categorization, thanks to its efficiency and adaptability.

$$_{norm} fi_i = \frac{fi_i}{\sum j \in features fi_i}$$

IV.Evaluation and Result Analysis

As shown in Table 1, Random forest with recall, precision, F1 Score and accuracy with SMOTE oversampling are 0.765957,0.186528,0.300000,0.994079 respectively and with no oversampling are 0.723404,0.918919,0.809524,0.999436 is depicted in figure 2.

Table 1: Comparison of the various metrics

Random Forest with	Recall	Precision	F1 Score	Accuracy
SMOTE Oversampling	0.765957	0.186528	0.300000	0.994079
No under/oversampling	0.723404	0.918919	0.809524	0.999436

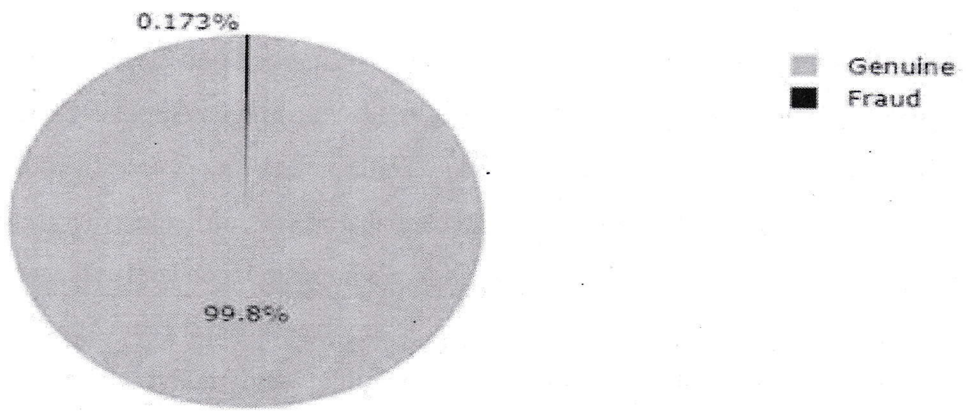


Figure 2. Fraud and Genuine transaction

V.Conclusion

Even though there are many fraud detection techniques we can't say that this particular algorithm detects the fraud completely. From our analysis, we can conclude that the accuracy for the Random Forest with SMOTE Oversampling is 99.4079 and Random Forest with No Under/Oversampling is 99.9436. When we consider the precision, recall, and the F1-score the Random Forest algorithm has the highest value . Hence we conclude that the Random Forest Algorithm works with best accuracy to detect credit card fraud.

REFERENCES

1) Ruttala Sailusha,V. Gnaneswar,R. Ramesh,G. Ramakoteswara Rao,"Credit Card Fraud Detection Using Machine Learning",Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2020).
2) Kanika,Dr Jimmy Singla,"A Survey of Deep Learning based Online Transactions Fraud Detection Systems",2020 International Conference on Intelligent Engineering and Management (ICIEM).
3) Vinod Jain,Mayank Agrawal,Anuj Kumar,"Performance Analysis of Machine Learning Algorithms in Credit Cards Fraud Detection",2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO).
4) Parth Roy,Prateek Rao,Jay Gajre,Kanchan Katake,Arvind Jagtap,Yogesh Gajmal,"Comprehensive Analysis for Fraud Detection of Credit Card through Machine Learning",2021 International Conference on Emerging Smart Computing and Informatics (ESCI).
5) Yixuan Zhang,Jialiang Tong,Ziyi Wang,Fengqiang Gao,"Customer Transaction Fraud Detection Using Xgboost Model",2020 International Conference on Computer Engineering and Application (ICCEA).
6) Imane SADGALI,Nawal SAEL,Fouzia BENABBOU,"Fraud detection in credit card transaction using machine learning techniques",2019 978-1-7281-4368-2/19/$31.00 ©2019 IEEE.
7) V.Kamesh, M.Karthick,K.Kavin,M.Velusamy, R.Vidhya,"Real-Time Fraud Anamaly Detection in E-banking Using Data Mining Algorithm",2017 South Asian Journel of engineering and technology
8) I. Sadgali, N. Sael, F. Benabbou, "Detection of credit card fraud:State of art",International Journal of computer science and network security, Vol.18, No.11, pp.76-83, 2018.
9) S. Askari, A. Hussain, "Credit Card Fraud Detection Using Fuzzy ID3", Proc. of International Conf.On Computing, Communication and Automation (ICCCA), Greater Noida, India, pp.446-452, 2017.
10) V. Sapovadia, "Financial Inclusion, Digital Currency, and Mobile Technology," in Handbook of Blockchain, Digital Finance, and Inclusion, Volume 2, Elsevier, 2018, pp. 361–385.

11) Arvind M Jagtap, Prof.Dr.Gomathi N, "Meta-Heuristic based Trained Deep Convolutional Neural Network for Crop Classification" ,International Journal of Emerging Trends in Engineering Research (IJETER) Volume 8. No. 7, July 2020.

12) ULB (2018), Kaggle, "Machine Learning Group-Credit Card Fraud Detection"

13) Minastireanu, E. A., & Mesnita, G. (2019). Light gbm machine learning algorithm to online click fraud detection. J. Inform. Assur. Cybersecur, 2019.

14) Fabrizio Carcillo, Yann-Aël Le Borgne, Olivier Caelen, Yacine Kessaci, Frédéric Oblé, Gianluca Bontempi, Combining unsupervised and supervised learning in credit card fraud detection, Information Sciences, 2019, ISSN 0020-0255.

15) Ugo Fiore, Alfredo De Santis, Francesca Perla, Paolo Zanetti, Francesco Palmieri, Using generative adversarial networks for improving classification effectiveness in credit card fraud detection, Information Sciences, Volume 479, 2019, Pages 448-455, ISSN 0020-0255.

16) ARMEL and D. ZAIDOUNI, "Fraud Detection Using Apache Spark," 2019 5th International Conference on Optimization and Applications (ICOA), Kenitra, Morocco, 2019, pp. 1-6. doi: 10.1109/ICOA.2019.8727610.

17) Bocheng Liu, Xiang Chen* and Kaizhi Yu ,"Online Transaction Fraud Detection System Based on Machine Learning",ICCTPE 2021.

DNA COMPUTING AND ITS APPLICATIONS: SURVEY

Dr. Ambily Merlin Kuruvilla

Associate Professor

PG Department of Computer Applications and Artificial Intelligence

Saintgits College of Applied Sciences.

E Mail: Ambilykuruvilla@gmail.com

Phone No: 9447661714

Aiswarya Ajukumar

Final year BCA student

PG Department of Computer Applications & AI

Saintgits College of Applied Sciences.

E Mail :aiswaryaajukumar6@gmail.com

Phone No: 9778513593

Akshara Sunilkumar

Final year BCA student

PG Department of Computer Applications & AI

Saintgits College of Applied Sciences.

E Mail: aksharasunilkumar87@gmail.com

Phone No:6282805998

Anagha Nair

Final year BCA student

PG Department of Computer Applications & AI

Saintgits College of Applied Sciences.

E Mail: anaghanairsreenivas@gmail.com

Phone No: 7306715101

Kalailekshmi M

Final year BCA student

PG Department of Computer Applications & AI

Saintgits College of Applied Sciences.

E Mail: kalailakshmim123@gmail.com

Phone No:7034464072

ABSTRACT

This research project focuses on DNA computing using Strategies, a widely accepted technology for computer advancement. It simply reflects a range of ideas within the discipline.. This presentation will focus on the different duties and modes in the rapidly expanding technical sector that this invention has made simpler. This will encompass the workings of the method, the different ways it could potentially be utilized in practice, and future possibilities in the relevant area. This study describes the whole colorful process of doing the work at the molecular level. This encompasses all the characteristics and uses relevant to this degree, resulting in a new age in computer systems. DNA computing presents enormous opportunities for solving difficult computational problems because of its ground-breaking incorporation of biology and computer technology. DNA computing primarily makes use of DNA molecules' exceptional information processing and storage capabilities.. The technology consists of embedding data inside DNA sequences and using transcription and DNA replication, two biological processes, to do computations in parallel.. The capacity of DNA computing to do massively simultaneous tasks has allowed for the solution of many difficult or difficult problems that would be inconceivable to tackle with traditional computational techniques. This is one of the most interesting features of DNA computing. In addition, the intrinsic scalability and energy efficiency of DNA-based systems suggest promise for fulfilling increasing computational demands in an era marked by the proliferation of data and resource constraints.

Key words: DNA, HPP: Deoxyribonucleic acid, Hamiltonian Path Problem.

INTRODUCTION

Considering the importance of DNA molecules in computation, DNA computing is an interdisciplinary field of study that is growing rapidly. The creation of a biologically inspired DNA computer that can one day replace silicon-based computers or at the very least operate well in conjunction with them is one of the main objectives of this branch of study. since R. Feynman's 1964 proposal for building a computer from molecules. It took twenty years to finish Adleman's proof of principle in 1994 that DNA molecules may utilize a biological process to solve an NP problem of the Hamiltonian Path Problem (HPP). DNA was the main information storing material used by all living organisms.

The primary function of DNA is to transfer and store life's knowledge throughout millions of years. Approximately 10 trillion DNA molecules may be contained in an area the size of a marble. It is theoretically possible to calculate 10 trillion times in a tiny area at once because all of these molecules have simultaneous data processing capabilities. Molecular computing is the term used more frequently to refer to DNA computing. This is a multifaceted subject that includes mathematics, biology, chemistry, and computer science. Computing with DNA offers a completely new paradigm for computation. Encoding data in the form of DNA strands, which are subsequently modified in an incubator using molecular biology methods known as "bio-operations" to mimic arithmetical and logical operations, is the basic notion behind DNA computing. It is thought that 1018 DNA strands together can function 104 times faster than the most powerful supercomputer on sale today.

After then, DNA computing has become the subject of significant cross-disciplinary study. Two elementary domains of research in DNA computing were distinguished by Rozenberg et al. (1999): (i) the theoretical domain, which deals with models, algorithms, and paradigms for DNA computing; and (ii) the experimental domain, which plans lab tests to assess the feasibility of biochemical processes. While more work remains to be done in order to adapt the DNA algorithm to practical problems, Scientists are eager to model and test the solution in a case study, nevertheless, in order to disprove DNA's fundamental limitation. These days, a lot of active research groups are digging into this topic, creating models and running lab tests, particularly when it comes to the barriers to biochemical survival. In an effort to conquer technical or application-related obstacles, other teams are developing DNA computers and algorithms.

The paper is structured as follows generally. Section 1 provides a summary of the research topic. In this section, we give a quick definition of DNA computing. Section 2 will discuss DNA computation techniques used in molecular biology for DNA processing in order to improve knowledge of a DNA computing methodology.

1. The DNA Molecule

Deoxyribonucleic acid is often referred to as DNA. DNA molecules are polymers made of nucleotide monomers. To the reasons we have, these have a very basic structure that encompasses three separate components: phosphate, sugar, and base. The four different bases are identified as adenine, guanine, cytosine, and thymine, or A, G, C, and T, respectively. Since the only thing that differentiates nucleotides is their base, we can identify them by introducing "G nucleotides" or similar base abbreviations.

Simply put, a single strand DNA fragments contain chains of nucleotide that possess a sugar-phosphate "backbone" that establishes a potent ionic link that connects two recurrent nucleotides. Chemical convention asserts that every single strand has a 5 and a 3 end, meaning that every individual thread has an original configuration. The Watson–Crick alignment of sequences is the most fundamental aspect of DNA.

When distinct strands are joined together, bases are pulled to one other in combinations; for example, A bonds with T and G joins with C. Consequently, it is common to refer to the pairs (A; T) and (G; C) as complimentary base pairs (Fig.2).DNA develops its typical double helix when the two separate strands join together (Fig.1). Two hydrogen bonds are created between the two base pairings, two among A and T, and three between G and C. In order for an idea to materialise, the subsequent conditions must be satisfied: In the first stage, the threads must be complimentary, and in the next phase, they need to possess different polarity.

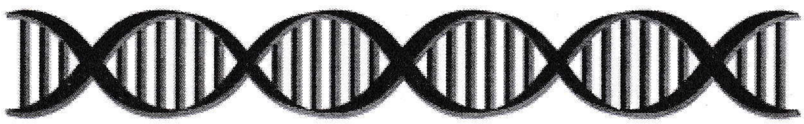


Fig.1 Double Helical DNA's structure

2. DNA Operations

Each DNA computational framework delivers a particular set of biological reactions on an accumulation of strands. Cellular biologists use all those techniques often. Keep in mind that certain DNA computation models require computations.

2.1. Integration

A device roughly that's the equivalent of a microwave oven is capable of synthesising oligonucleotides on demand. The four nucleotide bases in solution are given to the synthesiser, and they are combined with each other in accordance with a sequence which the user selects. Millions of copies of the appropriate oligonucleotide can be generated by the tool and then added to solution in an extremely small vial.

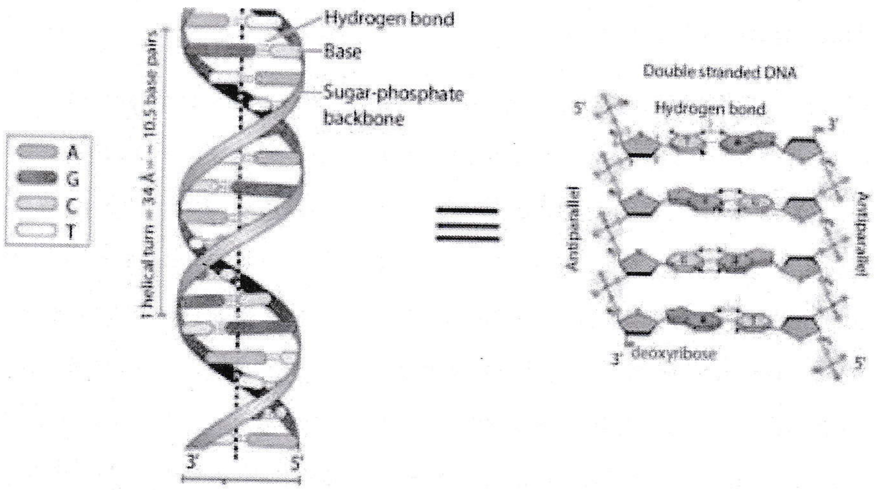


Fig.2 Detailed picture of DNA Strand

2.2. Annealing, ligation, and denaturing

When the double-stranded DNA is subjected to a temperature that depending on the strand's make-up, it may break down into single strands, a phenomenon known as denatured DNA. The hydrogen bonds that are dividing complementary strands are ruptured by heating (Fig.3). Considering the strands are unaffected by this process because the hydrogen bonds that hold strands together are far weaker than the covalent connections that hold strands together. Since the connecting link of a G–C pair which is joined by three hydrogen bonds. It is crucial to consider this factor while constructing sequences that represent computational units. Annealing is the counterpart of melting; it requires cooling a solution of individual strands so that complementary strands can stick to one another (Fig.3).

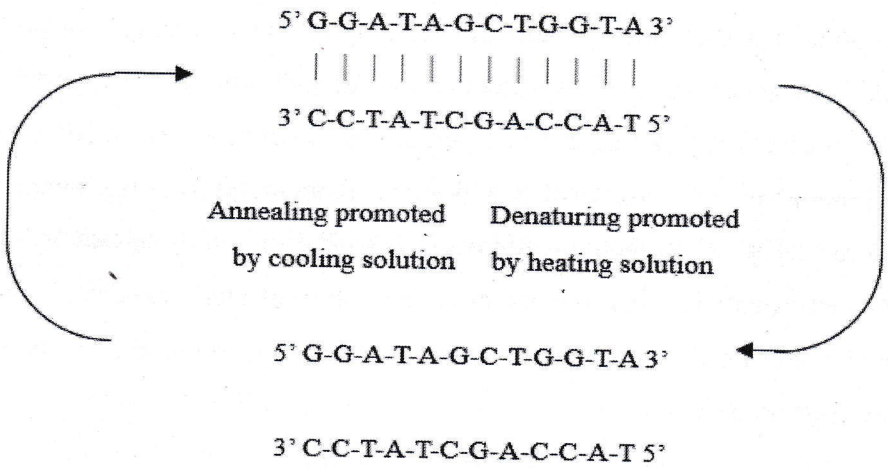


Fig.3 Annealing and Denaturing

2.3. The division of hybridization

In DNA computation, separation by hybridization is an ordinary method that involves removing all single strands carrying a certain short sequence (e.g., extract all strands bearing the sequence TAGACT) from a test tube. We must first make multiple copies of the sequence's counterpart in order to extract individual strands bearing the sequence x. We affix a biotin molecule 1 to these oligonucleotides, and it eventually binds to a fixed matrix. Strands with x will become annealed to the anchored complementary strands if the contents of the test tube are poured over this matrix. Only the strands carrying x remain after washing the matrix to eradicate any strands that did not anneal. The matrix may then be cleared of them.

2.4. Gel Electrophoresis

One major technique for distinguishing DNA strands by size is gel electrophoresis. The motion of molecules that are charged in an electrical field is known as electrophoresis. DNA molecules gravitate towards the positive pole in an electrical field because they are charged negatively. The shape and electric charge of a molecule determine how rapidly it migrates in an aqueous solution. The molecules of DNA flow in the same direction in an aqueous solution because they all have identical charges per unit length. However, the size of a molecule may influence the speed at which it travels if electrophoresis is performed on a gel, which is often made of agarose, polyacrylamide, or a combination of the two.

2.5. PCR and primer extension

The DNA polymerases perform a variety of tasks, including duplicating and repairing DNA. When nucleotide triphosphates are present, the polymerase will only expand the primer oligonucleotide p (always in the $5 \rightarrow 3\}$ direction) if and only if it is joined to the larger template oligonucleotide, t. A frequent obstacle in DNA computation is reading out the final answer to a problem encoded in a DNA strand, since laboratory procedures can result in a very diluted solution. The PCR method addresses this "needle in a haystack" problem by substantially (exponentially) multiplying any desirable molecules existing in the solution, causing the volume of the solution to "visibly" grow. Consequently, the problems with detection are addressed.

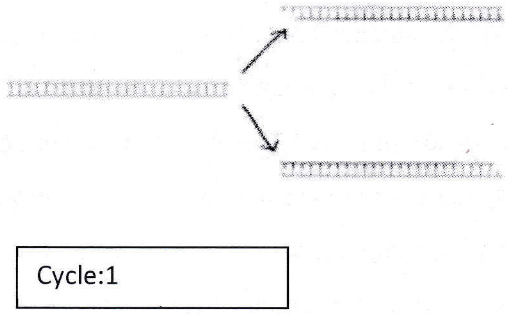


Fig 4: PCR Cycle

2.6 Restriction Enzymes

Often termed restriction enzymes, restriction endonucleases are sequences of DNA that are known as restriction sites. At that point, the enzyme cuts any double-stranded DNA containing the restriction site beneath its structure. Researchers were able to separate gene-containing fragments and recombine them with other molecules of DNA, or clone genes, because the enzymes could precisely cut DNA at certain spots. Since the names of the enzymes originate from the genus, species, and variant identities of the bacteria that create them, Escherichia coli strain RY13 is the producer of the restriction enzyme EcoRI. In concept, restriction enzymes—which have the capability of determining sequences—evolved through an identifiable origin protein through processes which includes genetic recombination and gene amplification.

3. **Literature Survey**

- Through his 1994 experiment, "Adleman" established the constant length based methodology for illustrating the distances between two cities. However, Adleman failed to identify the arcs in this experiment that indicated the distances between the cities. Although Adleman's model was updated in 1995, no data concerning distances between cities at the time was added. The first models that manage data or labels on arcs were invented by Narayanan and Zorbalas in 1998 during tackling a weighted graph problem.
- In addition to communicate information about arcs, "Narayanan and Zorbalas" suggested that use constant-based length when expressing the distances between cities. By using this method, the 3-mer of DNA at distance 1 and the corresponding data at distance 2 as 6-mer DNA, for example, will be presented. Consequently, shorter DNA strands will present a shorter distance and longer DNA strands a longer one. At the conclusion of the process, the optimal resolution to the problem will be presented by the shortest strand.

- "Ibrahim et al." have proposed an inventive method to overcome the constant-proportional length-based disadvantage: the direct-proportional length-based strategy. The cost of an edge is encoded using this method using a direct-proportional length oligos.
- After the first pool generation and amplification, a number of viable candidates are generated. Using standard biomolecular laboratory techniques, the ideal combination indicating a solution to the shortest path problem can be extracted.
- "Yamamoto et al." suggested utilize concentration control in 2000 to resolve the weighted graph problem. DNA concentrations serve as input and offer data in this technique since they are used to influence chemical reactions. Yamamoto et al. conjectured that this strategy can reduce the experiment operating expenditures in the DNA computing detection procedure since it only requires the extraction and analysis of bands with a reasonable intensity. This method sets the concentrations of complementary oligonucleotide encoding vertices to the same values as given, then computes the relative concentration Dij of each oligonucleotide encoding edge i → j with cost Cij.
- "Lee et al." presented a novel encoding method in 2004 that uses temperature gradients to address weighted graph problems. Costs are expressed using the fixed-length DNA strand melting temperature approach. More affordable solutions were produced by lower melting temperatures, which also led to cheaper costs. Every city sequence contributed equally to path stability because the melting temperatures were constant. Road sequences connecting the cities were constructed based on expenses and the cities of departure and arrival. The never-ending pursuit of the best solution demonstrates how technical and application difficulties can be addressed in this discipline.
- In 2021, homomorphic encryption techniques are used in a system architecture presented by Michael Johnson and Emily Davis. Homomorphic encryption maintains privacy by enabling computations[11] to be done directly on encrypted data. Verifiable computation techniques are employed by the system to provide data integrity tests without disclosing the actual data. Cryptographic techniques and zero-knowledge proofs are used to provide safe data retrieval and verification.
- David Lee and Jessica Thompson offer a system design in 2021 that combines cloud data storage and ABE. Data is encrypted using ABE, and access policies are linked to the encrypted data. Users are given traits, and those that match are used to grant access to the encrypted data. To control encryption keys and implement access controls, the system makes use of a key management architecture.

- Sarah Johnson provides an extensive analysis of the literature and studies on DNA-based data storage in 2022. It examines the state of the art at the moment, covering readout technologies, encoding schemes, synthesis approaches, and error correction tactics. The limits and possible uses of DNA-based data storage are also covered in the article.

Author	Year	Technique used	Distinguishable Features studied
Adleman	1994	length-based paradigm	Showing the separations between two cities
Narayanan ,Zorbalas	1998	extended based length	The optimal way to determine the distances between two cities will be represented by the strand that is the shortest at the conclusion of the process.
Yamamoto et al	2000	concentration control	Bands with a reasonable intensity can reduce the costs associated with running the experiment for the DNA computing detection process.
Lee et al	2004	unique encoding technique	City sequences had consistent melting temperatures, each contributed equally to path stability
Michael Johnson , Emily Davis	2021	Homomorphic encryption technique	Homomorphic encryption allows

			computations to be performed on encrypted data directly, preserving privacy
David Lee, Jessica Thompson	2021	integrates ABE with cloud data storage	The system utilizes a key management infrastructure to manage encryption keys and enforce access policies
Sarah Johnson	2022	Encoding techniques, synthesis methods, readout technologies, and error correction strategies.	Discusses potential applications and limitations of DNA-based data storage.

CONCLUSION

The study focus on the most recent advances in the field of DNA computing research. We also looked over some recent techniques and algorithms recently developed in the field of DNA computing. These techniques have been used recently by DNA computing researchers to address industrial and computational challenges. Nonetheless, some academics are focused on creating innovative methods in wet technology. DNA technology have advanced dramatically during the past ten years, both theoretically and practically speaking.

In conclusion, one of the most recent and fascinating fields for research is DNA computing. Numerous potential exist for extending and modifying DNA functions and properties to address practical problems, particularly in industrial engineering and issues with management engineering. Given all of the benefits that DNA computing offers, it ought to emerge as a viable solution to the problems that the silicon computer of today faces. Nevertheless, there are still several barriers to applying this approach to engineering challenges as of right now.

REFERENCE

1) Leonard M. Adleman," Computing with DNA"ScientificAmerican,August 1998.
2) Tim Fyer, (2018) " Life After Silicon: Published in: Engineering & Technology.
3) "Selection of genes mediating certain cancers, using a Neuro fuzzy approach", Anupam ghosh, Bibhas Chandra dhara and Rajat k.ve, Elesevier, Science Direct, Neurocomputing 133 (2014)
4) "Neural network and SVM Classifiers accurately predict lipid binding proteins, irrespective of sequence homology", Mehdi Khashei, Ali Zeinal Hamadani, Mehdi Bijari, Elsevier, Science Direct, Journal of Theoretical Biology (2014)
5) "Support vector machines for quality control of DNA sequencing", Ersoy Oz and Huseyin Kaya, Journal of Inequalities and Application 2013 (a Springer Open journal)
6) "Michael Johnson, Emily Davis. (2021). Efficient and Verifiable Cloud Data Storage with Homomorphic Encryption"
7) P. Sharma, and M. Saini. (2021). A Review on DNA Computing: Its Application and Security Issues. International Journal of Advanced Research in Computer Science, 12(3), pp. 358-363.
8) Sarah Johnson. (2022). DNA-Based Data Storage: Challenges and Opportunities
9) David Lee, Jessica Thompson. (2021). Secure Cloud Data Storage System with Attribute-based Encryption for Fine-grained Access Control.
10) E. T. Ordentlich, et al. (2020). DNA storage for archiving data in the cloud. Proceedings of the 2020 IEEE International Conference on Cloud Computing Technology and Science, pp. 392-396.
11) Ambily Merlin Kuruvilla,2021, Improved Artificial Neural Network Through Metaheuristic Methods And Rough Set Theory For Modern Medical Diagnosis, Indian Journal of Computer Science and Engineering.

EXPLORING THE FUTURE ENHANCEMENT OF HUMAN LIFE THROUGH IOT: A COMPREHENSIVE REVIEW

Mr. DENNY DAVID

Assistant Professor, Department of Computer Science and Applications, Naipunnya School of Management, Cherthala, Kerala

dennydavid2008@gmail.com

&

Mr. JOEL JOSEPH

Assistant Professor, Department of Computer Science and Applications, Naipunnya School of Management, Cherthala, Kerala

joeljoseph882@gmail.com

Abstract

The Internet of Things (IoT) has emerged as a transformative technology with the potential to revolutionize various aspects of human life. This research paper aims to investigate the role of IoT as a future enhancement for humans. By examining existing literature, case studies, and technological advancements, this study provides insights into the applications, benefits, challenges, and ethical considerations associated with IoT deployment. Additionally, it explores the potential impact of IoT on different domains such as healthcare, smart cities, agriculture, and industrial automation. Through a multidisciplinary approach, this paper seeks to contribute to the understanding of how IoT can positively influence human life and shape the future of society.

Keywords: Internet of Things (IoT), future enhancement, human life, applications, benefits, challenges, ethical considerations, healthcare, smart cities, agriculture, industrial automation.

Introduction

The background and significance of the Internet of Things (IoT) lie in its transformative potential to connect physical devices, sensors, and systems, enabling them to communicate and exchange data seamlessly over the internet. This technological paradigm shift has profound implications for various industries and societal domains, making it a significant area of research and development. The rapid evolution of computing, communication, and sensor technologies has paved the way for the proliferation of IoT devices. Smaller, more powerful, and energy-efficient sensors, coupled with advances in wireless connectivity and cloud computing, have made it feasible to deploy IoT solutions across diverse environments. IoT involves establishing communication links between physical objects or "things" and the internet or other devices. This connectivity enables real-time data exchange and remote monitoring, facilitating seamless interaction between the physical and digital worlds. IoT devices collect vast amounts of data from their surroundings, including environmental conditions, user interactions, and system performance metrics. This data is then processed, analyzed, and interpreted to derive meaningful insights, optimize operations, and support decision-making. IoT enables automation and control of processes, systems, and devices through remote monitoring and management. This includes tasks such as remotely controlling smart home appliances, monitoring industrial equipment for predictive maintenance, and optimizing energy usage in smart buildings.

Literature Review

The evolution of IoT (Internet of Things) technologies has been a subject of extensive research and analysis in academic literature over the past two decades. Researchers have explored various aspects of IoT technologies, including their development, applications, challenges, and future trends.

Early literature often focuses on defining IoT and conceptualizing its potential applications. For example, the term "Internet of Things" was first coined by Kevin Ashton in 1999, and subsequent research aimed to establish a common understanding of IoT as a network of interconnected devices capable of sensing, communicating, and sharing data. Research has explored the technological foundations of IoT, including wireless sensor networks (WSNs), RFID (Radio Frequency Identification), and various communication protocols (e.g., Zigbee, Bluetooth, Wi-Fi). Studies have investigated the capabilities and limitations of these technologies in enabling IoT applications. Literature has discussed different IoT architectures and standards to facilitate interoperability and scalability. For example, the reference architecture proposed by the International Telecommunication Union (ITU-T) provides a framework for designing and implementing IoT systems, while standards bodies like the IEEE and IETF have developed protocols and specifications for IoT communication and data exchange. Research has explored a wide range of IoT applications across various domains, including healthcare, smart cities, agriculture, industrial automation, and smart homes. Case studies and empirical research have examined the impact of IoT technologies on improving efficiency, enhancing services, and enabling new business models. With the proliferation of IoT devices generating vast amounts of data, research has focused on data management strategies and analytics techniques for extracting actionable insights from IoT data. This includes data preprocessing, storage solutions (e.g., cloud, edge, fog computing), and analytics algorithms (e.g., machine learning, data mining). Recent literature has emphasized the importance of edge computing and fog computing paradigms in IoT systems. Edge computing involves processing data closer to the source (i.e., IoT devices), while fog computing extends this concept by distributing computing resources across the network edge. Research has explored the benefits of these approaches in reducing latency, improving scalability, and enhancing data privacy. Current literature discusses emerging trends and future directions in IoT technologies, including 5G connectivity, AI (Artificial Intelligence) integration, blockchain for IoT security, and edge intelligence. Researchers anticipate continued advancements in IoT technologies, leading to more pervasive and intelligent IoT deployments across various industries.

The applications of IoT (Internet of Things) technology span across numerous domains, and extensive research has been conducted to explore its potential impact and benefits in various sectors. IoT has revolutionized healthcare through applications such as remote patient monitoring, smart medical devices, and telemedicine. Research in this domain explores the use of IoT for real-time health monitoring, chronic disease management, elderly care, and hospital asset tracking to improve patient outcomes and reduce healthcare costs. IoT plays a crucial role in building smart and sustainable cities by optimizing resource utilization, improving public services, and enhancing quality of life. Literature in this domain discusses IoT applications for traffic management, waste management, smart grids, environmental monitoring, urban mobility, and public safety. Research explores IoT solutions for smart irrigation, crop monitoring, pest detection, livestock tracking, and supply chain optimization to increase agricultural productivity and sustainability. Literature in this domain of Industrial IoT discusses IoT-enabled solutions for asset management, predictive maintenance, condition monitoring, supply chain optimization, and industrial automation to enhance productivity and reduce downtime.

The impact of IoT (Internet of Things) on human life has been a subject of extensive research and analysis in academic literature. Scholars have explored various dimensions of this impact, including social, economic, environmental, and ethical considerations. Research has investigated how IoT technologies influence social interactions, behaviour, and relationships.

Studies have examined the implications of IoT-enabled devices such as smartphones, wearables, and smart home devices on individuals' daily lives, social connectivity, and well-being. Literature has analyzed the economic implications of IoT adoption for individuals, businesses, and economies. Researchers have studied the potential economic benefits of IoT in terms of productivity gains, cost savings, revenue generation, and job creation across various industries. Scholars have examined the environmental implications of IoT technologies in terms of energy consumption, resource efficiency, and sustainability. Research has explored how IoT-enabled solutions such as smart grids, smart buildings, and smart transportation systems contribute to energy conservation, waste reduction, and greenhouse gas emissions reduction.

Objectives of the Study

1. To Evaluate Current IoT Implementations and Innovations
2. To Identify Key Domains of Human Enhancement
3. To Propose Strategies for Maximizing Human Enhancement through IoT
4. To Analyze Potential Future Impacts and Challenges

Research Methodology, Data Collection and Interpretation

Conducting surveys, experiments, and statistical analyses to quantify the impact of IoT on various aspects of human life such as healthcare, transportation, agriculture, and home automation. Utilizing interviews, case studies, and focus groups to understand the qualitative experiences and perceptions of individuals interacting with IoT devices and systems. Combining quantitative and qualitative approaches to gain a comprehensive understanding of how IoT technologies affect different aspects of human life. Deploying IoT sensor networks to collect real-time data on environmental parameters, human behaviours, and system performance. Utilizing wearable IoT devices to monitor health metrics, track physical activity, and provide personalized feedback to users. Integrating IoT sensors into infrastructure such as buildings, roads, and utilities to optimize resource usage, improve safety, and enhance overall efficiency.

1. Applications of IoT Devices

Since Internet of Things technology can be understood as devices/objects that are connected to a network, the applications of IoT devices are endless. Leaders of different companies/organizations are recognizing the potential of IoT technology to make an impact and are investing more in these key pieces of technology in order to reap the benefits. Nearly any object can be outfitted with the appropriate technology that will be involved in the data transmission from IoT devices and their connected networks. Writing about the different applications would be a long and arduous process and it would be more beneficial to first understand the most common applications of IoT devices before exploring what the future holds. Figure 1 shows growth in IoT globally.

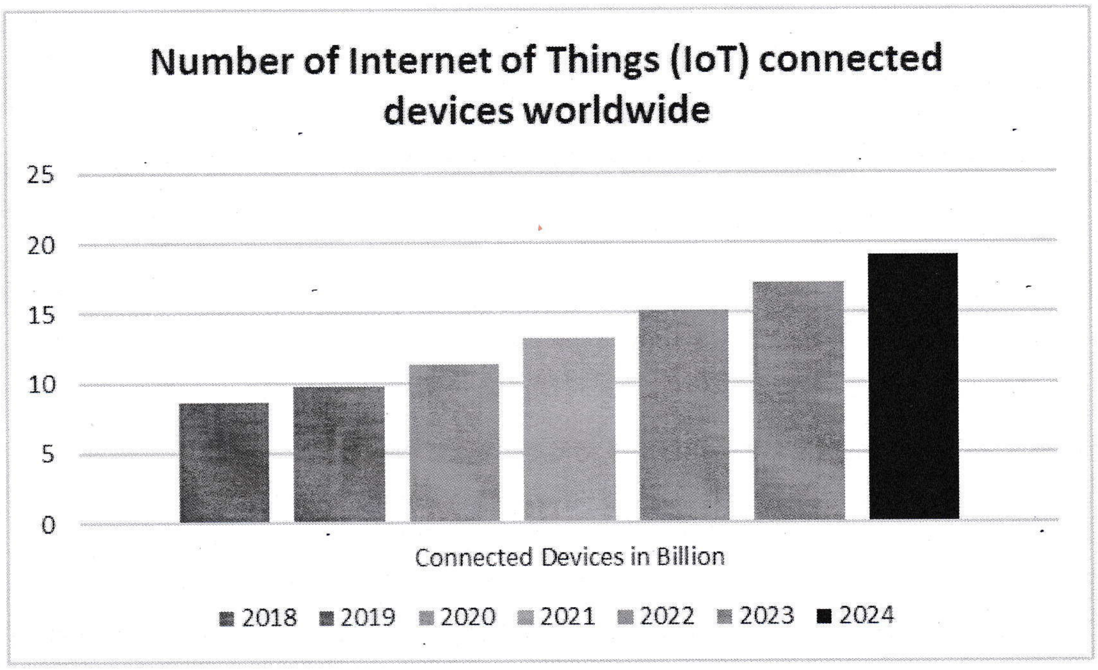


Figure 1: Number of IoT connected devices worldwide

1.1. Healthcare Applications

The Internet of Medical Things (IoTM) is an emerging subfield that is changing the way healthcare is being delivered through the use of IoT technology (Joyia). The use of IoT technology in healthcare has come a long way and continues to be a promising area for growth. Essential innovations, such as the AliveCor heart monitor, which relies on IoT sensors, show how useful technology can be when applied to healthcare in efforts to save lives. Advances in technology have consistently played a major role in the healthcare industry, and IoT devices have found numerous applications in healthcare settings. One way that IoT devices are useful in healthcare is through the use of remote health monitoring in order to monitor patients at home rather than in hospitals. The information that is collected from IoT devices is helpful in medical settings because it can be analyzed and used in ways such as early disease prediction. IoT sensors even played a critical role during the COVID-19 pandemic in helping healthcare workers better monitor critical parameters that could save lives if changes were detected right away. By examining these different IoT device applications in the healthcare industry, researchers can find additional ways to advance this field of research. Table 1 shows IoT enabled healthcare helpful during COVID-19 Pandemic.

Sl. No	Application
1	Treatment of COVID-19 patient
2	Smart Hospital
3	Storage of COVID-19 patient data
4	Originate multiple sources and devices
5	Accurate decision making
6	Monitor status of the COVID-19 patient
7	Alert about the COVID-19 disease

Table 1: IoT enabled healthcare helpful during COVID-19

Medical sensors are important in collecting useful information about a patient's health; however, this information is often very sensitive in nature, and this makes privacy a major concern moving forward. Security has always played a vital role in IoT technology; however, it matters even more in a situation such as healthcare, where IoT devices will be collecting sensitive information about patients that is private in nature. If a patient's medical information was compromised, this could lead to consequences for hospital organizations that did not employ the proper security measures to prevent it. The privacy and confidentiality of a patient's medical information are core concerns when addressing the security vulnerabilities of healthcare IoT devices.

1.2. Agriculture Applications

As the population of the world grows at an exponential rate, the need for efficient food delivery systems is becoming a core issue that is a driver behind the advancements in smart agriculture. In addition to the growing demand, factors such as climate change and water scarcity have also played roles in the increasing demand for more efficient agriculture systems. Much of the technology around IoT implementation aims to reduce agricultural resource waste as shown on Figure 2. The use of IoT technology in agricultural settings is critical to maintaining efficient operations and represents another common use case of IoT technology. Food supply chains that deliver quality and quantity are important to feed the world, and having efficient systems built around these supply chains will benefit people all over the world. The need for more efficient food-delivery systems has helped to promote IoT use in agriculture because stakeholders saw the benefits that technology could provide.

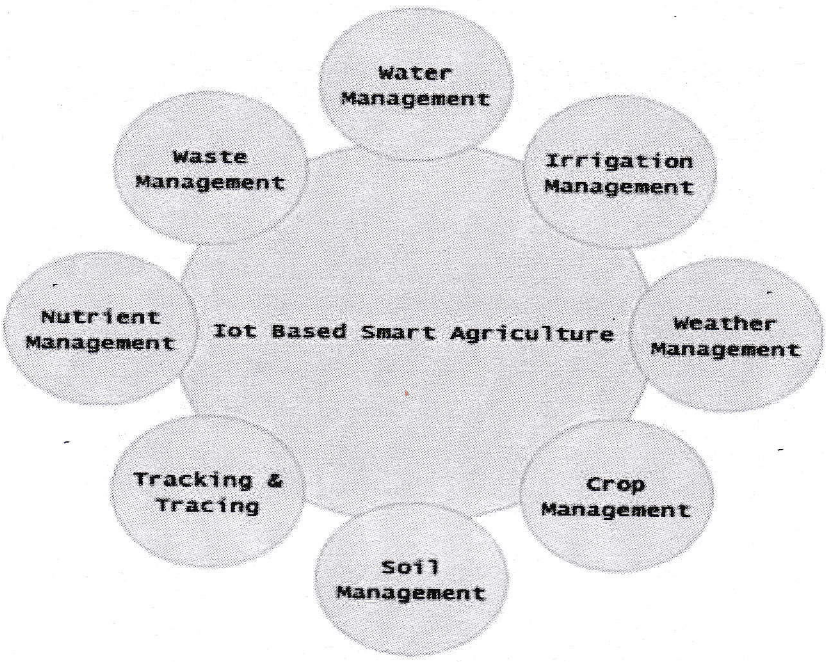


Figure 2: Different types of agriculture applications for IoT

1.3. Industrial automation and smart manufacturing

We've entered a new era for manufacturing, dubbed Industry 4.0, and characterized by widespread digitalization. Prior to this fourth major transformation in modern manufacturing, there was the lean revolution of the 1970s, the outsourcing trend of the 1990s, and the automation boom that began in the 2000s.

Even at this early stage, manufacturer commitment to digital transformation is strong. Preliminary findings from Aberdeen Group's analysis found that 35% of manufacturers plan to achieve digital transformation (industrial IoT, Industry 4.0, smart manufacturing). A key part of digital transformation is the Internet of Things, which is positioned to revolutionize the entire manufacturing value chain by providing an unprecedented level of connectedness and functionality. For consumers, this change comes in the form of small, highly connected devices (smartphones, tablets, GPS devices) and sophisticated electronics embedded into our transport means, living spaces, and workplaces. For manufacturing firms, this change empowers them with new ways to develop, innovate, and manufacture due to the endless connections that can take place. Indeed, Industrial IoT (IIoT) is the subset of IoT that concerns itself with connected manufacturing operations to develop products and services.

The transformative potential of Industrial IoT in manufacturing is staggering; overall, 91% of survey respondents see manufacturing benefits in the IoT.

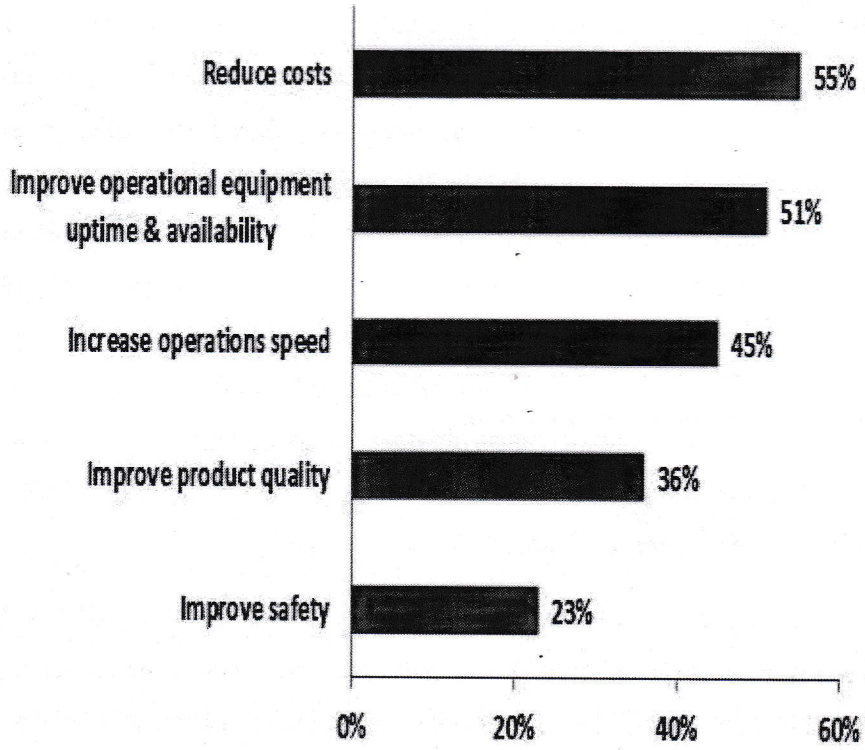


Figure 3: Manufacturing Benefits of Industrial IoT

Figure 3 Shows Benefits of Industrial IoT based on the survey, Survey respondents are eager to "operationalize" the benefits of the Industrial IoT. While they most certainly see the cost benefits of the IIoT, the bigger benefit is perhaps the IIoT's ability to enable connected operations. In particular, manufacturers wish to improve operational equipment uptime and availability. Obviously, the potential of IIoT-enabled data feeds from connected equipment and processes will drive a whole new wave of predictive analytics that will be instrumental in achieving this goal.

1.4. Home automation and smart appliances

Smart home applications represent promising use cases in which people benefit from IoT technology and there are numerous advantages/disadvantages to consider. Smart home devices date back to the 1970s when the X10 protocol was first conceived; this technology allowed for smart home devices to communicate properly. IoT devices in smart homes can be used in a variety of ways, such as measuring home conditions, managing home appliances, and controlling home access. Home automation remains a core feature around which IoT technology is applied. For example, there are numerous home appliances that can be turned on and equipped with IoT technology in order to become more efficient and convenient. There are many benefits that extend beyond convenience. The use of IoT sensors in smart homes can be used to assist the elderly in turning hard-to-reach devices on/off and even detect falls through the use of floor or camera sensors. The market is being driven by the rising popularity of smart devices, such as speakers offered by Amazon and Google. According to a recently released report by Strategy Analytics, the global smart home market has had a positive outcome in recent years. The report further estimates a compound annual growth rate (CAGR) of 10% from 2018 to 2023, leading to a market value of USD 155 billion.

2. Benefits of IoT for Human Enhancement

The benefits of IoT for human enhancement are vast and transformative, spanning various aspects of daily life, healthcare, productivity, and sustainability.

Healthcare Monitoring and Management: Remote Patient Monitoring (RPM): IoT devices enable continuous monitoring of vital signs, medication adherence, and disease progression from the comfort of one's home. This facilitates early detection of health issues and proactive intervention, reducing hospitalizations and improving patient outcomes. Chronic Disease Management: IoT technologies support the management of chronic conditions such as diabetes, hypertension, and heart disease by providing real-time data on patients' health status and enabling personalized interventions.

Safety and Security: Smart Home Automation: IoT-enabled smart home devices enhance safety and security by monitoring for intrusions, fires, gas leaks, and other hazards. They also allow for remote control of lighting, heating, and appliances, improving energy efficiency and reducing accidents. Personal Safety Wearables: Wearable IoT devices such as smartwatches and fitness trackers incorporate features like GPS tracking, fall detection, and emergency alerts, providing peace of mind for users and their caregivers.

Productivity and Efficiency: Industrial Automation: In manufacturing and logistics, IoT facilitates process optimization, predictive maintenance, and real-time monitoring of equipment and assets, leading to increased efficiency, reduced downtime, and cost savings. Smart Agriculture: IoT sensors and monitoring systems enable precision agriculture techniques such as soil moisture monitoring, crop health tracking, and automated irrigation, optimizing resource usage and improving crop yields.

Environmental Sustainability: Smart Energy Management: IoT-enabled energy monitoring and control systems optimize energy consumption in buildings and industrial facilities, reducing waste and carbon emissions. Environmental Monitoring: IoT sensors collect data on air quality, water quality, and pollution levels, enabling proactive measures to mitigate environmental degradation and protect public health.

Accessibility and Inclusivity: Assistive Technologies: IoT devices and applications enhance accessibility for individuals with disabilities, providing features such as speech recognition, gesture control, and voice commands to interact with technology and navigate the environment independently. Smart Assistants and Voice Interfaces: IoT-powered virtual assistants and voice-controlled devices offer intuitive interfaces for people with limited mobility or visual impairments, improving their ability to access information, communicate, and perform daily tasks.

3. Potential Future Impacts and Challenges

3.1. Security of IoT Devices

The use of Internet of Things devices is progressively becoming more prevalent in the daily lives of people around the world; however, cyberattacks remain a large threat to the safe use of IoT. Different examples of these devices are mobile phones, alarms, medical sensors, smartwatches, security systems, and more. The use of these devices continues to expand, and the need for strong security is vital to their success. Although these devices bring convenience, they come with many security issues and vulnerabilities. Since IoT devices often collect sensitive data, these data transmissions can be intercepted by third parties who intend to conduct harm or use these data for nefarious purposes. In one case, there were even attacks that could target IoT devices, such as the Mirai malware, which would hack/convert devices into its botnet and carry out DDoS attacks as shown in Figure 4. Malware, such as the Mirai and others, take advantage of the vast amounts of poorly protected IoT devices, which commonly suffer from poor configurations and open designs, making them targets. Detecting malware and IoT botnets remains an active research area, and many techniques are being applied. The use of a lightweight approach in the classification of IoT malware through the use of image recognition is just one example. Other ways include the use of machine learning algorithms that use supervised, unsupervised, and reinforcement learning in order to handle tasks such as authentication, access control, and malware detection.

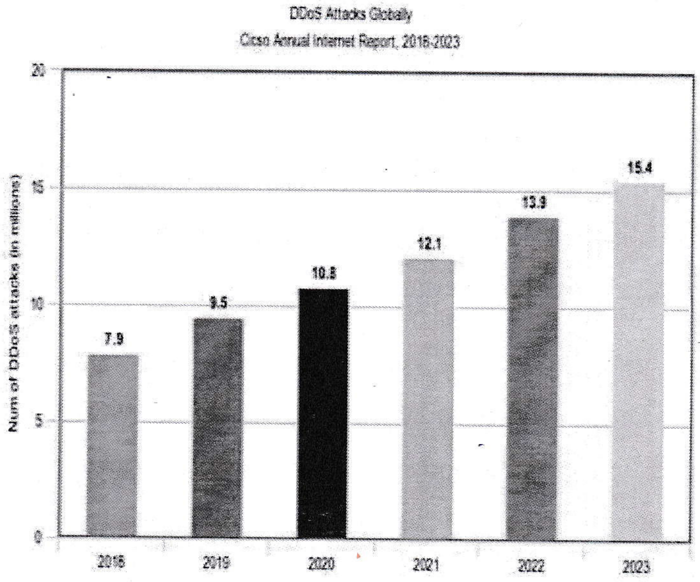


Figure 4: DDoS Attacks Globally

3.2. Authentication and Password Security

One security issue is the lack of security in regard to authentication and passwords. Many IoT devices rely on password security in order to stay protected from cyber criminals who are attempting to gain access to them. These passwords can often be weak, and criminals can have easy access to IoT devices. There is a lack of standardization revolving around how complex passwords should be. Research shows that having more complex password combinations in IoT devices can prevent more cyber-attacks. Even with stronger passwords, there would need to be additional security measures to prevent cyber attackers.

3.3. Interoperability Challenges

Using IoT devices smoothly and without compatibility issues remains a challenge both now and in the future. Interoperability is important because it allows IoT devices to communicate with each other more efficiently. It is a challenge to have IoT devices work together seamlessly because many operate on different infrastructures, devices, APIs, and even data formats. The need for the safe interoperability of IoT devices has even led to the creation of international organizations that develop standards that IoT devices can adopt with the intention of becoming more compatible. The use of protocols and standards, such as

Bluetooth and ZigBee, is critical to the rise of IoT technology because it essentially establishes the rules for use and communication, helping to address interoperability concerns. International organizations, such as IEEE, the Internet Engineering Task Force (IETF), OneM2M, and others, have developed important standards and protocols and play integral roles in influencing the IoT. One important contribution to addressing these challenges is the BiG IoT project, which is an initiative that seeks to create a common API that different IoT devices could communicate through.

3.4. Potential Future Directions and Developments

The potential future directions and developments of IoT in healthcare, agriculture, smart homes, smart cities, and Industry 4.0 are poised to bring about transformative changes and advancements. In healthcare, the integration of IoT with telemedicine, wearable health technology, and advanced data analytics holds promise for revolutionizing patient care. Real-time monitoring, personalized treatment plans, and early disease detection can be facilitated through IoT-enabled devices, leading to improved health outcomes. In agriculture, IoT-driven precision farming techniques, such as smart irrigation and crop disease detection, have the potential to optimize resource utilization, conserve water, and enhance crop yield. By leveraging real-time data from IoT sensors, farmers can make informed decisions and implement timely interventions. In the realm of smart homes, the focus will be on seamless integration, energy efficiency, and personalized automation. IoT-enabled smart home solutions will allow for centralized control and management of various devices, optimizing energy consumption and providing tailored experiences for inhabitants. In smart cities, IoT applications will enhance transportation systems, environmental monitoring, and citizen engagement. Intelligent traffic management, real-time tracking of air quality, and participatory governance will contribute to improved mobility, sustainability, and quality of life. Finally, Industry 4.0 will witness the integration of IoT in industrial automation, predictive maintenance, and supply chain optimization. IoT-driven technologies will enable the real-time monitoring of machines, predictive maintenance strategies, and streamlined logistics, leading to enhanced productivity and reduced downtime. Continued research and development in these areas will shape the future of IoT, paving the way for innovative solutions and transformative advancements across sectors.

Conclusion

In conclusion, exploring the future enhancement of human life through the Internet of Things (IoT) reveals a vast landscape of opportunities and possibilities. Through this comprehensive review, we have delved into various aspects of how IoT technologies are poised to revolutionize healthcare, safety, productivity, sustainability, and inclusivity. The benefits of IoT for human enhancement are evident across multiple domains. In healthcare, IoT enables remote patient monitoring, personalized medicine, and telemedicine, leading to improved health outcomes and better access to care. In safety and security, IoT-powered smart home devices and personal safety wearables enhance protection and peace of mind for individuals and families. Productivity and efficiency are augmented through industrial automation and smart agriculture, optimizing processes and resource usage. Environmental sustainability is promoted through smart energy management and environmental monitoring, mitigating environmental risks and preserving natural resources. Additionally, IoT fosters accessibility and inclusivity through assistive technologies and intuitive interfaces, empowering individuals with disabilities to live more independently and participate fully in society.

While the potential of IoT for human enhancement is vast, challenges such as data privacy, security, interoperability, and ethical considerations must be addressed to ensure responsible deployment and usage. Collaboration among stakeholders, including policymakers, industry leaders, researchers, and communities, is essential to harness the full potential of IoT while mitigating risks and addressing concerns. As we continue to innovate and integrate IoT technologies into our daily lives, it is crucial to prioritize human-centric design, ethical principles, and sustainability to create a future where IoT enhances human well-being, fosters social inclusion, and promotes environmental stewardship. By embracing the transformative power of IoT with a holistic and conscientious approach, we can create a future where technology truly enhances the human experience and enables individuals and communities to thrive.

References

1. Pardini K., Rodrigues J.J., Diallo O., Das A.K., de Albuquerque V.H.C., Kozlov S.A. A smart waste management solution geared towards citizens. Sensors. 2020; 20:2380. doi: 10.3390/s20082380. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

2. Halcox J.P., Wareham K., Cardew A., Gilmore M., Barry J.P., Phillips C., Gravenor M.B. Assessment of remote heart rhythm sampling using the AliveCor heart monitor to screen for

atrial fibrillation: The REHEARSE-AF study. Circulation. 2017;136:1784–1794. doi: 10.1161/CIRCULATIONAHA.117.030583. [PubMed] [CrossRef] [Google Scholar]

3. Tao W., Zhao L., Wang G., Liang R. Review of the internet of things communication technologies in smart agriculture and challenges. Comput. Electron. Agric. 2021;189:106352. doi: 10.1016/j.compag.2021.106352. [CrossRef] [Google Scholar]

4. Greg Cline. Industry 4.0 and Industrial IoT in Manufacturing: A Sneak Peek. He is the head of research for Aberdeen's Product Innovation and Engineering (PIE) and Manufacturing research practices.

5. Jo T.H., Ma J.H., Cha S.H.N. Elderly perception on the internet of things-based integrated smart-home system. Sensors. 2021;21:1284. doi: 10.3390/s21041284. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

6. Xiao L., Wan X., Lu X., Zhang Y., Wu D. IoT security techniques based on machine learning: How do IoT devices use AI to enhance security? IEEE Signal Process. Mag. 2018;35:41–49. doi: 10.1109/MSP.2018.2825478. [CrossRef] [Google Scholar]

7. Bröring A., Ziller A., Charpenay V., Thuluva A.S., Anicic D., Schmid S., Seidel C. The big iot api-semantically enabling iot interoperability. IEEE Pervasive Comput. 2018;17:41–51. doi: 10.1109/MPRV.2018.2873566. [CrossRef] [Google Scholar]

PNEUMONIA DETECTION USING DEEP LEARNING

Ms Ashna Satheesh

BCA Student

PG Department of CA & AI

Saintgits College Of Applied Sciences, Kottayam, Kerala

ashnasatheesh15@gmail.com

7025537084

Ms Darsana Rajesh

BCA Student

PG Department of CA & AI

Saintgits College Of Applied Sciences, Kottayam, Kerala

darsanarajesh648@gmail.com

8590470618

Ms Krishnanjali J

BCA Student

PG Department of CA & AI

Saintgits College Of Applied Sciences, Kottayam, Kerala

krishnanjalij28@gmail.com

9539516327

Ms Sruthika R

BCA Student

PG Department of CA & AI

Saintgits College Of Applied Sciences, Kottayam, Kerala

shruthikareghu@gmail.com

6282519468

Dr Meenu Suresh

Assistant Professor

PG Department of CA & AI

Saintgits College Of Applied Sciences, Kottayam, Kerala

meenu@saintgits.org

Abstract:

This paper introduces an advanced deep learning approach for pneumonia detection using convolutional neural networks (CNNs). Trained on a large dataset of annotated chest X-ray images, our model leverages transfer learning and addresses class imbalance through data augmentation and weighted loss functions. Visualization techniques, including Grad-CAM, enhance interpretability, aiding clinicians in understanding the model's focus. Evaluation on a benchmark dataset demonstrates superior sensitivity and specificity compared to traditional methods. Our findings highlight the model's robustness across diverse demographics, emphasizing its potential for early diagnosis and improved patient outcomes. The study underscores the transformative impact of deep learning on pneumonia diagnosis, providing a valuable tool for efficient and accurate healthcare practices.

Keywords: Pneumonia detection, Deep Learning, Convolutional neural networks(CNNs), Transfer learning, Robustness, Grad-CAM, Diagnostic radiology, Specificity.

I. Introduction

Pneumonia remains a pervasive global health challenge, demanding swift and precise diagnostic methodologies for effective patient management. Traditional diagnostic approaches often face limitations in accuracy and timeliness, necessitating innovative solutions to augment the capabilities of healthcare professionals. In response to this critical need, our research on the application of deep learning techniques, specifically convolutional neural networks (CNNs), for pneumonia detection through the analysis of chest X-ray images.The advent of deep learning has ushered in a new era in medical image analysis, offering unprecedented potential to revolutionize the field of diagnostic radiology. Convolutional neural networks, in particular, have exhibited remarkable proficiency in discerning intricate patterns and subtle features within medical images. Leveraging these advancements, our study proposes a state-of-the-art methodology that amalgamates advanced image processing and machine learning algorithms. This integration aims to increase the accuracy and efficiency of pneumonia diagnosis, providing a robust and reliable tool for healthcare practitioners.In this test, we delve into the intricacies of our deep learning framework, meticulously trained on a substantial dataset of annotated chest X-ray images. The model's ability to discern nuanced patterns indicative of pneumonia presence is optimized through transfer learning from pre-trained architectures. To mitigate challenges posed by class imbalance inherent in diagnostic image datasets, we employ sophisticated strategies such as data augmentation and weighted loss functions, ensuring the model's resilience in diverse clinical scenarios.Furthermore, the interpretability of deep learning framework is a pivotical aspect often overlooked in medical diagnostics. Our research emphasizes the importance of providing clinicians with insights into the decision-making process. To this end, we employ visualization techniques, including Grad-CAM, to elucidate the regions of interest in X-ray images, fostering a deeper understanding of the model's diagnostic focus.As we proceed, this paper will present the comprehensive evaluation of our proposed approach on a benchmark dataset, showcasing its superior performance compared to traditional diagnostic methods. The study will not only highlight the model's sensitivity and specificity but also explore its robustness across varied patient demographics, affirming its potential applicability in real-world clinical settings.Our research contributes to the ongoing paradigm shift in pneumonia diagnosis through the fusion of deep learning and medical imaging. The transformative impact of this work extends beyond technological innovation, aiming to provide clinicians with a reliable and interpretable tool for accurate pneumonia detection, ultimately improving patient outcomes in the realm of respiratory health.

II. LITERATURE REVIEW

The recent emergence of deep learning has transformed medical imaging analysis, particularly in the realm of pneumonia detection. Convolutional neural networks (CNNs), a cornerstone of deep learning, have showcased impressive abilities to discern subtle patterns and features within medical images, often outperforming human experts in specific diagnostic tasks. This shift toward automated image analysis presents vast potential for enhancing diagnostic precision, mitigating interpretation inconsistencies, and streamlining patient care pathways.

The paper "Deep Learning for Automatic Pneumonia Detection by Tatiana Gabruseva" [1] presents a deep learning approach for automatic pneumonia detection, addressing its status as a major global cause of death, especially among young children. Current detection methods based on chest X-ray examinations by specialists are time-consuming and prone to disagreements. The proposed computational method utilizes deep convolutional neural networks, single-shot detectors, and squeeze-and-extinction mechanisms to automatically identify pneumonia areas. The approach showed promising results in the Radiological Society of North America Pneumonia Detection Challenge, achieving top performance.

The thesis by Alaa M. A. Barhoom and Prof. Dr. Samy S. Abu Naser [2] from Al-Azhar University focuses on using deep learning for pneumonia detection and classification using X-ray imaging. Deep learning, a subset of machine learning, enables computers to analyze data inputs and output values within a specific range. The aim is to develop an effective method for pneumonia detection based on X-rays to aid chest doctors in making accurate decisions. The thesis includes designing and implementing a model with deep learning algorithms and testing it on various chest X-ray images, with results discussed in detail. Deep learning, mostly used in driverless cars and consumer devices, is gaining attention in artificial intelligence and medical imaging.

The paper "Pneumonia Detection Using Deep Learning Based on Convolutional Neural Network "by Luka Racic, Tomo Popovic [13] discusses the rising use of artificial intelligence (AI) in medicine, particularly in analyzing chest X-ray images for pneumonia diagnosis using machine learning algorithms like deep learning. It highlights AI's role in supporting faster and more accurate decision-making processes. The research focuses on classifying X-ray images into pneumonia-related changes or not. It also mentions the emergence of fields like Machine Learning (ML) and Deep Learning (DL) and concludes with AI's significant advancements, such as Google's AlphaGo beating world champion Gary Kasparov in chess in 2016. Overall, AI facilitates better decision-making in medicine, especially in analyzing biological image formats.

The paper "Pneumonia Detection Using CNN based Feature Extraction" [16] discusses the importance of detecting pneumonia, a life-threatening infectious disease primarily caused by Streptococcus pneumoniae, which is a significant cause of mortality in India. Developing an automatic detection system for pneumonia is crucial, especially in remote areas lacking access to expert radiologists. Convolutional Neural Networks (CNNs) are highlighted as effective tools for disease classification, particularly in analyzing chest X-ray images. Pre-trained CNN models facilitate feature extraction and classification, enhancing the accuracy of pneumonia detection. The study underscores the potential of autonomous systems using CNN-based feature extraction to aid in prompt diagnosis and treatment of pneumonia.

The study conducted by Thawsifur Rahman on "Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray" [14] aims to detect bacterial and viral pneumonia using digital chest X-ray images. They employ four pre-trained deep Convolutional Neural Networks (CNNs) for transfer learning: AlexNet, ResNet18, DenseNet201, and SqueezeNet. With a dataset of 5247 chest X-ray images, including germ cases that contain bacteria, virus and normal cases, the study achieves impressive accuracy rates. Specifically, they report 98% correctness for normal vs. pneumonia images, 95% for bacterial vs. viral pneumonia images, and 93.3% for distinguishing viral, bacterial, and normal pneumonia. These findings offer potential for enhancing pneumonia diagnosis by radiologists and facilitating rapid airport screening of pneumonia patients, crucial for early and accurate treatment.

The paper "A Deep Feature Learning Model for Pneumonia Detection Applying a Combination of mRMR Feature Selection and Machine Learning Models" by M. Togaçar ,B. Ergen b, Z. Cömert [15] applied deep learning models and image augmentation techniques to discover pneumonia in chest X-ray images. Researchers combined features extracted from CNN models and reduced them using the mRMR feature selection method. They achieved optimal results with various classifiers, using lung X-ray images for diagnosis. For each deep model,the number of deep features was decreased from 1000 to 100 which gives a total of 300 deep features as result. The study found that deep features provided robust pneumonia detection, and the mRMR method effectively reduced feature set dimensionality.

III. METHODOLOGY

Three sections make up the approach of the work, data pre-processing, balancing the imbalanced classifiers and model descriptions.

Dataset Description

A well-known Machine Learning wearhouse for sharing and assessing datasets, Kaggle, provided the dataset used in this study to assess students' ability to adapt to online learning. Kaggle provides a wide range of datasets that have been contributed by individuals, organizations, and scholars who operate in various sectors. The dataset employed must be consistent with the goals and parameters of our investigation. Researchers thoroughly evaluated the dataset's dependability and quality. This assessment comprises reviewing the dataset for correctness, completeness, and relevance in order to ensure that it meets the need of our learning.

Figure 1 : Loss & Accuracy of VGG16

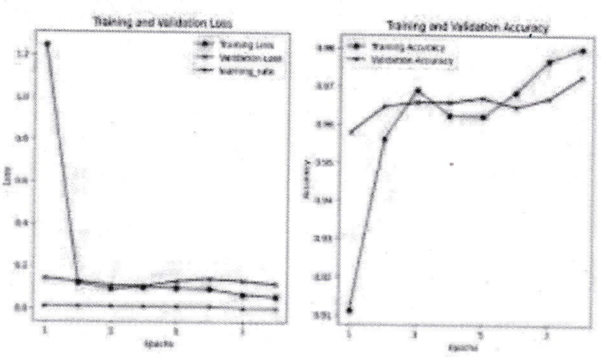


Figure 2: Confusion matrix of VGG16

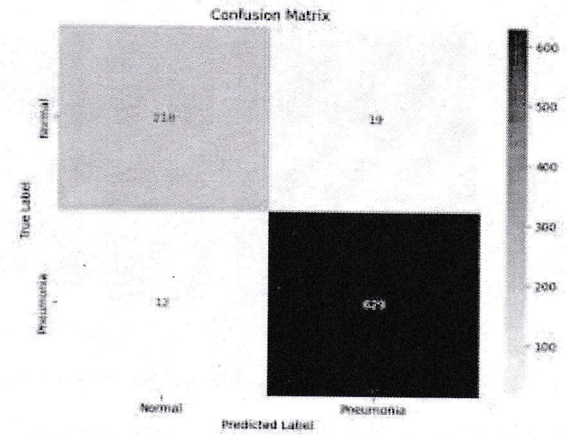


Figure3: Loss & Accuracy of Xception Model

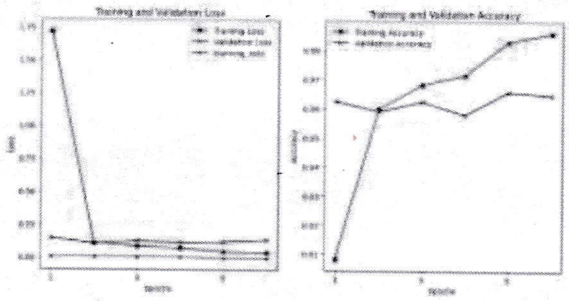


Figure 4: Confusion Matrix for Xception model.

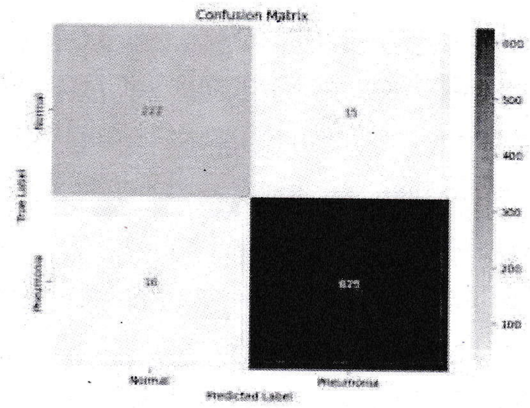


Figure 5: Normal Chest X-Ray

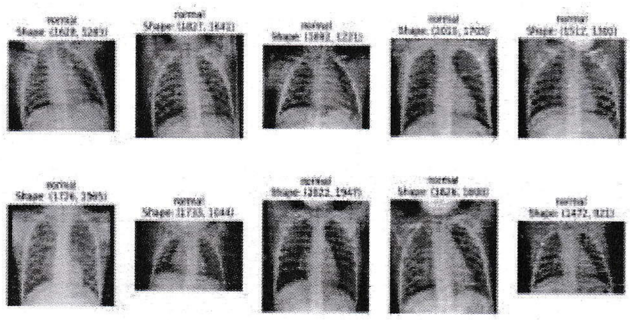
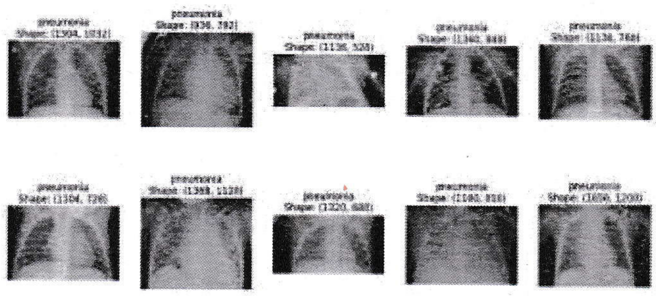


Figure 6: X-Ray detected with pneumonia



The dataset includes both normal chest X-rays and X-rays detected with pneumonia. Figure 5 shows the normal chest X-Rays and Figure 6 shows X-Rays detected with pneumonia

A. *Data Preprocessing*

In this study before the analysis, cleaning, transforming, and organizing the data to ensure its quality and suitability for the research objective has been completed. Through meticulous data preprocessing, including collection, cleaning, augmentation, normalization, and class balancing, the project seeks to ensure the quality and suitability of the dataset for training robust CNN models, ultimately facilitating timely and accurate pneumonia diagnosis.

B. *Description of Models*

This study explores various machine learning and deep learning methodologies for pneumonia detection. The algorithms and techniques investigated include CNN + Data Augmentation (color jitter),CNN + Data Augmentation (Color jitter + Layers Augmentation),CNN (batch normalization & padding + color jitter),VGG16,Xception Model.

These strategies each offer a unique way to handle and examine medical imaging data in order to identify pneumonia. The study compares the performance and effectiveness of these methodologies in accurately identifying pneumonia cases from chest X-rays or CT scans.

1. *CNN + Data Augmentation (color jitter)*

Convolutional Neural Networks (CNNs) are more resilient and more generalizable in computer vision applications when used with data augmentation methods such as color jittering.CNNs automatically learn spatial hierarchies of features from images, making them effective in tasks like image recognition and object detection. Data augmentation diversifies the learning dataset by applying transformations, reducing overfitting, and improving generalization. Color jittering randomly adjusts image colors, making the model more resilient to variations in lighting and color distribution. This combination improves model performance in real-world scenarios where object appearance may vary due to environmental factors..

2.CNN + Data Augmentation (Color jitter + Layers Augmentation)

Convolutional Neural Networks (CNNs) perform much better in computer vision tasks when combined with data augmentation methods like layer augmentation and color jittering. This combination also improves model robustness. CNNs excel at processing visual data by extracting hierarchical features. Data augmentation, including color jittering, diversifies training data to improve generalization. Layer augmentation enhances model complexity by adding or modifying layers. Color jittering introduces variability in color distributions, making models more resilient to changes in lighting and color. Layer augmentation increases model capacity and adaptability. Together, these techniques create versatile models capable of handling diverse data and real-world scenarios effectively.

3. CNN (batch normalization & padding + color jitter)

Combining Convolutional Neural Networks (CNNs) with techniques such as batch normalization, padding, and color jittering significantly improves model performance in computer vision tasks. CNNs process visual data effectively, extracting hierarchical features crucial for tasks like image classification and object detection. Batch normalization stabilizes training by normalizing layer activations. Padding preserves spatial dimensions during convolutional operations, preventing information loss at image borders. Color jittering introduces variability in color distributions during training, enhancing the model's robustness to changes in lighting and color. Together, these techniques create more stable, robust, and adaptable models capable of handling diverse data distributions and real-world scenarios effectively.

4.VGG16

VGG16 is a widely-used convolutional neural network architecture known for its simplicity and effectiveness in image recognition tasks. Developed by the Visual Geometry Group at the

University of Oxford, it comprises 16 layers, including 13 convolutional and 3 fully connected layers. VGG16 employs 3x3 filters with ReLU activation functions, followed by max-pooling layers for downsampling. With approximately 138 million parameters, it achieved remarkable performance in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2014. VGG16 serves as a base model for transfer learning and continues to be influential in various computer vision applications.

5. Xception Model

Google created Xception, a convolutional neural network architecture, as an expansion of the Inception model.. It replaces standard convolutional layers with depthwise separable convolutions, leading to improved efficiency and performance. Depthwise separable convolutions consist of depthwise and pointwise convolutions, reducing parameters and computational complexity while maintaining or improving model accuracy. Xception is well-suited for various computer vision tasks, including image classification, object detection, and segmentation. Its efficiency and effectiveness make it a valuable tool for both research and practical applications in the area of deep learning.

I. RESULT AND DISCUSSION

The effectiveness of models can be evaluated using a variety of metrics. Precision, Recall, F1 score,Support and Accuracy are the most important characteristics used to assess a model's performance. The value of the confusion matrix which is generated during the testing of the model is considered to calculate the score of the precision, recall, F1- Score, and accuracy.

The performance metrics of various machine learning models utilized in this study are displayed on TABLE 1, encompassing a diverse array of methodologies. The models examined include VGG16, and the Xception Model. The evaluation metrics, such as accuracy, precision, recall, and F1 Score, provide valuable insights into the efficacy of each approach in pneumonia detection from chest X-ray or CT scan images. The study enhances understanding of deep learning architectures in medical imaging, aiding healthcare professionals in making informed decisions regarding pneumonia diagnosis and treatment.

Model	Class Name	Precision	Recall	FI Score	Support
VGG16	Normal	0.95	0.92	0.93	237
	Pneumonia	0.97	0.98	0.98	641
	Accuracy			0.96	878
	Macro avg	0.96	0.95	0.95	878
	Weighted avg	0.96	0.96	0.96	878
Xception	Normal	0.93	0.94	0.93	237
	Pneumonia	0.98	0.98	0.98	641
	Accuracy			0.96	878
	Macro avg	0.95	0.96	0.96	878
	Weighted avg	0.96	0.96	0.96	878

TABLE1: Results of Deep Learning Prototypes

II. FUTURE SCOPE

The paper introduces an advanced deep learning method for pneumonia detection using CNNs. Trained on a large dataset, the model addresses class imbalance and uses visualization techniques for interpretability. It shows superior performance compared to traditional methods. Future directions include refining the model, integrating it into clinical workflows, and expanding its scope for longitudinal monitoring and prognostication.

III. REFERENCES

[1] Tatiana Gabruseva,Dmytro Poplavskiy,Alexandr Kalinin "Deep Learning for Automatic Pneumonia Detection",2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),pp.350-351

[2] Alaa M. A. Barhoom and Prof. Dr. Samy S. Abu Naser "Diagnosis of Pneumonia Using Deep Learning",International Journal of Academic Engineering Research (IJAER) 6 (2):48-68 (2022)

[3] Chouhan, V.; Singh, S.K.; Khamparia, A.; Gupta, N.; Tiwari, P.; Moreira, C.; Damasevicius, R.; De Albuquerque, V.H.C. A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images. Appl. Sci. 2020, 10, 559.

[4] Mohammad Tariqul Islam, Md Abdul Aowal, Ahmed Tahseen Minhaz, and Khalid Ashraf. 2017. Abnormality detection and localization in chest x-rays using deep convolutional neural networks. arXiv preprint arXiv:1705.09850 (2017).

[5] Sirazitdinov, I.; Kholiavchenko, M.; Mustafaev, T.; Yixuan, Y.; Kuleev, R.; Ibragimov, B. Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database. Comput. Electr. Eng. 2019, 78, 388–399.

[6] D. Kermany, K. Zhang, M. Goldbaum, "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification", Mendeley Data, v2, 2018. Accessed July 2020. [Online]

[7] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[8] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, et al., Chexnet: radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv:1711.05225.

[9] D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan, A. Mittal, "Pneumonia Detection Using CNN based Feature Extraction," 2019 IEEE International Conf. on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 2019, pp 1-7.

[10] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, Matthew P. Lungren, and An drewY.Ng. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv:1711.05225v1, 2017.

[11] O. Stephen, M. Sain, U. J. Maduh and D.-U. Jeong, "An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare," Journal of HealthCare Engineering, vol. 2019, 2019.

[12] Benjamin Antin, Joshua Kravitz, and Emil Martayan. 2017. Detecting Pneumonia in Chest X-Rays with Supervised Learning. (2017).

[13] Luka Račić; Tomo Popović; Stevan čakić; Stevan Šandi, Pneumonia Detection Using Deep Learning Based on Convolutional Neural Network, 2021 25th International Conference on Information Technology (IT). 10.1109/IT51528.2021.9390137

[14] Saad Kashem, Amith Khandakar, Tawsifur Rahman: Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray 10.3390/app10093233

[15] M. Toğaçar, B. Ergen, Z. Cömert, F. Özyurt: A Deep Feature Learning Model for Pneumonia Detection Applying a Combination of mRMR Feature Selection and Machine Learning Models 10.1016/j.irbm.2019.10.006

[16] DimpyVarshni; KartikThakral; LuckyAgarwal; RahulNijhawan; Ankush Mittal, Pneumonia Detection Using CNN based Feature Extraction, 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT).10.1109/ICECCT.2019.8869364

EXPLORING THE ETHICAL IMPLICATIONS OF AI DECISION-MAKING ALGORITHMS

Mr. Alex Thankachan
BCA Student
Department of Computer Science and Applications,
Naipunnya School of Management, Cherthala, Kerala
University of Kerala
alexthankachan95@gmail.com

Ms. Amulia P.M
Assistant Professor
Department of Computer Science and Applications, Naipunnya School of Management,
Cherthala, Kerala
University of Kerala
csa.amuliapm@naipunnyacollege.ac.in

Ms. Girly Antony
Assistant Professor
Department of Computer Science and Applications, Naipunnya School of Management,
Cherthala, Kerala
University of Kerala
girlyantony.ga@gmail.com

Abstract

AI research has had a profound impact on various industries, revolutionizing the way they operate and opening up new possibilities. In healthcare, AI-powered systems are helping doctors diagnose diseases, predict patient outcomes, and develop personalized treatment plans. In finance, AI algorithms are being used to detect fraud, optimize investment portfolios, and provide personalized financial advice. The manufacturing industry is benefiting from AI-driven automation, improving productivity and efficiency. AI is also transforming transportation, retail, agriculture, and many other sectors, making processes more efficient and enhancing customer experiences. The origins of AI research can be traced back to the Dartmouth Conference in 1956, where the term "Artificial Intelligence" was coined. Since then, researchers have been exploring various aspects of AI, including natural language

processing, computer vision, machine learning, and robotics. Over the years, AI research has evolved, thanks to advancements in computing power, algorithms, and the availability of massive amounts of data. This evolution has led to breakthroughs such as deep learning and reinforcement learning, enabling AI systems to perform tasks that were once considered impossible.

The paper discusses about some ethics in artificial intelligence and various ethical implications and corresponding algorithms related to real world applications.

Keywords: - Computer Systems, Algorithms, Human Intelligence, Biased Outcomes, Transparency, Real-World Applications

INTRODUCTION

Artificial Intelligence (AI) has emerged as a transformative technology, revolutionizing industries and reshaping the way we live and work. AI is typically implemented as a system comprised of both software and hardware. From a software standpoint, AI is mainly concerned with algorithms. An artificial neural network (ANN) is a conceptual framework for developing AI algorithms. It is a human brain model made up of an interconnected network of neurons connected by weighted communication channels.AI uses various algorithms to find complex non-linear correlations in massive datasets (analytics). Machines learn by correcting mirror algorithmic errors (training),thereby boosting prediction model accuracy[1]. Artificial intelligence (AI) is a term used in computing to describe a computer programs capacity to execute tasks associated with human intelligence such as reasoning and learning.

RESEARCH METHODOLOGY

The Method used in this research is internet-based research. Internet-based research methods involve using the Internet to gather data for a research study. These methods can include the analysis of online data such as social media posts or website. Internet-based research methods offer the advantage of being able to reach a large and diverse area quickly and inexpensively and can be especially useful for studying topics that may be sensitive or difficult to access through other means.

APPLICATIONS OF ARTIFICIAL INTELLIGENCE ACROSS VARIOUS INDUSTRIES

KEY TAKEAWAYS:

- Potential applications of artificial intelligence continue to expand as more people adopt the technology.

- AI usage is particularly prominent in finance, digital spaces like social media, ecommerce, and e-marketing and even healthcare.
- For investors who want to put AI to work for them.

CURRENT APPLICATIONS OF AI RESEARCH

The current applications of AI research are vast and diverse. Natural language processing and machine learning techniques are powering virtual assistants like Siri and Alexa, enabling them to understand and respond to human commands. Computer vision is being used in self-driving cars, surveillance systems, and facial recognition technology. AI algorithms are helping e-commerce platforms recommend products based on user preferences and behavior. Chatbots are becoming increasingly sophisticated, providing customer support and enhancing user experiences. AI-powered robots are assisting in manufacturing, healthcare, and even household chores. The possibilities are endless, and we are only scratching the surface of what AI can do.

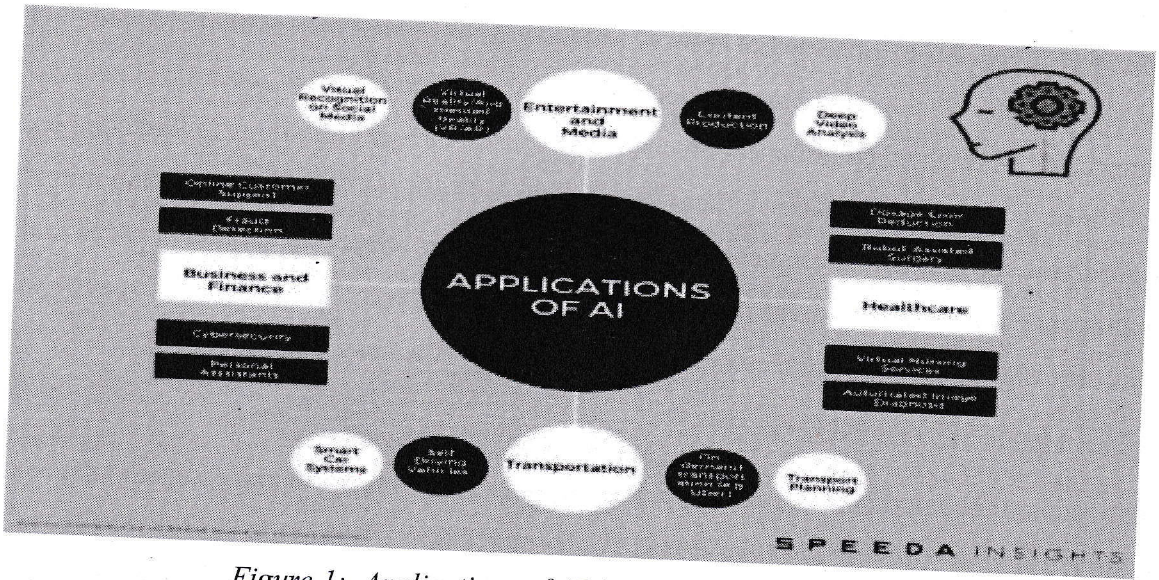


Figure 1: Applications of AI (Source: Speeda insights)

CURRENT STATUS OF AI AND THE FUTURE SCOPE IN THE MEDICAL FIELD AND HEALTHCARE

Artificial intelligence has developed very rapidly in recent years, resulting in a huge impact on all sectors of society. People are also paying more and more attention to AI. Artificial *intelligence application in the field of medical treatment are more common, this paper introduces the main applications of artificial intelligence in medicine, common applications include* virtual assistant ,AI auxiliary diagnosis and treatment, medical robots feedthrough in depth analysis of the various applications found applications in the field of artificial intelligence in the process of all kinds of problems, There are common problems such as lack of unified norms and shortage of artificial intelligence talents, and the above problems are discussed in detail. Finding and solving problems provides a strong guarantee for the development of artificial intelligence in the medical field. It is believed that AI will provide greater help to both doctors and patients [2]. In this paper, researchers have summarized recent enlargements of AI in e-health in many important applications, which is effectively materialized in COVID-19 also.

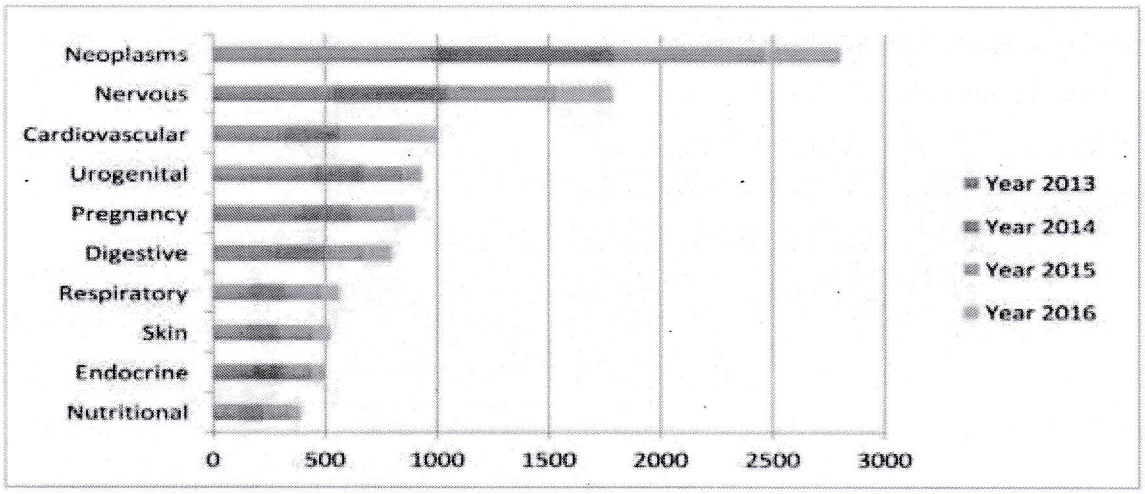


Figure 2: Artificial intelligence in Healthcare; Past, Present and Future (Source: Jiang F, Jiang Y, Zhi H, et al Artificial intelligence in healthcare: past, present and future Stroke and Vascular Neurology 2017;2:DOI: 10.1136/svn-2017-000101)

CURRENT STATUS OF AI AND THE FUTURE SCOPE IN THE TRANSPORTATION AND MANUFACTURING SECTOR

Data-driven decision making is becoming an integral part of manufacturing companies.

Data is collected and commonly used to improve efficiency and produce high quality items for the customers. IoT-based and other forms of object tracking are an emerging tool for collecting movement data of objects/entities (e.g., human workers, moving vehicles, trolleys etc.) over space and time. Movement data can provide valuable insights like process bottlenecks, resource utilization, effective working time etc. that can be used for decision making and improving efficiency. Turning movement data into valuable information for industrial management and decision making requires analysis methods. We refer to this process as movement analytics. The purpose of this document is to review the current state of work for movement analytics both in manufacturing and more broadly. We survey relevant work from both a theoretical perspective and an application perspective. From the theoretical perspective, we put an emphasis on useful methods from two research areas: machine learning, and logic-based knowledge representation. We also review their combinations in view of movement analytics, and we discuss promising areas for future development and application [3]. The report also states that on-demand car services, like Uber, will make a complete transition to autonomous operations in the near future.

CURRENT STATUS AND THE FUTURE SCOPE OF AI IN SOCIAL MEDIA AND ENTERTAINMENT

In this digital era in which an unimaginable amount of data is generated daily, AI has become a significant part of the major social networks that contribute massively to managing this vast data coming to social media platforms. AI tools help enhance features of social media platforms and lead social media activities at scale across a number of use cases, including text and visual content creation, social media monitoring, ad management, influencer research, brand awareness campaigns and more. AI-powered visual content creation tools are also going mainstream. Text-to-image AI models like DALL-E, Midjourney and Stable Diffusion have been revolutionizing the way visual content is created. These systems use machine learning algorithms to create images from a text description or even to generate new variations of existing images. While many creatives have embraced AI text-to-image generators as an exciting tool for streamlining digital art creation, others have watched the arrival of this technology as controversial regarding the role of AI in visual art and issues concerning style appropriation [4]. Beyond the analysis of user-generated content for indexing and search purposes, monitoring content and users on social networks can provide valuable information and knowledge on specific communities, on people's behavior and opinions, on societal trends, etc.

CURRENT STATUS AND THE FUTURE SCOPE OF AI IN BUSINESS AND FINANCE

There are numerous obstacles to the long-term success of small and medium-sized businesses as a result of rising rivalry and pervasive internal inconsistencies caused by the economy's breakneck expansion (SMEs). Because it represents both the company's current development status and its possible future prospects, a scientific evaluation of the company's capacity for sustainable growth is crucial. To make the best decisions as a group, it's helpful to employ both linguistic term sets and tentative fuzzy preference relations. E-commerce activities have been encouraged as a means to improve corporate-efficiency thanks to the recent technological revolution and the growth of the digital economy. [5]. Increasing online store revenue by utilizing MIS and AI is a well-established practice. That technique has grown in significance for modern corporate operations.

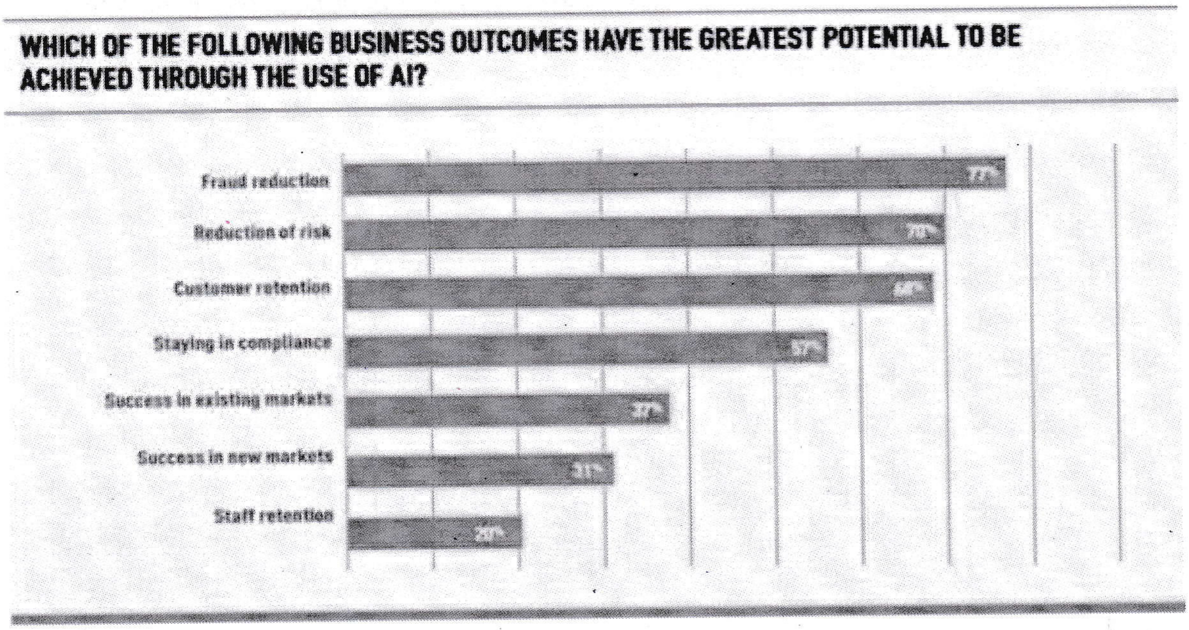


Figure 3: Artificial intelligence in Finance (Source: Business Intelligence Trends 2020)

EXCITING POSSIBILITIES OF AI RESEARCH IN THE FUTURE

The future of AI research holds even more exciting possibilities. One area that is gaining significant attention is the development of general AI, which is AI capable of performing any intellectual task that a human can do. Imagine a world where AI systems can reason, learn, and adapt to new situations like humans. This could lead to advancements in fields such as education, scientific research, and creativity. Another exciting possibility is the integration of AI with other emerging technologies like the Internet of Things (IoT) and blockchain. This could enable smart cities, autonomous vehicles, and secure decentralized systems. The potential for AI to transform society is immense, and the future is ripe with opportunities.

CHALLENGES AND ETHICAL CONSIDERATIONS IN AI RESEARCH

While the possibilities of AI research are exciting, it is crucial to address the challenges and ethical considerations associated with this technology. One of the key challenges is the ethical use of AI. There are concerns about bias in AI algorithms, privacy issues, and the potential for AI systems to be used for malicious purposes. Transparency, accountability, and responsible AI development are essential to ensure that AI benefits society as a whole. Additionally, there is a need for regulations and policies to govern the use of AI and protect individual rights. As researchers, we must be mindful of these challenges and work towards developing AI systems that are fair, transparent, and accountable.

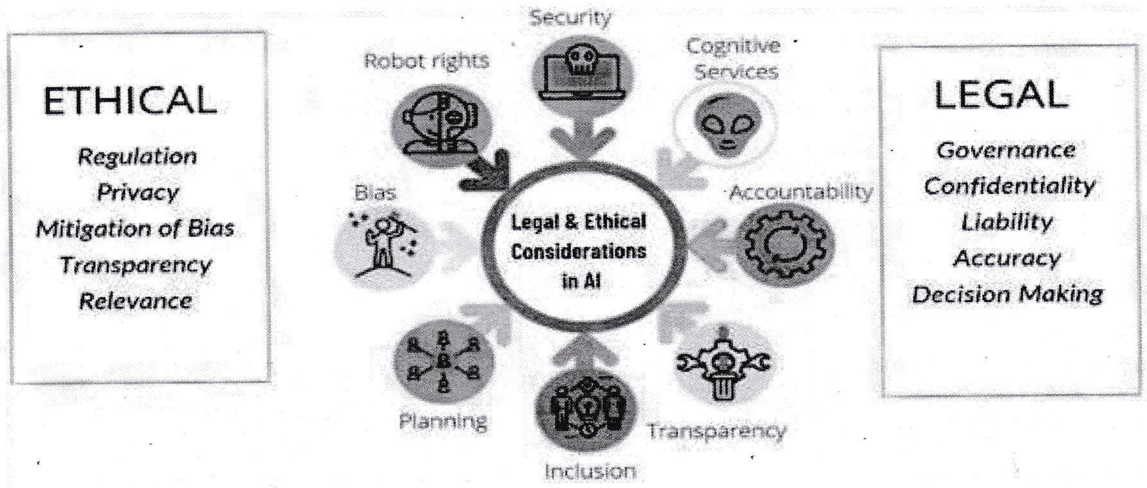


Figure 4: The various ethical and legal concerns associated with the use of AI in healthcare settings (Source: https://www.frontiersin.org/articles/10.3389/fsurg.2022.862322/full#F1)

THE ROLE OF GOVERNMENT AND ORGANIZATIONS IN SUPPORTING AI RESEARCH

To fully realize the potential of AI research, it is essential for governments and organizations to provide support and investment. Governments can play a crucial role in funding research initiatives, creating policies that encourage innovation, and fostering collaboration between academia, industry, and research institutions. Organizations can contribute by allocating resources for AI research, promoting a culture of innovation, and providing platforms for researchers to share their findings. By working together, we can accelerate the development and adoption of AI technologies, driving economic growth and improving the quality of life for people worldwide.

LEADING INSTITUTIONS AND RESEARCHERS IN AI RESEARCH

There are several leading institutions and researchers at the forefront of AI research. Institutions like Stanford University, Massachusetts Institute of Technology (MIT), and University of California, Berkeley, have established themselves as leaders in AI research, producing groundbreaking work in various subfields. Researchers such as Geoffrey Hinton, Yoshua Bengio, and Andrew Ng have made significant contributions to the field and continue to push the boundaries of AI. It is important to stay updated on the work of these institutions and researchers to understand the latest advancements and trends in AI research.

RESOURCES FOR STAYING UPDATED ON AI RESEARCH

As a researcher, it is crucial to stay updated on the latest developments in AI research. Fortunately, there are several resources available to help us stay informed. Academic journals such as "Artificial Intelligence" and "Journal of Machine Learning Research" publish cutting-edge research papers in the field. Online platforms like arXiv.org and OpenAI provide access to preprints and technical reports. Conferences and workshops, such as the Conference on Neural Information Processing Systems (NeurIPS) and the International Conference on Machine Learning (ICML), bring together researchers to share their work and exchange ideas. It is important to leverage these resources to stay at the forefront of AI research.

CONCLUSION

THE IMPORTANCE OF CONTINUED INVESTMENT IN AI RESEARCH FOR BUILDING THE FUTURE

In conclusion, AI research has the power to shape the future and unlock exciting possibilities. From its humble beginnings to its current applications, AI has come a long way. However, there are still challenges to overcome and ethical considerations to address.

Governments, organizations, and researchers must collaborate to support AI research and ensure its responsible and ethical development. By investing in AI research, we can build a future where AI technologies benefit society, improve lives, and drive innovation. Let us embrace the potential of AI research and work towards a future that is built on the foundations of artificial intelligence.

REFERENCES

[1] Ethical Consideration in Artificial Intelligence (Source: Frontiers in Surgery, Volume9,2022)

[2] ZHAO Feng and XU Hua 2022 J. Phys.: Conf. Ser. 2289 012030.

[3] Peter Baumgartner [v1] Tue, 4 Oct 2022 03:27:17 UTC (210 KB)

[4] How AI Transforms Social Media Rem Darbinyam Former Forbes Councils Member Forbes Technology Council

[5] J. R, S. H. Krishna, M. G.M, S. Mohammed, K. B. Raj and G. Manoharan, "Fuzzy Evaluation Method on the Financing Efficiency of Small and Medium-Sized Enterprises," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-7, DOI: 10.1109/ICECONF57129.2023.10083731

COMPARATIVE ANALYSIS OF C AND PYTHON FOR NOVICE PROGRAMMERS: EXPLORING FACTORS AFFECTING SUITABILITY AND LEARNING EXPERIENCE

Mr.Vinod Chandran

Associate Professor, Dept. of Computer Science and Applications, Naipunnya School of Management, Cherthala , Kerala

vinodchandranvc@gmail.com

&

Ms. Mary Joseph

Assistant Professor, Dept. of Computer Science and Applications, Naipunnya School of Management Cherthala

maryjoseph91@gmail.com

Abstract

Selecting the right programming language for beginners is crucial for their learning journey and future career prospects in computer programming. This research paper thoroughly investigates whether C or Python is more suitable for novice programmers, considering multiple critical factors. It analyzes aspects like ease of learning, memory efficiency, engagement level, setup process, syntax, and code readability. Additionally, the study examines job market demand, job prospects, and project portability. It evaluates standard library availability, support for data structures, error handling, scalability, and continuity of learning in both languages. Special emphasis is placed on assessing built-in features for real-world applications and relevance to the modern IT industry. By synthesizing these factors, the research aims to offer valuable insights into C and Python's suitability for beginners. The findings are anticipated to guide individuals starting their programming journey, helping them make informed decisions and establish a strong foundation for future endeavours in software development.

KEYWORDS: Novice programmers, Memory efficiency, Learning experience, Skill development, Job market demand, Code readability, Project portability

Introduction

In the ever-evolving landscape of computer programming, the selection of the initial programming language for beginners holds profound significance. The choice between C and Python, two prominent languages with distinctive characteristics, presents a fundamental decision for those embarking on their journey into the world of coding. This research endeavor seeks to comprehensively explore and compare the suitability of C and Python for novice programmers, aiming to provide valuable insights to aid in informed decision-making. By delving into factors such as ease of learning, versatility, community support, industry relevance, and future career prospects, this study aims to illuminate the strengths and weaknesses of each language in catering to the needs and preferences of beginners. Through a thorough examination of these key aspects, this research endeavors to contribute to a deeper understanding of the factors influencing the choice between C and Python, ultimately empowering aspiring programmers to embark on their learning journey with confidence and clarity.

Background

The surge in computer programming's importance has spurred interest among individuals in learning to code. The choice of initial programming language significantly impacts beginners' learning journeys and career trajectories. C and Python, widely recognized languages, offer distinct features; C is renowned for its efficiency and foundational role, while Python is favored for its simplicity and versatility. This research delves into comparing C and Python's suitability for novice programmers, considering factors like ease of learning, versatility, community support, industry relevance, and career prospects. Understanding the strengths and weaknesses of each language is pivotal for beginners to align their choices with their goals. By examining these factors, this study aims to empower aspiring programmers with valuable insights and inform educational institutions and industry stakeholders on optimizing programming education and training initiatives.

Review of Literature

The comparative analysis of C and Python as introductory programming languages has been extensively explored in the literature, offering valuable insights into their suitability and learning experiences for novice programmers. Kafura and Moore (2018) investigated the effectiveness of C and Python in introductory programming, highlighting their differences in syntax, readability, and ease of use. Robins and Rountree (2014) conducted a comparative analysis focusing on teaching introductory programming, emphasizing Python's simplicity and C's performance-oriented nature. Orsini, McShea, and Zelesnik (2019) explored Python and C as introductory languages in computer science courses, shedding light on their application in academia. Monnappa and Rao (2016) provided a comprehensive comparison of Python and C, examining their features, performance, and application domains.

In addition to academic research, instructional materials such as Shaw's "Learn C the Hard Way" and Gries and Campbell's "The Science of Programming" offer practical exercises and theoretical foundations for learning C programming. These resources provide valuable insights into C's intricacies and its role in computational subjects. Conversely, VanderPlas's "Python Data Science Handbook" equips readers with essential tools for working with data using Python, emphasizing Python's versatility and relevance in data science applications. Van Rossum's "Python Tutorial" serves as a foundational resource for learning Python, offering guidance on syntax, data structures, and programming paradigms.

Overall, the literature review highlights the diverse perspectives and comprehensive analyses of C and Python as introductory programming languages. While C is renowned for its performance and low-level control, Python excels in simplicity, readability, and versatility, making it an attractive choice for beginners and professionals alike. By synthesizing findings from academic research and instructional materials, this review provides a robust foundation for understanding the factors influencing the suitability and learning experiences of novice programmers in C and Python.

Hypothesis

H0: There is no difference in the effectiveness of learning computer programming for beginners between C and Python.

H1: Either C or Python is better than the other for beginners in learning computer programming.

Objective of the Study

1. To determine which language, C or Python, offers a more accessible learning experience for novice programmers by analyzing factors such as syntax complexity and availability of learning resources.
2. To identify which language, C or Python, provides a more conducive environment for novice programmers to build and deploy projects effectively in real-world scenarios by assessing factors such as memory efficiency and readability of code.
3. To investigate the job market demand and career prospects associated with proficiency in C and Python to guide individuals in making informed decisions regarding their programming language choice for long-term career growth in the modern IT industry.

Research Methodology, Data Collection and Interpretation

The research was undertaken among undergraduate college students majoring in Computer Science /Computer Applications at the University of Kerala who had experience with both C and Python programming. The study took place between December 2023 and January 2024. A total of 139 respondents were selected through simple random sampling, and their responses were gathered via a structured questionnaire. Data entry and analysis, including the utilization of Simple Percentage and Chi-square tests, were conducted using Microsoft Office Excel.

Table-1: Easiness of Use

Language	Population N = 138	Percentage
	Frequency	
C	57	41.3%
Python	81	58.7%

According to Table 1, among the 138 respondents, 58.7% expressed that Python is more user-easier to use compared to C, while merely 41.3% favored C as a preferable option.

Table-2: Memory Efficiency

Language	Population N = 138	Percentage
	Frequency	
C	42	30.4%
Python	96	69.6%

Based on Table 2, out of the 138 respondents, 69.6% indicated that Python is more memory efficient than C, with only 30.4% favoring C as the preferable option.

Table- 3: Development of Skills.

Skills	C	Python	
Problem-solving	66	72	p value=0.0000016
Algorithmic thinking	72	66	
Code efficiency	30	108	
Debugging	45	93	

The chi-square test was employed to examine the correlation between C and Python regarding different skills. The calculated p-value was 0.00000016, indicating strong statistical significance. This result supports the acceptance of the alternative hypothesis, suggesting a difference in the effectiveness of learning computer programming for beginners between C and Python. Based on the statistics in Table 3, it is evident that Python is more effective than C in skill development.

Table- 5: Level of Engagement

Language	Population N = 138	Percentage
	Frequency	
C	60	43.5%
Python	78	56.5%

According to Table 5, out of the 138 respondents, 56.5% indicated that Python, as opposed to C, provided a higher level of engagement for programmers, while only 43.5% favored C as the preferable option.

Table- 6: Ease of setup process for beginners

Language	Population N = 138	Percentage
	Frequency	
C	36	26.1%
Python	27	19.6%
Both Equally	75	54.3%

Based on Table 6, regarding the ease of setup process, out of the 138 respondents, 19.6% preferred Python, 26.1% preferred C, and 54.3% favored both. Hence, there is little disparity in the ease of setup process between C and Python.

Table- 7: Beginner-Friendliness in Language Syntax

Language	Population N = 138	Percentage
	Frequency	
C	30	21.7%
Python	72	52.2%
Both Equally	36	26.1%

According to Table 7, concerning beginner-friendliness in language syntax, among the 138 respondents, 52.2% favored Python, 21.7% favored C, and 26.1% favored both. Consequently, Python exhibits superior beginner-friendly features.

Table- 8: Readability of code

Language	Population N = 138	Percentage
	Frequency	
C	21	15.2%
Python	72	52.2%
Both Equally	45	32.6%

Based on Table 8, regarding code readability, out of the 138 respondents, 52.2% preferred Python, 15.2% preferred C, and 32.6% favored both. Consequently, Python demonstrates superior code readability.

Table- 9: Demand for skills in Job market

Language	Population N = 138	Percentage
	Frequency	
	C	
Low	33	23.9%
High	21	15.2%
Moderate	84	60.9%
	Python	
Low	6	4.3%
High	114	82.6%
Moderate	18	13.0%

According to Table 9, among the 138 respondents, 82.6% signaled a strong demand for Python skills in the job market, as opposed to C skills, with only 15.2% favoring C as the preferable option.

Table- 10: Better portability for projects

Language	Population N = 138	Percentage
	Frequency	
C	6	4.3%
Python	111	80.4%
Both Equally	21	15.2%

Based on Table 10, out of the 138 respondents, 80.4% expressed that the superior portability of projects developed in Python is preferable compared to C, with only 4.3% favoring C and 12.2% supporting both equally.

Table- 11: Language's library for support for various Data Structures and their associated operations

Language	Population N = 138	Percentage
	Frequency	
C	9	6.5%
Python	99	71.7%
Both Equally	30	21.7%

According to Table 11, among the 138 respondents, 71.7% indicated that Python provides better library support for various data structures and their associated operations compared to C, with only 6.5% favoring C and 21.7% supporting both equally.

Table- 12: Beginner-friendly approach to handling errors

Language	Population N = 138	Percentage
	Frequency	
C	30	21.7%
Python	75	54.3%
Both Equally	33	23.9%

Based on Table 12, among the 138 respondents, 54.3% indicated that Python offers a better beginner-friendly approach to handling errors compared to C, with only 21.7% favoring C and 23.9% supporting both equally.

Table- 12: Support for learning and mastering Data Structures and Algorithms

Language	Population N = 138	Percentage
	Frequency	
C	45	32.6%
Python	36	26.1%
Both Equally	51	37.0%

According to Table 13, out of the 138 respondents, 37% equally support both Python and C for learning and mastering Data Structures and Algorithms. 32.6% indicated a preference for C, while 26.1% favored Python.

Table- 13: Availability of built-in features for beginners to rapid development of real-world applications

Language	Population N = 138	Percentage
	Frequency	
C	6	4.3%
Python	90	65.2%
Both Equally	39	28.3%

Based on Table 13, out of the 138 respondents, 65.2% noted that Python offers greater availability of built-in features for beginners to rapidly develop real-world applications compared to C, with only a small 4.3% favoring C and 28.3% supporting both equally.

Table- 14: Recommendation of Programming Language considering all aspects of the learning process

Language	Population N = 138	Percentage
	Frequency	
C	42	30.4%
Python	57	41.3%
Equally	39	28.3%

Based on Table 14, out of the 138 respondents, 41.3% recommended Python as the preferable choice for beginners as a programming language, considering all aspects of the learning process compared to C, with 30.4% favoring C and 28.3% supporting both equally.

Findings

Based on the analysis of the survey data, the findings of the research paper "Comparative Analysis of C and Python for Novice Programmers: Exploring Factors Affecting Suitability and Learning Experience" are as follows:

1. ***Ease of Use***: Among the 138 respondents, 58.7% found Python to be more user-easier to use compared to C, indicating a preference for Python in terms of usability.

2. ***Memory Efficiency***: A significant majority of respondents (69.6%) indicated that Python is more memory efficient than C, suggesting that Python consumes fewer resources for memory management tasks.
3. ***Skill Development***: The chi-square test revealed a strong statistical significance (p-value = 0.00000016), supporting the hypothesis that Python is more effective than C in skill development for novice programmers.
4. ***Engagement***: A higher level of engagement was reported by 56.5% of respondents for Python compared to C, indicating that Python provides a more engaging programming experience.
5. ***Ease of Setup***: While 19.6% preferred Python and 26.1% preferred C for the setup process, a majority (54.3%) favored both equally, suggesting little disparity in the ease of setup between the two languages.
6. ***Beginner-Friendliness***: Python was favored by 52.2% of respondents for its beginner-friendly features, compared to 21.7% for C, further highlighting Python's suitability for novice programmers.
7. ***Code Readability***: Python demonstrated superior code readability, with 52.2% of respondents preferring Python over C (15.2%).
8. ***Job Market Demand***: A strong demand for Python skills (82.6%) was observed in the job market compared to C skills (15.2%), indicating better career prospects for Python programmers.
9. ***Project Portability:*** Python was favored by 80.4% of respondents for the superior portability of projects, compared to only 4.3% for C.
10. ***Library Support:*** Python provides better library support for various data structures and operations, as indicated by 71.7% of respondents, compared to only 6.5% for C.
11. ***Error Handling***: Python's beginner-friendly approach to error handling was preferred by 54.3% of respondents, compared to 21.7% for C.
12. ***Learning Data Structures and Algorithms***: While 37% equally supported both Python and C, Python was favored by 26.1% of respondents, compared to 32.6% for C.
13. ***Availability of Built-in Features***: Python offers greater availability of built-in features for rapid development of real-world applications, according to 65.2% of respondents.
14. ***Recommendation for Beginners***: 41.3% of respondents recommended Python as the preferable choice for beginners, considering all aspects of the learning process, compared to 30.4% for C.

These findings collectively suggest that Python is more suitable for novice programmers than C, considering various factors affecting the learning experience and future career prospects.

Conclusion

In conclusion, this research comprehensively explored and compared the suitability of C and Python for novice programmers, considering multiple critical factors affecting the learning experience and future career prospects. The findings reveal significant advantages of Python over C in various dimensions. Python emerged as the preferred choice due to its ease of use, memory efficiency, higher level of engagement, beginner-friendliness, superior code readability, and strong demand in the job market. Moreover, Python demonstrated better support for project portability, library availability, error handling, and learning data structures and algorithms. These findings underscore Python's dominance as an introductory programming language, offering a more conducive environment for novice programmers to build skills and deploy projects effectively in real-world scenarios. The statistical analysis further supported Python's superiority in skill development compared to C. Ultimately, this research provides valuable insights for individuals starting their programming journey, enabling them to make informed decisions and establish a strong foundation for future endeavors in software development.

References

1. Kafura, D., & Moore, D. (2018). "A comparative analysis of C and Python as introductory programming languages." Journal of Computing Sciences in Colleges, 33(2), 175-184.
2. Robins, A., & Rountree, J. (2014). "A comparative analysis of Python and C for teaching introductory computer programming." In Proceedings of the 45th ACM technical symposium on Computer science education (pp. 481-486).
3. Orsini, A., McShea, R., & Zelesnik, G. (2019). "A comparative analysis o**Vol-15 No-1 Special Issue-03 (April 2024) UGC CARE Journals**f Python and C as introductory programming languages in computer science courses." In Proceedings of the 50th ACM technical symposium on Computer science education (pp. 282-288).
4. Monnappa, A., & Rao, B. (2016). "A comparative analysis of Python and C programming languages." International Journal of Computer Applications, 144(1), 1-4.
5. Shaw, Zed A. (2017). "Learn C the Hard Way: Practical Exercises on the Computational Subjects You Keep Avoiding (Like C)." Addison-Wesley Professional.
6. Van Rossum, Guido. (1995). "Python Tutorial."
7. Gries, Paul and Campbell, Jennifer. (2012). "The Science of Programming." Springer.

"A COMPARATIVE ANALYSIS ON PROCEDURE ORIENTED PROGRAMMING AND OBJECT-ORIENTED PROGRAMMING"

Ms.Mary P K

Assistant Professor

Department of Computer Science and Applications

Naipunnya School of Management,

Cherthala, Kerala

University of Kerala

marypk88@gmail.com

Ms. Kavitha Santhosh

Assistant Professor

Department of Computer Science and Applications

Naipunnya School of Management

Cherthala, Kerala

University of Kerala

kavithasarath27@gmail.com

Abstract

The meaning of the term object oriented is examined in the context of the general – purpose programming language C++. This language supports data abstraction, object –oriented programming and traditional programming techniques. This language is discussed eight paradigms are discussed: Procedural, data hiding, data abstraction, Inheritance, polymorphism, Message passing and dynamic binding. It is a methodology which is essentially centered on the way object collaborate to convey and share the data.

Keywords— C++, Inheritance, polymorphism, data abstraction, dynamic binding.

I. INTRODUCTION

OOP is a programming paradigm based on the concept of objects, which may contain data, in the form of fields often known as attributes, and code in the form of procedures often known as methods.

In OOP, computer programs are designed by making them out of objects that interact with one another. There is a significant diversity of OOP languages but the most popular ones are class – based, meaning that objects are instances of classes which typically also determine their type.

The most widely used programming languages (such as C++, object Pascal, Java, Python etc) are multi – paradigm programming languages that support OOP to a greater or lesser degree, typically in combination with imperative, procedural programming. A programming model is the primary genre of computer programming, and they vary in the way different elements of the program are represented and how steps for fixing obstacles are defined. There are several programming languages but two of the most important methods are Object Oriented Programming and Procedural Programming languages.

Therefore, this comparison seeks to compare and contrast Procedural and Object-Oriented Programming Languages. This comparison will first discuss about each programming language separately, then later give a brief explanation for each one of them and how they are similar.

Object-Oriented Programming was developed out of the need to write the logic instead of how to define the data. Object Oriented Programming or OOP is a programming language that uses the concept of classes and objects to construct models based on the real world surrounding. An object is defined as a composition of nouns like strings, variables or numbers and verbs like functions, at the same time an object is a constituent of a program that recognizes how to execute certain actions and how to interrelate with other elements of the program. Objects are the foundation of object-oriented programming. An object-oriented program uses a set of objects, which will communicate by sending and receiving messages to request services or information. A class is a collection of objects with similar properties and behaviours. A method in Object-Oriented language is like a procedure in Procedural Language.

Finally, an object or a collection of objects attempts to complete its goals (goals such as displaying 'hello world' on to the screen) by communicating by swapping messages. In fact, displaying 'Hello World' is a method. The popularity of Object-Oriented Programming was mainly because of its simplicity in writing code, and easy for programmers to understand. Its adoption by many software development companies was because of its power and simplicity. Some examples for Object-Oriented Programming languages include Java, C#.NET, C++, Python and Perl.

[1] Procedural Programming focuses mainly on the actions or procedures that will take place within the program; therefore, Procedural Programming languages follow a sequence of instructions and conveys it to the computer. Procedural programming depends on procedures. As procedural programming language follows a method of solving problems from the top of the code to the bottom of the code, if a change is required to the program, the developer has to change every line of code that links to the main or the original code.

If the user wants to code a program, they would have to follow a sequence of instructions and thereby enter the instructions. In addition, when a problem needs to be fixed using procedural programming, the developer will be required to start with the problem and then he logically fragment the problem down into sub problems. Subsequently, this process will continue until a sub-procedure is simple enough to be solved by itself. Examples for procedural programming languages include C, COBOL, FORTRAN and VB.

In view of the differences between Object-Oriented Programming and Procedural Programming it is obvious that Object-Oriented Programming is based on objects and classes while Procedural Programming is based on procedures. Using objects in Object-Oriented Programming rather than procedures as in Procedural Programming allow the developers to reuse a single code anywhere as needed. Thus, allowing coding methods that are more complicated with ease and using less code.

When we consider about the security of the data when using either of the programming languages, Object-Oriented Programming provides more security as it has a more improved data concealing mechanism rather than Procedural Programming Languages. Procedural Programming uses global data for sharing data within functions therefore data can be accessed from function to function without any access limits. However, Object-Oriented Programming does not allow global data but instead the developer has the ability to set the functions to private or public so developers can control the access rights for data.

In Procedural Programming, it is quite difficult to add new data or functions to the program but Object-Oriented Programming offers an easy approach to add new data and functions. Additionally, in Procedural Programming data cannot be moved liberally from function to function but Object-Oriented Programming allow objects to move and communicate with each other via member functions.

OOPs History

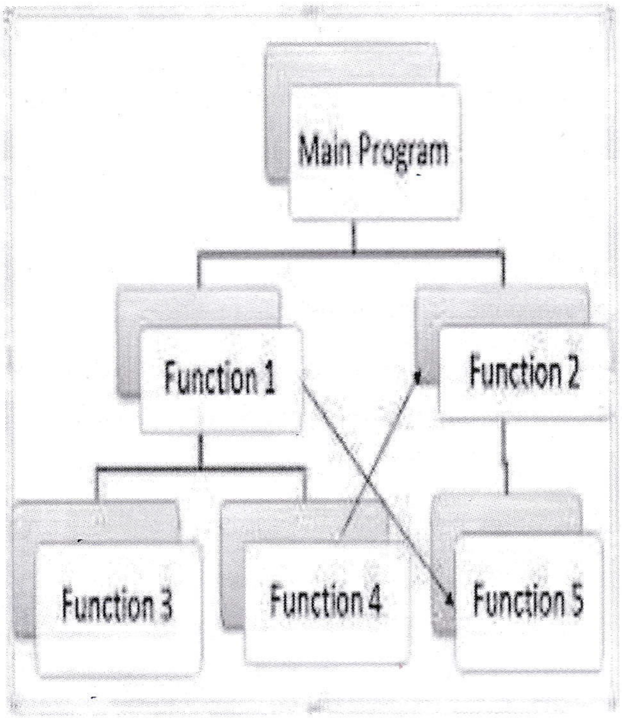
[2] The idea of object – oriented programming gained momentum in the 1970s and in the early 1980s Bjorn stroustrup integrated object-oriented programming into the C language. The resulting language was called C++ and it became the first object oriented language to be widely used commercially.

In the early 1990s a group at sun led by James Gosling developed a simpler version of C++ called Java that was meant to be a programming language for video – on – demand applications.

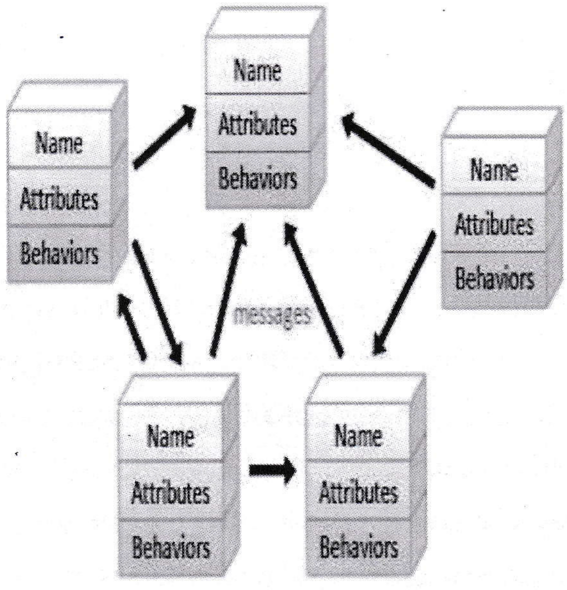
II. COMPARISON BETWEEN PROCEDURE ORIENTED AND OOPS

	Procedure Oriented Programming	Object oriented Programming
Divided into	In POP, program is divided into small parts called functions.	In OOP, program is divided into parts called objects.
Importance	In POP, importance is not given to data but to functions as well as sequence of actions to be done.	In OOP, importance is given to the data rather than procedures or functions because it works as a real world.
Approach	POP follows top down approach.	OOP follows bottom up approach.
Access specifies	POP does not have any access specifier.	OOP has access specifier named public, private, protected etc.
Data Moving	In POP, data can move freely from function to function in the system.	In OOP, objects can move and communicate with each other through member functions.
Expansion	To add a new data and function in POP is not so easy.	OOP provides an easy way to add new data and functions.
Data Access	In POP, most functions use global data for sharing that can be accessed freely from function to function in the system.	In OOP, data cannot move easily from function to function, it can be kept public or private so we can control the access of data.
Data Hiding	POP does not have any proper way for hiding data so it is less secure	OOP provides data hiding so provides more security.
Overloading	In POP, overloading is not possible.	In OOP, overloading is possible in the form of function overloading and operator overloading.

STRUCTURE OF PROCEDURE ORIENTED DIAGRAM



STRUCTURE OF OBJECT ORIENTED DIAGRAM



An object-oriented program consists of many well-encapsulated objects and interacting with each other by sending messages

FEATURES OF OOPS

- Problems are divided into objects
- It is not possible to access data freely.
- Data hiding is possible.
- It uses bottom – up programming technique.
- It is easy to add new data and functions.
- Object can exchange data through its functions.

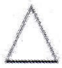
III. BASIC CONCEPTS OF OOPS

Object oriented program development is a new programming style having real world thinking. It is not a programming technique. So each and every programmer has its own way of thinking. To standardize the thinking the following concepts are defined.

1. Objects
2. Classes
3. Data Abstraction
4. Data Encapsulation
5. Inheritance
6. Polymorphism
7. Dynamic Binding
8. Message Passing

1. OBJECT

[1] Objects are the basic unit of OOP. They are instances of class, which have data members and uses various member functions to perform tasks. It is defined as an entity. It may be either physical or logical. It is a basic unit of Object-Oriented Programming and represents the real-life entities. An Object is an instance of a Class. When a class is defined, no memory is allocated but when it is instantiated (i.e. an object is created) memory is allocated. An object has an identity, state, and behavior. Each object contains data and code to manipulate the data. Objects can interact without having to know details of each other's data or code, it is sufficient to know the type of message accepted and type of response returned by the objects. For example, "Dog" is a real-life Object, which has some characteristics like color, Breed, Bark, Sleep, and Eats.

Object	Data (Attributes)	Functions
Triangle 	Side a Side b Side c Border color Fill color	Draw Fill Area Move

While representing objects in computer it occupies some memory space and have an associated address. During execution objects can exchange information.

2. CLASS

[1] It is similar to structures in c languages. It is defined as a collection of objects with same type of data and functions. It is defined as used defined data type but it also contains functions in it. It declares and defines what data variables the object will have and what operations can be performed on the class's object.

Class classname

{

Declaration of data

……………………

……………………. Definition of functions

……………………

……………………

}

3. DATA ABSTRACTION

Abstraction refers to showing only the essential details and hiding the other details. It is a named collection of data that describes a data object in a class. Class is also called the abstract data type because classes use the principle of data abstraction. Classes can provide methods to the outside world to access. This can be done using access specifier.

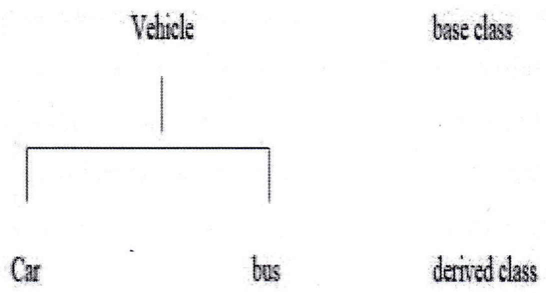
Consider a real-life example of a man driving a car. The man only knows that pressing the accelerators will increase the speed of the car or applying brakes will stop the car, but he does not know about how on pressing the accelerator the speed is increasing, he does not know about the inner mechanism of the car or the implementation of the accelerator, brakes, etc in the car. This is what abstraction is.

4. ENCAPSULATION

It can also be said data binding. It is all about binding the data variables and functions together in class. It is a technique used to protect the information in an object from other objects. The other objects and programs cannot access the data in an object directly. This concept is called data encapsulation or data hiding. Consider a real-life example of encapsulation, in a company, there are different sections like the accounts section, finance section, sales section, etc. The finance section handles all the financial transactions and keeps records of all the data related to finance. Similarly, the sales section handles all the sales-related activities and keeps records of all the sales. Now there may arise a situation when for some reason an official from the finance section needs all the data about sales in a particular month. In this case, he is not allowed to directly access the data of the sales section. He will first have to contact some other officer in the sales section and then request him to give the particular data. This is what encapsulation is. Here the data of the sales section and the employees that can manipulate them are wrapped under a single name "sales section".

5. INHERITANCE

Inheritance is an important pillar of OOP(Object-Oriented Programming). The capability of a class to derive properties and characteristics from another class is called Inheritance. When we write a class, we inherit properties from other classes. So when we create a class, we do not need to write all the properties and functions again and again, as these can be inherited from another class that possesses it. Inheritance allows the user to reuse the code whenever possible and reduce its redundancy. It is the process of creating a new classes from the existing classes. The new classes are called derived classes. The existing classes are called base classes. It is basically a method which provides a way that capabilities and properties from one class to come into another class. This technique provides code reusability to the programmers. We can form a new class from an existing class, where the existing class contains some properties or methods that also exist in the new class. Here the new class is called the derived class , and the existing class that is the class from which the new class is derived is called as base class.



6. POLYMORPHISM

The word [11] polymorphism means having many forms. In simple words, we can define polymorphism as the ability of a message to be displayed in more than one form. For example, A person at the same time can have different characteristics. Like a man at the same time is a father, a husband, an employee. So, the same person possesses different behaviour in different situations. This is called polymorphism. It is a technique used to write more than one function definition with same function name. The functions may be in the same class or in different derived classes. Polymorphism is the ability for a message or data to be processed in more than one form. It is an important concept of object-oriented programming which supports the capability of an object of a class to behave differently in response to a message or action.

7. DYNAMIC BINDING

[7] Binding is defined as the connection between the function call and its corresponding program code to be executed. It is also known as late binding. It is generally use with polymorphism and inheritance.

There are two ways of binding. They are

i) Static binding

ii) Dynamic binding

In static binding, the binding occurs during compilation time. In dynamic binding, the binding occurs during run time.

8. MESSAGE COMMUNICATION

[6] Message Passing in terms of computers is communication between processes. It is a form of communication used in object-oriented programming as well as parallel programming. Message passing in Java is like sending an object i.e. message from one thread to another thread. It is used when threads do not have shared memory and are unable to share monitors or semaphores or any other shared variables to communicate. Suppose we consider an example of producer and consumer, likewise what producer will produce, the consumer will be able to

consume that only. We mostly use Queue to implement communication between threads. It is defined as a process of sending request to execute a function for an object. The general form is Objectname. Message(information);

Objects can communicate with each other's by passing message same as people passing message with each other.

Comparative analysis

Approach to Problem-Solving:

One of the key differences between Object-Oriented Programming (OOP) and Procedure-Oriented Programming (POP) lies in their approach to problem-solving. OOP focuses on modelling real-world entities as objects, which interact with each other through methods and message passing. This approach emphasizes encapsulation, inheritance, and polymorphism, enabling developers to create modular, reusable, and extensible systems. In contrast, POP follows a more procedural approach, dividing programs into a series of procedures or functions that manipulate data step by step. While POP promotes structured and sequential programming, it may lead to less modular and reusable code compared to OOP.

Reusability and Modularity:

OOP promotes reusability and modularity through the use of classes and objects. By encapsulating data and behavior within objects, developers can reuse and extend existing code without modifying its underlying implementation. Inheritance allows classes to inherit properties and methods from base classes, further enhancing code reuse. Additionally, OOP supports modularity by organizing code into classes, each responsible for a specific set of functionalities. In contrast, POP may lack the same level of reusability and modularity, as procedures are typically designed for specific tasks and may not be easily adaptable to other contexts without modification.

Data Abstraction and Encapsulation:

OOP emphasizes data abstraction and encapsulation, allowing developers to model real-world entities as abstract data types with well-defined interfaces. Encapsulation hides the internal state of objects, exposing only the necessary operations to manipulate their data. This promotes information hiding and reduces coupling between components, making code more maintainable and extensible. In contrast, while POP also supports data abstraction to some extent through procedural abstraction, it may not provide the same level of encapsulation as OOP. Procedures in POP typically manipulate data directly, exposing implementation details and potentially leading to tighter coupling between components.

Inheritance vs. Procedural Abstraction:

Inheritance is a fundamental feature of Object- Oriented Programming (OOP), allowing classes to inherit properties and behaviors from other classes. This promotes code reuse and hierarchical organization, *enabling developers to create specialized classes based on existing* ones. In contrast, Procedure-Oriented Programming (POP) relies on procedural abstraction, where procedures encapsulate sets of instructions to perform specific tasks. While both paradigms support abstraction, inheritance in OOP offers a more structured and hierarchical approach to code organization compared to procedural abstraction in POP.

Scalability and Maintainability:

[7]Object-Oriented Programming (OOP) is often praised for its scalability and maintainability, as it promotes code reuse, modularity, and encapsulation. By organizing code into classes and objects, developers can easily extend and modify existing systems without affecting other components. Inheritance and polymorphism further enhance scalability, allowing for the creation of complex hierarchies and flexible architectures. In contrast, while Procedure-Oriented Programming (POP) may be suitable for smaller, less complex projects, it may lack the scalability and maintainability of OOP due to its procedural nature. As programs grow in size and complexity, maintaining and extending procedural codebases can become increasingly challenging.

Performance Considerations:

When comparing Object-Oriented Programming (OOP) and Procedure-Oriented Programming (POP) in terms of performance, it's essential to consider various factors such as execution speed, memory usage, and resource efficiency. While OOP may introduce some overhead due to features like dynamic dispatch and object instantiation, modern OOP languages and compilers have become increasingly optimized to minimize these costs. Additionally, the benefits of code reuse, modularity, and maintainability offered by OOP may outweigh any potential performance drawbacks in many cases. In contrast, POP may offer more predictable performance characteristics, as procedural code tends to be more straightforward and explicit. However, this simplicity may come at the cost of flexibility and extensibility, particularly in larger and more complex systems.

Real-World Applications

Object-Oriented Programming in Software Development:

[8]Object-Oriented Programming (OOP) is widely used in software development for various applications, including web development, mobile app development, game development, and enterprise systems. OOP promotes code reuse, modularity, and extensibility, making it well-

suited for building complex and scalable systems. Popular OOP languages such as Java, C++, and Python are widely used in industry for developing a wide range of software applications.

Procedure-Oriented Programming in Software Development:

While Procedure-Oriented Programming (POP) is less prevalent in modern software development compared to OOP, it still has its place in certain domains and applications. POP is often used in embedded systems, scientific computing, system programming, and legacy codebases where performance and predictability are paramount. Languages like C, Fortran, and Pascal are still used today for developing low-level and performance-critical software.

The choice between OOP and POP depends on various factors, including project requirements, team expertise, performance considerations, and scalability needs. While OOP may be more suitable for large, complex projects that require flexibility and maintainability, POP may be more appropriate for smaller, performance-critical applications where simplicity and control are prioritized. Ultimately, developers and organizations should carefully evaluate these factors and choose the programming paradigm that best aligns with their goals and constraints.

This paper presented a comprehensive comparative analysis between Object-Oriented Programming (OOP) and Procedure-Oriented Programming (POP). OOP emphasizes code reuse, modularity, and extensibility through classes, objects, inheritance, and polymorphism. POP, on the other hand, focuses on procedural abstraction, simplicity, and control through procedures, functions, variables, and control structures. By examining their fundamental principles, features, advantages, disadvantages, and real-world applications, we highlighted the strengths and weaknesses of each paradigm. Procedure-Oriented Programming (POP) emphasizes functions and procedural flow, offering simplicity and efficiency for smaller projects. Object-Oriented Programming (OOP), with its focus on encapsulation, modularity, and scalability, is better suited for larger, more complex systems.

IV CONCLUSIONS AND FUTURE SCOPE

In conclusion the purpose of this comparison was to compare Object-Oriented Programming and Procedural Programming with adequate descriptions for each programming languages. Firstly, the comparison looked into what Object-Oriented Programming is and the comparison defined what Object-Oriented Programming is. Furthermore, the comparison briefly explained what objects, classes and methods are and how they cooperate with each other to make Object-Oriented Programming work. Thereafter, the comparison showed what Procedural Programming is and clearly defined it. It further explained what procedures are, how procedures are used and how procedures work.

OOPS classes tend to be in generalized form which make relations among classes becomes artificial at times. The object-oriented programs are tricky in design. So to program with OOPS one needs to have proper design skills, programming skills. With the use of feature like class, objects, encapsulations, polymorphism, inheritance, and abstraction it can be seen that development of software is increase by using these capabilities.

Finally, the comparison differentiated Object-Oriented Programming and Procedural Programming Languages and the main differences between them. Which concluded that Object-Oriented Programming is easier to use, more secure and efficient than Procedural Programming language.

REFERENCES

[1] Laimek, R. and W. Pawgasame. Internal ballistics simulation based on object oriented programming. in Defence Technology (ACDT), 2015 Asian Conference on. 2015.

[2] Neyer, A., F.F. Wu, and K. Imhof, Object-oriented programming for flexible software: example of a load flow. Power Systems, IEEE Transactions on, 1990. 5(3): p. 689-696.

[3] Longshu, L. and X. Yi. The teaching research on a case of object-oriented programming. in Computer Science and Education (ICCSE), 2010 5th International Conference on. 2010.

[4] Dewar, R. and F. Gasperoni. Safety and OOP. in System Safety, 2006. The 1st Institution of Engineering and Technology International Conference on. 2006.

[5] Lei, Z. Study on comparison of AOP and OOP. in Computer Science and Service System (CSSS), 2011 International Conference on. 2011.

[6] de Lima, C.H.M., et al. A software development framework based on C++ OOP language for link-level simulation tools. in Telecommunications Symposium, 2006 International. 2006.

[7] kafura.Object-Oriented-Programmingand-Software,Engineering.1996, http://people.cs.vt.edu/kafura/cs2704/oop.swe.html.

[8] Procedure Oriented Programming(POP) vs Object Oriented Programming(OOP). 2011.

[9] Sasso, D & Biles, W.E (2008). An Object-Oriented Programming Approach For A GIS Data-Driven Simulation Model Of Traffic On An Inland Waterway. Proceedings of the 2008 Winter Simulation Conference

[10] Rentsch, T. Mathematics Education: Taking a Clue From the Recent Technological Revolution -A Very Brief History of Object-Oriented Computing. Accessed July 10, 2016. http://www.cut-the- knot.org/Mset99/OOhistory.shtml

[11] Exforsys. The History of Object Oriented Programming. http://www.exforsys.com/tutorials/oopsconcepts/the-history-of-object-oriented-programming.html

[12] TechTerm (2007). OOP Definition. Accessed July 11, 2016 http://techterms.com/definition/oop .

[13] Egenhofer, Max. J & Frank, Andrew. U (1992). Object-Oriented Modeling for GIS. URISA journal 4 (2): 3-19

[14] Nino J., Hosch, (2001): Programming and Object Oriented Design Using Java. [15]. Wiley, Stroustrup, B.(1994.): The Design and Evolution of C++.Addison-Wesley.

[15] Brosgol, M. (1998). A comparison of the concurrency and real-time features of Ada 95 and Java. 1998 Proceedings of Association for Computing Machinery.

[16] Cary, J. R. & Shasharina S. G.. Comparison of C++ and FORTRAN

[17] 90 for object-oriented scientific programming. Computer Physics Communications, Nov. Issue.

[18] Malik, D. S. C++ programming: Program design including data structures. Boston: Course Technology.

[19] Morrison, R., . Persistent Programming: A Programming Language and Database Integration Technology. Proceedings of the Newcastle International Seminars, P 34-56.

[20] Mosli, S. Comparison of C++, FORTRAN 90 and Oberon-2 for scientific programming. GISI 95, Editors: Friedbert Huber-Waschle, Helmut Schauer, Peter Widmayer Berlin, P 740-748.

[21] Satzinger, J. & Orvik T. U. The object-oriented approach. Boston: Course Technology.

[22] Sintes, A. Teaching yourself object-oriented programming. Indianapolis: SAMS Publishing.

[23] Stotts, P. D. A Comparative survey of concurrent programming languages. ACM SIGPLAN Notice, Vol. 17 Is 10, P. 50-61.

[24] Stroustrup, B. The C++ programming language. Reading: Addison-Wesley.

[25] Stroustrup, B. The design and evolution of C++. Reading: Addison-Wesley.

[26] Taylor, D.. Object-oriented technology: A manager's guide. Reading: Addison-Wesley.

A COMPARATIVE ANALYSIS OF SOCIAL MEDIA PREDICTION USING TEXT MINING APPROACH

Ms. Priyanka. G

Assistant Professor

Assistant Professor Department of Computer Science and Application

Naipunnya School of Management, Cherthala, India

University of Kerala

priyabhuvan.priya@gmail.com

Ms. Gayathri. A

Assistant Professor

Assistant Professor Department of Computer Science and Application

Naipunnya School of Management, Cherthala, India

University of Kerala

gayu.anju89@gmail.com

Abstract

With the advancement of internet technology social media revolutionized the society and the culture by bringing the people under one crown of widespread connected network. E-mails and Instant Messaging (IM) technologies were popular with the advent of internet technology that allowed people to exchange information quickly and efficiently. Changes in technological development compel the analytics industry to formulate procedures that provide companies with data required to make crucial business decisions. People are interacting directly with each other with the outbreak of online social media, commonly known as Web 2.0. Consumers of common interest, group together through social media and share opinions pertaining to a particular topic. Since people express their opinions, interests, likes and dislikes openly in social media, noteworthy opportunities exist for Business Intelligence (BI) in analyzing social media content. The outburst of social media has produced exceptional opportunities for the people, which has revolutionized the way to express their opinions. Now-a-days people not only share the information, they also comment on the interesting information such as Product rating, important decisions, movies, health care, Traffic Analysis, weather, etc. Opinions are mostly articulated in comments or reviews. Many internet users use Microblogging sites such as Twitter, Facebook, LinkedIn, Pinterest, etc., to share their content. That information would be informal descriptions, mostly unstructured and does not follow any language grammar.

Social Media Mining is an emerging field that attracts the researchers to discover new patterns from human generated content.

This research focuses on building a powerful model to extract, analyze and gain insights of such user-generated content from social media websites. The model is built by applying information extraction, NLP and text mining techniques. The primary objective of the research is to construct an influential model to extract, analyze and develop a social media mining model to place emphasis on user-generated content from microblogs. This model helps in identifying the extracted information using NLP and text mining techniques to enhance BI. The secondary objectives of this research work are:

- *To extract the unstructured text data such as opinions and reviews contained in social media.*
- *To apply language processing techniques to convert the unstructured data into an intermediate form suitable for application of core text mining operations.*
- *To study and analyze the existing text mining techniques that is best suited for developing a mode with supervised learning method.*
- *To develop an integrated framework to extract knowledge hidden in unstructured text data such as opinions and reviews contained in social media by applying IE techniques for the enhancement of business decision-making.*
- *To identity an appropriate tool to implement the model which supports text pre-processing, Part-of- Speech (PoS) tagging, feature extraction and text classification.*
- *To collect the domain-specific knowledgebase pertaining to a domain.*

Keywords—Media, Social, Information, Analysis, Text Mining.

I. INTRODUCTION

Traditional means of expressing opinions or reviews was through oral or written medium. In the recent years, because of the growth of electronic media and ease of access to internet technology there is abundant availability of opinions in the electronic form. Subsequently e-mails and other electronic media became popular. One's thoughts, comments, likes, dislikes are shared as their opinions (user-generated content) in various social media sites, which can be used to extract valuable information more objectively. Social media can be understood as a form of knowledge. Enormous amount of information propagates through social media and the number of social media users are growing exponentially, this gives an interesting opportunity to extract the data present, convert into an intermediate form, which enables specific predictions to be made about expected outcomes. The extraction of information hidden in the unstructured text gives an organization a competitive edge hence the knowledge gained from the social media data contributes for effective decision-making [1]

Emotions and opinions are shared or exchanged by an individual in the form of information. The receiver of the information reads it, actively participate in the social media and also contributes additional information, resulting in voluminous data in social media sites. The opinionated information is an important part of textual data, which influences better decision-making [2].

Opinion may be about an individual, a topic, a product or a movie etc. Opinion mining mainly identifies the orientation of the opinion by the person who provides it, by interpreting the features or attributes or components present in the sentence. Opinions can be very vital whenever a decision has to be taken or to choose one from among multiple options. People often depend upon the experiences of their peer groups or friends for such advice. Generally, opinion mining aims to determine the attitude of the user with respect to some topic or the overall contextual polarity in social media data. The attitude of a person is based on the judgment, evaluation, affective state, or the intended emotional communication. Opinion mining helps in overcoming the challenges in accessing the data available in social media sites and transforming them into usable and actionable information. Opinion mining is different from pure data and text mining since opinion mining deals with subjective i.e., personal or individual statements. Opinion mining is a controlled Natural Language Processing (NLP) problem, since it requires understanding of the polarity of the sentiment like whether it is positive or negative and the entities involved. The fundamental task of opinion mining is classification of polarity of an opinion. Polarity classification is the positive or negative sentiment of each sentence, entity or topic under discussion. For example, like, good, thumbs up are all of positive polarity whereas dislike, bad, thumbs down is of negative polarity. Polarity classification helps to make an evaluation trustworthy. It is also very essential to differentiate between subjective and objective description since it helps to classify the sentiment. A sentence or a text is subjective if it indicates a person's feelings while an objective sentence indicates some facts and known information about the world. Since any text or informal description may have a polarity without having any opinion. For example, a news item can be classified as good or bad without being subjective. Opinion mining systems may sometimes serve as a warning system or a recommender system that advices business organization whether to proceed in a given direction or not. Performing a mining operation on user-generated content may provide BI to organizations by extracting significant knowledge concealed in form of opinions. Figure 1 is a architectures of the text mining and figure 2 is process of the text mining.

Opinion mining tasks can be classified into four levels,

i.e., Document level, Sentence level, Phrase level and Feature level.

- Document level task is mainly concerned with arranging a document into a set of predefined classes.
- Sentence level task selects a set of sentences which summarizes the opinion.
- Phrase level opinion mining narrows down on a phrase to arrive at the opinion expressed which is more accurate than document and sentence level.
- Feature level opinion mining is carried out whenever there is a need to look in for a feature or particular attribute rather than the overall feedback or opinion.

II. CHALLENGES IN SOCIAL MEDIA MINING

Social media mining faces many challenges when evaluating text given as opinions [3].

Polarity Scale: Defining a scale for representing the polarity of an opinion is not easy.

Keyword Spotting: Identifying the right set of keywords to correlate the presence of strong or tough opinions or phrases in the sentences can be a difficult task.

Heterogeneous nature of data: The social media data considered for opinion mining may vary widely in style, presentation, content, short forms of words may be used, special characters may be used, and may not follow sentence grammar.

Visualization: Visualization of sentiment information rather than document summarization is complicated since summary should include aggregation, highlighting important opinions, plotting graph depicting points of agreement and disagreement, identification and accounting details of opinion holders.

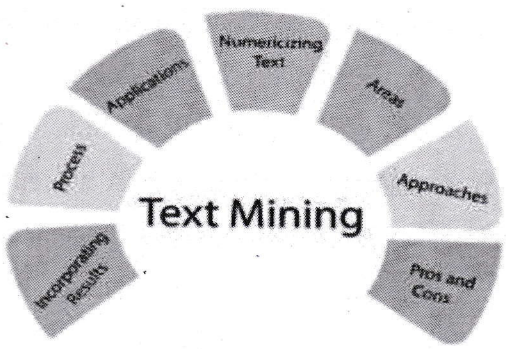


Fig 1: Text mining Architecture

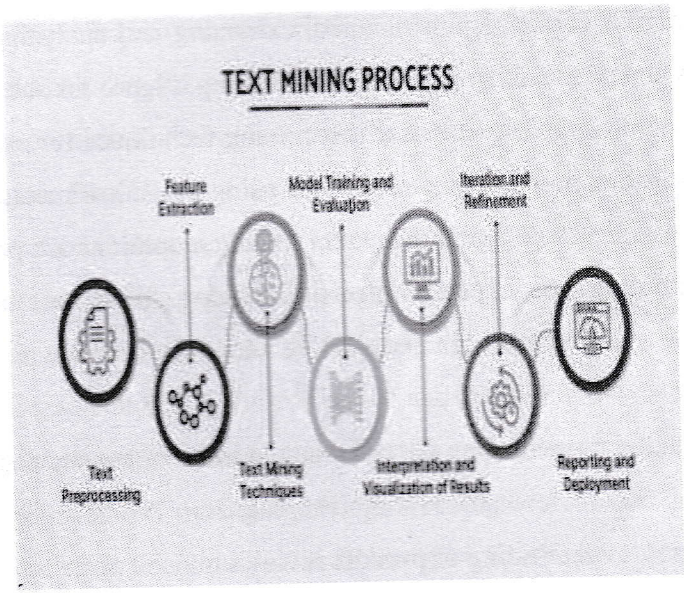


Fig 2: Text minign Process

III. IMPORTANCE OF SOCIAL MEDIA MINING

Generally, organizations use the social media data to understand the needs and behavior of their customers or a specific set of individuals with respect to the organizations current or future products or services. NLP, text analysis and computational linguistics are used collectively to mine opinions and extract information pertaining to a subject contained in the input source i.e., World Wide Web's (WWW) voluminous text data. The focus of sentiment analysis is to classify the polarity of the input text as positive or negative or neutral, at various levels like document, sentence, phrase or feature. Sentiments are classified based on emotions expressed in the sentence like "angry", "sad" or "happy" [4]. Opinion mining focuses on polarity detection and sentiment analysis focuses on recognition of the emotion. Since the identification of the sentiment is required to determine the polarity these two major disciplines are combined.

Social Media Mining refers to the process of representing, analyzing and extracting actionable patterns from social media data. Mining opinions and sentiments from natural language text given by users is challenging. It demands an in-depth understanding of the explicit and implicit rules, syntax and semantics of the language. Some of the basic concepts and primary algorithms suitable for mining social media data are introduced. Social media mining also borrows theories and methodologies from data mining, machine learning, text mining, sociology and statistics. Tools are also used to represent, measure and mine useful patterns.

The discussion on social media mainly pertains to a particular topic and the user generated content is generally semi-structured or unstructured. The data found may vary in formats [5]. Since the user-generated content is unstructured, extracting and analyzing the data is a time-consuming process and this leads to poor decision making. Figure 1.1 shows the framework of Social Media Mining by applying classical text mining techniques for extracting information and converting the information to structured form using the suitable categorization methods. Polarity values are assigned to classify the extent to which opinions are positive and negative. This plays a major role in the impact of decision making. To extract the context sensitive meaning of a given opinion, the grammatical relationship between words are taken into consideration.

A new dimension can be formed using the opinionated information such as:

- Summarizing user reviews: user ratings ranging from 1 to 5 can be suggested to get the opinions from the user. Higher rating of product reviews may be preferred.
- Recommendation systems : A system which recommends products that receives only positive feedback.
- Business and Government intelligence: public viewpoints are monitored for trend predictions; social media data can be used to predict real world outcomes.
- Automatic analysis of politics: analysis of voters' opinion to guess politicians' positions during elections and pending Government policy.

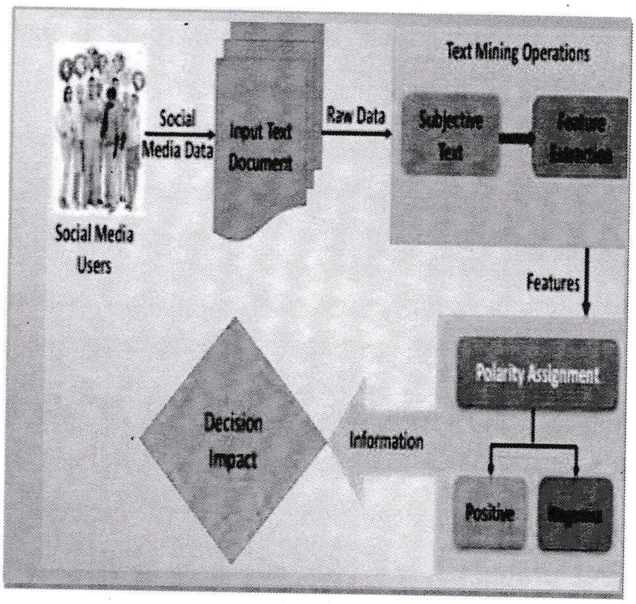


Fig 3: Framework of Social Media Mining

Many real-life applications require comprehensive analysis of information which is gathered from blogs, reviews, feedbacks and comments.

To influence BI and to improve performance, analyzing the data and plan for courses of action that brings about improvement is vital. All findings should be analyzed to determine whether there is a political, cultural or financial bias involved in it. A company may implement a BI solution that provides intuitive access to data. If this data access is not leveraged for decision making and acted upon, then BI has done nothing to improve performance. Huge availability of digital data has resulted in data explosion and information overload.

Traditionally handwritten or data typed on a piece of paper was used to process an order, now the data is available in electronic form. With the growth of electronic media, the volume of data has increased resulting in complexity in processing and analyzing the data. So, there is a need to utilize an intelligent mining technique that results in effective and efficient analysis of the extracted data. A key sign of successful BI is the degree to which it impacts business performance, linking insight to action. When BI is deployed effectively, all that data becomes a strategic asset to be exploited [7].

IV. INFLUENCE OF TEXT MINING IN SOCIAL MEDIA MINING

One of the most important and leading tools for BI is Text Mining [8]. Text mining can be applied to enhance BI in the areas like Scientific Data Analysis, Marketing, identifying potential customers, Market Segmentation, Biomedical sciences, Document warehouse for SAP and Fraud detection. The theory of text mining may seem to be complicated, but understanding the task is easy if the process is broken down step by step. A simplified approach can be adopted that breaks down the overall process into steps that offers insight into how text mining works. Further organizations can identify the specific areas, which best suit their requirements and how to best apply text mining within their current processes. Text mining is generally defined as the process of deriving high quality information from text, especially unstructured text. Text mining extracts useful information from input data by identifying and exploring interesting patterns. Basically, there are four basic types of elements that can be extracted from text:

- Entities are the basic building blocks like people, organization, locations.
- Attributes are features of extracted entities like designation of a person, his age etc.
- Facts are relations that exist between entities like "employment" relationship between a person and an organization.
- Event is an activity in which entities participate.

Common operations in text mining includes application of statistical or linguistic techniques to recognize, tag and extract entities, concepts and relationships in the input data source. Preprocessing operations play a vital role in a text mining system for identification and extraction of representative features and for transforming unstructured data into a structured intermediate format. Preprocessing operations attempts to leverage many key document features like characters, words, terms and concepts. Use of domain or background knowledge in preprocessing operations enhances the extraction of concepts and validation activities. A set of keywords or feature words are selected based on the domain that enhances the concept extraction and validation. Access to background or domain knowledge helps in the development of a more meaningful and consistent concept hierarchy. Background knowledge constraints helps in meaningful knowledge discovery operations. As a result, background knowledge source is important for classification and concept extraction.

Data that is found on social media sites are different from the conventional attribute-value data for classical data mining. Tagging documents for text mining systems begins with application of Part-of-Speech tagging to identify the entities and relationships. Relationships mean facts or events involving two or more entities. Most text mining systems insists us to define in advance the types of semantic information to be extracted from the document.

The social media describes any Website or service that facilitates using a piece of media to share an idea, advertise, promote or deliver content. Media in this sense could be documents, presentations, photos or videos. Social Media is already a powerful source of information transmission on a worldwide scale. The data collected from social media are in the form of semi-structured or unstructured data. The reason that can be attributed for the presence of unstructured text is that human beings have the ability to distinguish and apply various linguistic patterns whereas computers cannot easily handle spelling variations, meaning with reference to a particular context and slang. So unstructured data are for human consumption and are not machine processable. If such data are considered for extracting information, then it leads to poor decision making. So converting unstructured text to intermediate form suitable for further processing text is a challenging task. NLP technique helps to bridge the gap in human and computer language by providing the computer the ability to analyze, understand and generate text [9].

Structured data are in form of records and there is a well- defined data model where transactions are stored, that facilitates query, search and analysis easier. Structured data models are present in form of tables and there exists relations among models. However unstructured text is just plain text that doesn't follow any grammar and is without explicit formatting that makes it very difficult either extract, query, search or analyze. To capture meaningful information from the free form unstructured texts, Information Extraction (IE) is used. IE is a process to generate structured data from unstructured or semi-structured document collection [10]. The input of the IE process can be unstructured documents like free text written in natural language or semi-structured documents, which are commonly found on the internet. The output of the IE process is structured form data, which can be processed automatically by machines.

IE aims to find and understand limited relevant parts of texts and gather information from them, produce a structured representation of relevant information i.e., relation and a knowledge base, organize information so that it is useful to people, present information in a semantically precise form that allows further inferences to be made by computer algorithms [11]. The two main steps of IE are Identification and Classification. Identification is recognizing the given input text as information that is of use and Classification is locating the information into a pre-established semantic category.

V. SOCIETAL CONTRIBUTION

- This research will help to improve efficiency of some social media mining model based on the opinions expressed in social media.
- Predictions can be carried out linking user behavior to the opinions expressed.
- The model can be used to improve the utility of reviews and feedback results based on review rankings that reflect the distribution of opinions.
- Reduce the human effort required to analyze the voluminous content generated by social media users.
- Corpus can be built by accumulating feature words relating to specific domain.

VI. EVALUATION MEASURES

The expected result of an IE system can be defined very accurately by giving an input text data or a collection of text data with the ease of an ideal evaluation approach of IE. After completion of the extraction of information, the question arises is "Did the approach extract the most relevant information or missed some important information?" Also the model should not extract junk information. Unfortunately extracting correct information by avoiding junk information is very difficult. However, it is possible to measure the performance of the information extraction model with respect to these two factors [4].

Precision and Recall: In particular, the measures precision and recall were adopted to evaluate the model implemented. Precision, measures only the appropriate output and Recall, measures to the extent in which the system produces the entire output. Thus, the two measures precision and recall are used to measure completeness and correctness. For the binary classification problem, the confusion matrix must be presented with the predicted results of a classifier with two rows and two columns. The row represents the observed class labels and the column represents the predicted class labels. Each cell represents the number of predictions made by the classifier.

Predicted Class Labels	Observed class labels	
	Positive	Negative
Positive	True Positive (tp)	False Positive (fp)
Negative	False Negative (fn)	True Negative (tn)

Fig 4: Framework of Social Media Mining

$$\text{Precision} = \frac{tp}{tp+fp}$$

$$\text{Recall} = \frac{tp}{tp+fn}$$

F-Measure: The f-measure score is calculated as the weighted harmonic mean of precision and recall. This gives the combined balance score between Precision and Recall. F- Measure is used as the evaluation metric for sentiment classification and is used as follows:

Accuracy: A common measure for evaluating the classification performance is accuracy. Accuracy is the proportion of correctly classified samples to the total number of samples.

$$\text{F-Measure} = 2\,\frac{Precision \times Recall}{Precision + Recall}$$

$$\text{Accuracy} = \frac{tp+tn}{tp+tn+fp+fn}$$

VII. RESULTS AND DISCUSSIONS

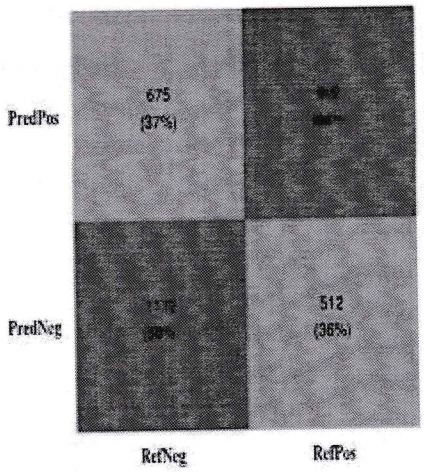


Fig 5: Confusion matrix for the predictive results showing the predicted (Pred) values against the reference tweets (Ref)

Method	Reference			
	Correct rate (%)	Sensitivity (%)	Specificity (%)	Balanced accuracy (%)
Random Forest	71	66	77	72
Decision Tree	65	62	68	65
Logistic	64	63	64	64
Naive Bayes	60	51	70	61

Fig 6: Various results of Social Media Mining

In binary classification problems, such as those related to social media predictions. In such scenarios, predictions can be classified into four categories based on the comparison between the predicted and actual values:

1. True Positive (TP): Instances where the model correctly predicts the positive class. For social media predictions, it could be correctly identifying a post as relevant or trending when it actually is.
2. True Negative (TN): Instances where the model correctly predicts the negative class. In social media, this could be correctly identifying a post as non-relevant or not trending when it actually isn't.
3. False Positive (FP): Instances where the model incorrectly predicts the positive class. In social media, this could be incorrectly identifying a post as relevant or trending when

it's not.

4. False Negative (FN): Instances where the model incorrectly predicts the negative class. In social media, this could be incorrectly identifying a post as non-relevant or not trending when it actually is.

These metrics are commonly used to evaluate the performance of a classification model. According to the figure 6 the RF algorithm provides 71% accuracy, DT algorithm provides 65% accuracy, logistic algorithm provides 64% accuracy and Naïve Bayes provides 60% accuracy.

According to the figure 6 the RF algorithm provides 66% sensitivity, DT algorithm provides 62% sensitivity, logistic algorithm provides 63% sensitivity and Naïve Bayes provides 51% sensitivity.

According to the figure 6 the RF algorithm provides 77% specificity, DT algorithm provides 68% specificity, logistic algorithm provides 64% specificity and Naïve Bayes provides 70% specificity.

According to the figure 6 the RF algorithm provides 72% balanced accuracy, DT algorithm provides 65% balanced accuracy, logistic algorithm provides 64% balanced accuracy and Naïve Bayes provides 61% balanced accuracy.

VIII. CONCLUSIONS AND FUTURE SCOPE

The research is concerned with extracting opinions from social media data and analyze the opinions of the users on a particular topic and provide with the interpreted results useful for decision-making. This research developed an opinion mining system to extract opinions from Twitter data on a particular domain and using classification methods, a training data set were built to improve the performance of the model and obtained improvements in performance when compared with the existing models. The problem faced during the extraction of opinions was in selecting the measure of interestingness or appropriate feature words, which mainly affect the interpretation of the opinion expressed which influence the decision-making.

As a future work for the research, the main criteria and methodology considered in this model can be augmented to develop an extended model and fine tune to arrive at a generic opinion mining model. Also, the model can be extended with multi-lingual features for better extraction of cross language platforms. An ontology-based approach for opinion mining can be developed to improve the semantic efficiency of the model. Further research work on this model is planned and will continue to be improvised

REFERENCES

[1] Kanimozhi KV, Venkatesan M. Unstructured data analysis-a survey. Int J Adv Res Comput Commun Eng.2015;4:223–5.

[2] Grar M, Cherepnalkoski D, Mozeti I, Kralj Novak P. Stance and infuence of twitter users regarding the brexit referendum. Comput Social Netw. 2017;4:6.

[3] Fan W, Wallace L, Rich S, Zhang Z. Tapping the power of text mining. Commun ACM. 2006;49(9):76–82.

[4] Kaplan Andreas M, Haenlein M. Users of the world, unite! the challenges and opportunities of social media. Business Horiz. 2010;53(1):59–68.

[5] McNab C. What social media ofers to health professionals and citizens. Bull World Health Org. 2009;87(8):566.

[6] Dwivedi Mridula, Yadav Anil, Venkatesh Umashankar. Use of social media by national tourism organizations: a preliminary analysis. Inf Technol Tour. 2011;13(2):93–103.

[7] Greenwood S, Perrin A, Duggan M. Social media update 2016. Pew Res Center. 2016;11:2.

[8] Pang B, Lee L, et al. Opinion mining and sentiment analysis. Found Trend Inf Retriev. 2008;2(1–2):1–135.

[9] Ding X, Liu B, Yu PS. A holistic lexicon-based approach to opinion mining. In Proceedings of the 2008 international conference on web search and data mining. ACM; 2008, p. 231–40.

[10] Taboada M, Brooke J, Tofloski M, Voll K, Stede M. Lexicon-based methods for sentiment analysis. Comput Linguist. 2011;37(2):267– 307.

[11] Bhuta S, Doshi A, Doshi U, Narvekar M. A review of techniques for sentiment analysis of twitter data. In: 2014 International Conference on issues and challenges in intelligent computing techniques (ICICT), IEEE; 2014. p. 583–91.

[12] Balage F, Pardo T. Nilc_usp: a hybrid system for sentiment analysis in twitter messages. In SemEval@ NAACL-HLT.2013, p. 568–72.

[13] Lau KN, Lee KH, Ho Y. Text mining for the hotel industry. Cornell Hotel Restaur Adm Q. 2005;46(3):344–62.

[14] Clark J. Text mining and scholarly publishing. In: Publishing Research Consortium, 2013.

[15] Gémar G, Jiménez-Quintero JA. Text mining social media for competitive analysis. Tour Manag Stud. 2015;11(1):84–90.

[16] Xiang Z, Qianzhou D, Ma Y, Fan W. A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. Tour Manag. 2017;58:51–65.

[17] Demetris A, Constantine D. Co-evolutionary dynamics in social networks: a case study of twitter. Comput Soc Netw.2015;2:14.

[18] Asur S, Huberman BA. Predicting the future with social media. In: 2010 IEEE/WIC/ACM International conference on web intelligence and intelligent agent technology (WI-IAT), IEEE. 2010, p. 492–99.

[19] Zhao Y. Analysing twitter data with text mining and social network analysis. In Proceedings of the 11th Australasian data mining and analytics conference (AusDM 2013), 2013.

[20] Hong L, Davison BD. Empirical study of topic modeling in twitter. In: Proceedings of the frst workshop on social media analytics. ACM; 2010, p. 80–8.

[21] Yang Y, Wang X. World cup 2014 in the twitter world: a big data analysis of sentiments in us sports fans' tweets. Comput Hum Behav. 2015;48:392–400.

[22] O'Connor B, Balasubramanyan R, Routledge BR, Smith NA. From tweets to polls: linking text sentiment to public opinion time series. ICWSM. 2010;11(122–129):1–2.

[23] Enz CA, Verma R. Introduction to the Cornell hospitality research summit special issue: the new science of service innovation in a multipartner world, 2016.

[24] Kotu V, Deshpande B. Predictive analytics and data mining: concepts and practice with rapidminer. New York: Morgan Kaufmann; 2014.

[25] Alghamdi R, Alfalqi K. A survey of topic modeling in text mining. I J ACSA. 2015;6(1):147–53.

[26] Blei David M, Ng Andrew Y, Jordan Michael I. Latent dirichlet allocation. J Mach Learn Res. 2003;3(Jan):993–1022.

[27] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. In: Advances in neural information processing systems. 2002, p. 601–8.

[28] Yang J, Counts S. Predicting the speed, scale, and range of information difusion in twitter. 2010.

[29] Liu B. Sentiment analysis and human language technologies. Synth Lect Hum Lang Techn. 2012;51(1):1–167.

[30] Ahmad M, Aftab S, Ali I. Sentiment analysis of tweets using svm. Int J Comput Appl. 2017;177(5):25–9.

TECHNIQUES FOR ENHANCING PRIVACY CONSERVATION IN RELATIONAL DATABASE

Jithin Babu

Asst. Professor,

Department of Computer Science and Applications,

Naipunnya School of Management,

Cherthala, Alappuzha, Kerala

Jithinbabu007@gmail.com

Abstract—Data confidentiality, Availability and Integrity are essential part of database security. The access control mechanisms and privacy preservation mechanisms protect the data from unauthorized or third party user. When there is a lack in privacy preserving mechanism(PPM) and data is shared with others, the authorized user may need to compromise the privacy of data. The privacy preservation can be achieved through anonymization techniques like generalization or suppression. Along with privacy the precision of the authorized data is important. The aim of the work is to provide better security and minimum level of precision to the retrieved data, for that in this paper an accuracy constrained privacy preserving access control mechanism is implemented with additional constraint on each selection predicate called imprecision bounds. The accuracy constraints are satisfied for multiple roles also. Along with that a workload aware anonymization concept is used for selection predicates. The privacy preservation method introduced includes a multilevel anonymization in order to provide better results. The experimental results shows that the proposed system with multilevel anonymization works better in terms of precision, privacy for more permissions than current state of art.

Index Terms-Access Control, Anonymization, Precision, Privacy preservation, Query evaluation

I. Introduction

Every organization keeps a set of databases to store their information and there may be several situations to share those information with others. As we are living in the information age there are large sources of data around us. To improve the services the organizations collect and analyze the data. The Confidentiality, Integrity and Availability are termed as the [CIA-triad] designed to enable the information security within the organization. They are considered to be the essential components of the security. To ensure that only the authorized information are available only to the authorized users and access control mechanism is implemented in the databases. However there may happen the misuse of sensitive information by the authorized users to compromise the privacy of the customers. For the enhancement of the protection against the identity disclosure and enforcing the privacy policies, the concept of privacy preservation of sensitive data introduced by satisfying some privacy requirements [2].

The paper deals with the privacy preservation in the anonymity aspects. Sensitive information is the essential part of every database and even if we are implementing the privacy protection mechanisms [2] there may have the chance if linkage attacks [4] by the authorized users even after the removal of identifying attributes. This problem got discussed in the area of micro data publishing[3]and privacy definitions like k-anonymity [2], l-diversity [5] etc

Imprecision is a problem in information retrieval. A concept to imprecision bound is introduced in order to solve the problem of imprecision where a minimal level of tolerance or a threshold is defined for each permission [1]. The imprecision added to each permission/query and the aggregate imprecision for all queries got minimized in the workload aware anonymization techniques [5],[6]. The privacy of xml data are also discussed in[15],and also about the spatial database [16].

The topic of satisfying accuracy constraints for the individual permissions in the workload aware anonymization had not discussed yet. The accuracy constrained privacy preserving access control mechanism introduced is relevant in the workload aware anonymizations [1]. The topic of continuous data publishing anonymization [4] also introduced. A static relational table is used in this paper where the table is anonymized only once. A role based access control is also discussed where a concept like accuracy constraints for permissions applied to any privacy preserving security policy.

Normally the anonymization techniques are used to ensure the privacy and security for the data. In many of the works the anonymization technique like generalization is used[1]. The access control mechanisms are essential in order to preserve the confidentiality of the data by providing authentications where as the privacy preservation is also important because it prevents the micro data or the sensitive information not to disclose with a third party user.

To improve the efficiency of the security methods, a privacy preservation module and a accuracy preserving module is combined [1]. The proposed system deals with a multilevel anonymization technique. Here instead of using single level anonymization like generalization or suppression, a combined form of anonymization technique introduced, which include generalization and suppression together.

The rest of this paper proceeds as follows. In section 2 preliminaries are discussed. The section 3 discusses the previous works. The problem definition and privacy preserving access control framework is discussed in section 4. Section5 discusses the proposed system with multi level anonymization. Section6 discusses the experimental results. The paper concludes in section 7.

II. Preliminaries

Given a relation T={A1,A2,A3An}, where A1is an attribute, T* is the anonymized version of the relation T. We assume that T is a static relational table. The attributes can be of the following types:

- *Identifier*. Attributes, e.g., name and social security that can uniquely identify an individual. These attributes are completely removed from the anonymized relation.
- *Quasi-identifier (QI)*. Attributes, e.g., gender, zip code, birth date that can potentially identify an individual based on other information available to an adversary. QI attributes are generalized to satisfy the anonymity requirements.
- *Sensitive attribute*. Attributes, e.g., disease or salary, that if associated to a unique individual will cause a privacy breach.

A. *Access control mechanism for relational databases*

To define tuple-level permissions fine-grained access control like Oracle VPD [8] and SQL [9] are introduced in relational databases. Truman model [10] is introduced for evaluating user queries. A user query is modified by the access control mechanism and only authorized tuples are returned in this model. Column level access control allows queries to execute on the authorized column of the In relational data[8], [11] column level access control mechanisms allow queries to execute on authorized column, by replacing the unauthorized cell values by NULL values [12] cell level access control for relational data is achieved. For defining permissions on objects based on roles in an organization a Role-based Access Control (RBAC) was introduced. An RBAC policy configuration includes a set of Users(U),a set of Roles(R), and a set of Permissions(P).We assume that the selection predicates on the QI attributes define a permission for the relational RBAC model,[11].UA is a user-to-role(U_R)assignment relation and PA is a role to- permission (R _ P) assignment relation.

B. *Privacy definitions*

Here, privacy definitions related to anonymity are introduced.

Definition 1 (Equivalence Class (EC)). An equivalence class is a set of tuples having the same QI attribute values.
Definition 2 (k-anonymity Property). A table T* satisfies the k-anonymity property if each equivalence class has k or more tuples [2].

III. Previous works

Access control mechanisms for databases are an important concept that allows queries only on the authorized part of the database [8], [10]. Later a user authorization is limited to pre-defined predicates in a Predicate based fine-grained access mechanism [11]. Many techniques introduced for the enforcement of access control and privacy policies, they got discussed in[11]. The interaction between the access control mechanisms and the privacy protection mechanisms was missing in those studies.

Recently, Chaudhurietal. Have studied access control with privacy mechanisms [12].Random noise was added to original query in differential privacy and the results which satisfy privacy constraints. But they do not considered the accuracy constraints for permissions. Li et al. [5] defined privacy in terms of K-anonymity where after sampling; k-anonymity offers similar privacy guarantees as those of differential privacy.

The accuracy-constrained privacy preserving access control framework[1] allows the access control administrator to specify imprecision constraints that the privacy protection mechanism is required to meet along with the privacy requirements. Both privacy-aware access control and problem of workload-aware anonymization are similar.

In our analysis of the related work, we focus on query-aware anonymization and a multilevel privacy ensurance which include the combination of the two anonymization techniques generalization and suppression.

We refer the recent survey paper [3] for k-anonymity techniques and algorithms. LeFevre et al. [5] in his work the workload aware anonymization techniques discussed for the first time, they proposed an algorithm named Selection Mondrian algorithm, it is a modification to the greedy multidimensional partitioning algorithm Mondrian [10]. In their algorithm, the greedy splitting heuristic minimizes the sum of imprecision for all queries on the basis of given query workload. a R+tree based anonymization algorithm was introduced by Iwuchukwu and Naughton in [7].

The anonymized data using biased R+tree based on the given query work load is more accurate for queries. Based on space filling curves fork-anonymity and l-diversity Ghinitaet al. have proposed several algorithms [13]. They also introduce the problem of accuracy-constrained anonymization for a given bound of acceptable information loss for each equivalence class [13].

Similarly, Xiaoetal.[14] propose to add noise to queries according to the size of the queries in a given workload to satisfy differential privacy. In any of these works bounds for query imprecision have not been considered. The existing literature on workload-aware anonymization has a focus to minimize the overall imprecision for a given set of queries, but the anonymization with imprecision constraints for individual queries has not been discussed before. We follow the imprecision definition of LeFevreetal.[6] and introduce the constrain to imprecision bound for each query in a given query workload.

IV. Problem Definition and Privacy Preserving Access Control Framework

A. Thek-PIB problem

The optimal k-anonymity problem is discussed in this section. This problem has been shown to be NP-complete for suppression[3] and generalization[4]. The hardness result for k-PIB follows the construction of LeFevreetal. [9] that shows the hardness of k-anonymous multi-dimensional with the smallest average equivalence class size. We show that finding k-anonymous partitioning that violates imprecision bounds for minimum number of queries is also NP-hard. A multi set of tuples is transformed into an equivalent set of distinct (tuple, count) pairs. The cardinality of Query Qi is the sum of count values of tuples falling inside the query hyper-rectangle. The constant qv defines an upper bound for the number of queries that can violate the bounds.

B. The Privacy preserving access control framework

The section describes about how the privacy preserving access control framework [1] works. It is illustrated using Fig.1 (The arrows represents the direction of information flow).Here an access control mechanism as well as a privacy preservation system is introduced together to improve security of the data. The privacy protection mechanism ensures that the privacy and accuracy of the data before the data available to the access control module. The permissions of the access control policy are based on the selection predicates.

The original data or tuple values in a relationship are replaced with the anonymized data normally the generalized data. Here a generalization concept is used for the anonymizations. Here the relaxed and strict access control are discussed, they enforce mechanisms over anonymized data. The access control by reference monitor can be of two types like

1. *Relaxed.* Use overlap semantics to allow success to all partitions that are overlapping the permissions.

2. *Strict.* Use enclosed semantics to allow access only to those partitions that are fully enclosed by the permissions.

V. Proposed System With Multilevel Anonymization

Privacy preservation and access control mechanism are important concept in every information sharing system. As the privacy preserving access control mechanism discussed in [1], it ensures the privacy and access control frameworks in one system. It provides efficient protection of information

Before getting the original data to the access control module the sensitive information passes through the privacy protection module and get anonymized, so that privacy become efficient when compared to the other mechanisms in[2].

Anonymization techniques replace the original data with some other text or symbols which cannot be easily identifiable by the users. In the privacy preserving access control mechanism[1],the anonymization technique used was generalization method. In generalization the values are replaced with a range of value (Foreg.Lettheageofbob20, it becomes <10-30> range after generalization).

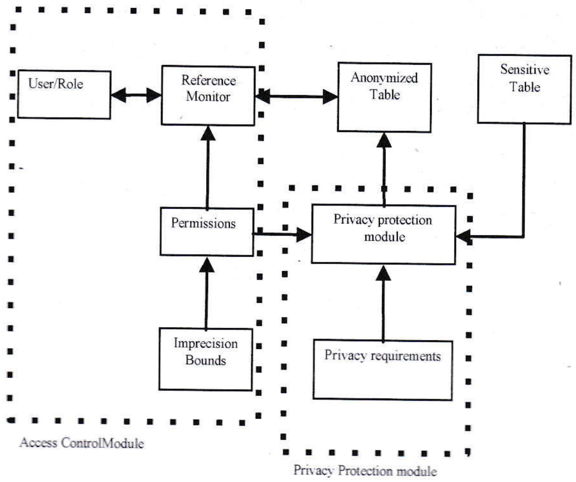


Fig1 Accuracy-constrainedprivacy-preservingaccess control mechanism.

Here the anonymization is applied only once to the data for the security enhancement. The proposed system in which a multilevel anonymization is introduced in order to enhance the security more. In this system, a suppression technique also implemented along with generalization, where the data will be replaced with any symbols or letters(eg. Let the zip code of Alice be 812345, after suppression the value can be represented as 8123**, 81****,******, etc).

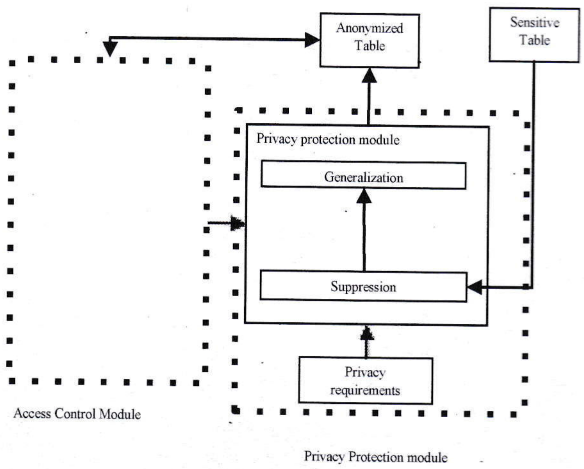


Fig2.Accuracy-constrainedprivacy-preservingaccess control mechanism with multilevel

In the first level anonymization the suppressed information of original table is used and in the second level of anonymization a generalized value is used. So that there it can assure a multilevel security for the data. This also provides a minimum level of precision to the data, along with that the sensitive information will get protected. The proposed system depicted in fig. 2

VI. Experimental Evaluation

The system is implemented in dreamweaver8 with JSP support. Tomcat server is used as application server and mySQL as backend database.

The section describes about the experimental evaluation done in a medical dataset. The privacy preserving access control model and the multilevel access control model are used to show the experimental results. The fig. 3 shows the experimental result.

The fig3 shows that the number of tuples retrieved by the query given is increased with the increase in the predicates of the query. The blue line shows the number of tuples retrieved when the number of predicate is one, two, and three with three predicates in privacy preserving access control module. The red line shows the number of tuples retrieved by the queries with multiple predicates in multilevel access control model. It clearly shows that the proposed method performs better than the previous methods in terms of number of tuples retrieved.

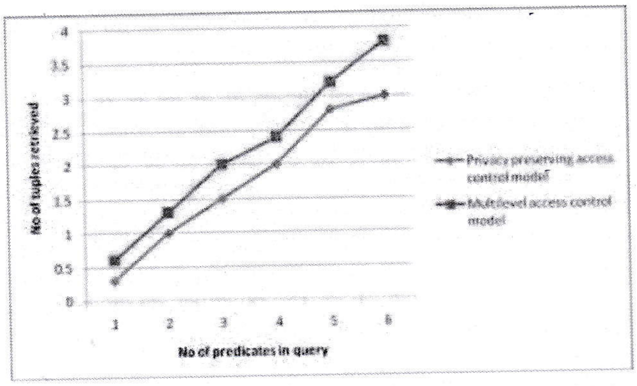


Fig.3.Experimentalresultsonnumberofquerypredicates and number of tuples retrieved

The fig 4 shows the pie chart that depict the relation between the filter count (number of predicates) and result count(number of tuples retrieved).

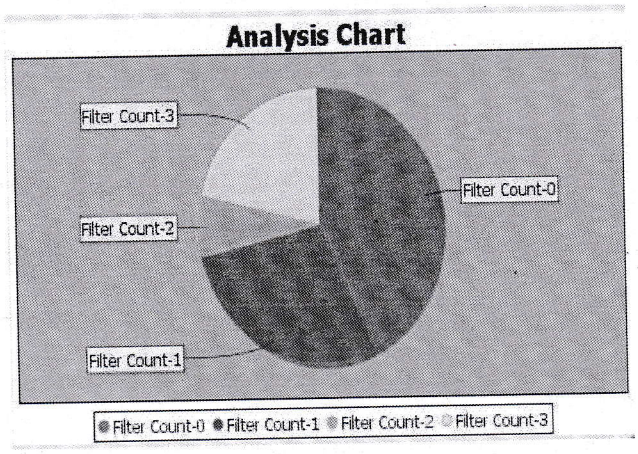


Fig4.Analysis chart on the basis of filter count and result counts

The pie chart in the fig4 is drawn with the history of search in the month of February 2015. The Fig 5 shows the details with which the chart was drawn, which is the history of search with different number of predicates in the month of February 2015.

Start Date 2015-02-01 End date 2015-02-28 Search

Filter Count	Date	Result Count
3	26-02-2015	2
3	26-02-2015	2
0	26-02-2015	4
0	26-02-2015	4
1	26-02-2015	1
1	26-02-2015	4
2	26-02-2015	4
3	26-02-2015	0
0	27-02-2015	4
1	28-02-2015	4
1	28-02-2015	4
0	28-02-2015	4
0	28-02-2015	4
0	28-02-2015	2

Fig5.HistoryofsearchinFebruary

VII. Conclusion

The paper discusses about the need of the security in the databases. The access control and privacy preserving modules are combined in order to provide better results. Here the sensitive information in original database will only be available to the access control modules after providing some privacy to data. Mainly anonymization techniques are used to enforce the privacy and mainly generalization techniques are implemented. In the proposed system a multilevel anonymization is introduced in order to improve the efficiency in privacy preserving access control mechanism. The proposed method works better in terms of tuple retrieval and privacy preservation and it is showed by the experimental results.

Acknowledgment

The author would like to thanks all those people, who guided and supported. Without their valuable guidance and support, this task was not possible and also likes to thank colleagues for their discussions and suggestions.

References

[1] Zahid Pervaiz, Walid G.Aref, Arif Gafoor, Nagabhushana Prabhu"Accuracy constrained privacy preserving access control mechanism for relational databases" IEEE Transaction on Knowledge Engineering, vol.26, No.4, April 2014, pp.795-807

[2] E. Bertinoand R. Sandhu, "Database Security-Concepts, Approaches,andChallenges,"IEEETrans.DependableandSecureComputing,vol.2, no. 1, pp. 2-19, Jan.-Mar. 2005.

[3] P. Samarati, "Protecting Respondents' Identities in Micro data Release," IEEE Trans. Knowledge and Data Eng., vol. 13, no. 6, pp.1010-1027, Nov. 2001.

[4] B. Fung, K. Wang, R. Chen, and P. Yu, "Privacy-Preserving Data Publishing: A Survey of Recent Developments," ACM Computing Surveys, vol. 42, no. 4, article 14, 2010..

[5] A. Machanavajjhala, D. Kifer, J. Gehrke, and M.Venkitasubramaniam, "L-Diversity: Privacy Beyondk-anonymity,"

ACM Trans. Knowledge Discovery from Data, vol. 1, no. 1, article 3,2007.

[6] K. LeFevre, D. DeWitt, and R. Ramakrishnan, "Workload-Aware Anonymization Techniques for Large-Scale Datasets," ACM Trans. Database Systems, vol. 33, no. 3, pp. 1-47, 2008.

[7] T. Iwuchukwu and J. Naughton, "K-Anonymization as Spatial Indexing: Toward Scalable and Incremental Anonymization," Proc.33rd Int'l Conf. Very Large Data Bases, pp. 746-757, 2007.

[8] K. Browder and M. Davidson, "The Virtual Private Database inoracle9ir2," Oracle TechnicalWhite Paper, vol. 500, 2002.

[9] S. Rizvi, A. Mendelzon, S. Sudarshan, and P. Roy, "Extending Query Rewriting Techniques for Fine-Grained Access Control," Proc. ACMSIGMOD Int'l Conf. Management of Data, pp. 551-562, 2004.

[10] S. Chaudhuri, T. Dutta, and S. Sudarshan, "Fine GrainedAuthorizationthroughPredicatedGrants,"Proc.IEEE23rdInt'lConf.Data Eng., pp. 1174-1183, 2007.

[11] K. LeFevre, D. DeWitt, and R. Ramakrishnan, "Mondrian Multidimensional K-Anonymity," Proc. 22nd Int'l Conf. Data Eng., pp.25-25,2006.

[12] N.Li,W.Qardaji,andD.Su, "Provably Private Data Anonymization: Or, k-Anonymity Meets Differential Privacy," Arxiv preprintarXiv:1101.2604, 2011.

[13] G. Ghinita, P. Karras, P. Kalnis, and N. Mamoulis, "Fast Data Anonymization with Low Information Loss," Proc. 33rd Int'l Conf.Very Large Data Bases, pp. 758-769, 2007.

[14] G.Ghinita,P.Karras,P. Kalnis, and N. Mamoulis, "A Framework for Efficient Data

[15] Anonymization under Privacy and Accuracy Constraints," ACMTrans. Database Systems, vol. 34, no. 2, article 9, 2009.

[16] E.Bertino,S.Castano,andE.Ferrari"SecuringXMLDocumentswithAuthor-X", IEEE Internet Computing, vol. 5, no. 3, pp. 21-30, 2001.

[17] NehaSehtaetal"A Fine Grained Access Control Model for Relational Databases", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (1) , 2012, 3183 – 3186, 3183.