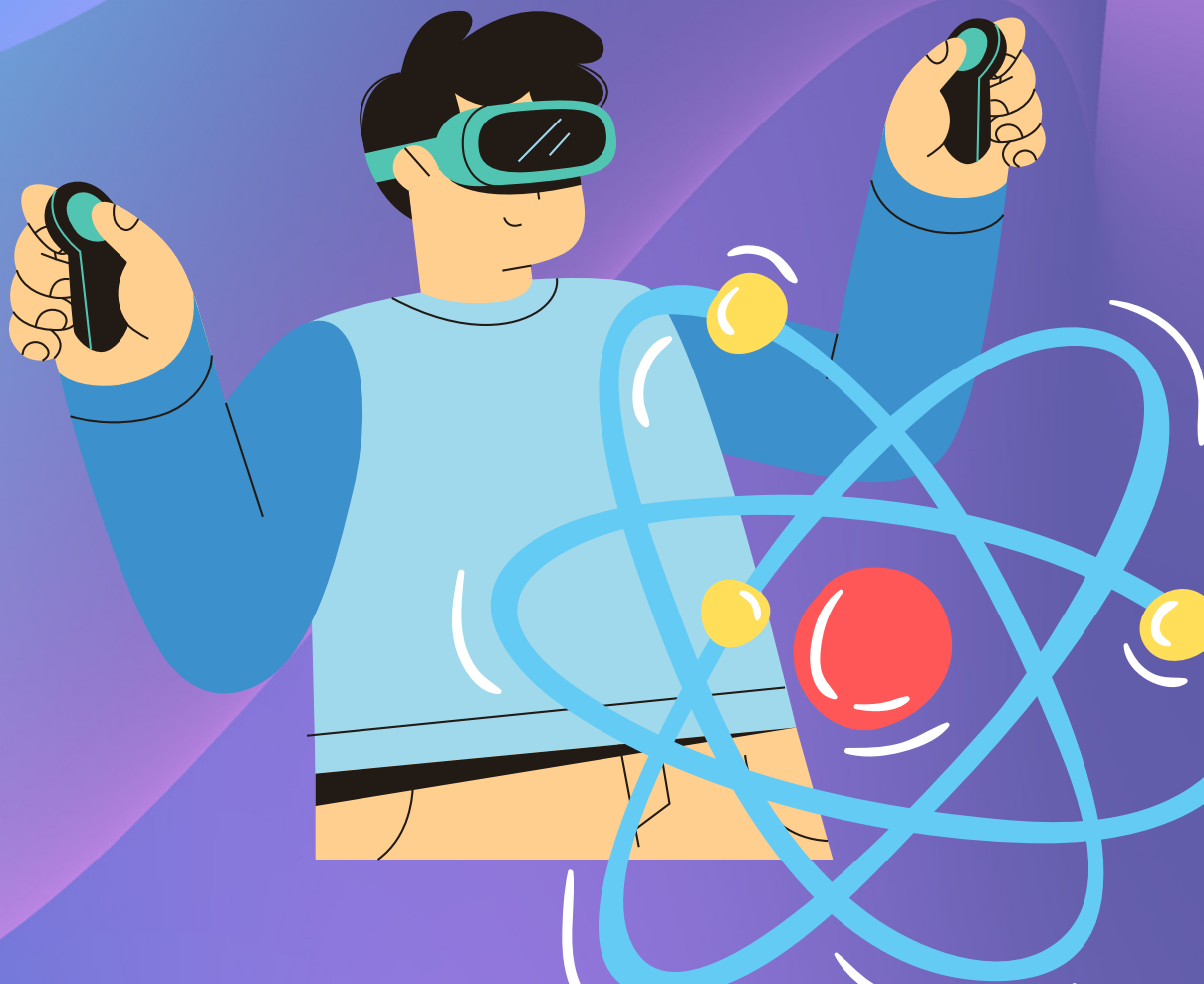
Emerging Research in Science & Engineering

Dr. Shraddha Prasad
Dr. Harmeet Kaur



Emerging Research in Science & Engineering

Parab Publications
Sharing Knowledge, Shaping Future

India | UAE | Nigeria | Malaysia | Montenegro | Iraq | Egypt | Thailand | Uganda | Philippines | Indonesia

www.parabpublications.com

Emerging Research in Science & Engineering

Edited By:

Dr. Shraddha Prasad
Associate Professor
Department of Applied Science
Faculty of Science & Engineering
Jharkhand Rai University, Ranchi

Dr. Harmeet Kaur
Associate Professor
Department of Management
Faculty of Commerce & Management
Jharkhand Rai University, Ranchi

First Impression: August 2023

Emerging Research in Science & Engineering

ISBN: 978-81-19585-14-4

Rs. 1000/- ($80)

DISCLAIMER

Information contained in this book has been published by Parab Publications and has been obtained by the editors from sources believed to be reliable and correct to the best of their knowledge. The authors are solely responsible for the contents of the articles compiled in this book. Responsibility of authenticity of the work or the concepts/views presented by the authors through this book shall lie with the authors and the publisher has no role or claim or any responsibility in this regard. Errors, if any, are purely unintentional and readers are requested to communicate such error to the author to avoid discrepancies in future.

Published by:
Parab Publications

Preface

Emerging research in science & engineering contributes to the creation of new technologies, this edited book covers certain areas of emerging research in science & engineering. *The contributions by the authors includes a review on H-TCP – Congestion Control Mechanism* in which the authors have explore multiple techniques and methods to enhance network performance.

In the chapter *Machine Learning based IoT application for Patient Support System*, its author has highlighted the potential benefits of machine learning based IoT in healthcare, including personalized and cost-effective care for patients, real-time monitoring and analysis of patient data, and improved patient outcomes.

In another chapter *Modelling an Intrusion Detection System in Distributed Networks* the authors highlights about intrusion detection system in distributed networks useful for the security concern.

The Chapter *Basic Understanding of Security Techniques in IoT based on Machine Learning* present the significant potential for enhancing the security of IoT systems by providing intelligent analysis, anomaly detection, and predictive modelling.

The chapter titled "*Design the Real Time Driver Drowsiness and Yawning Detection System using IOT"* the authors illustrate about the technology used to wake up the driver from drowsiness so to avoid severe accidents.

Some of the authors have highlighted on *Cyber Threat Intelligence* and its importance in helping organizations stay ahead of cybercriminals

Few chapters include *Machine Learning Algorithms to Classify Medication for Patients, The Impact of Artificial Intelligence on Higher Education*, *A Survey on security protocols for internet of things*, *The Impact of ChatGPT on student learning*, *Design of Real Time Automatic Emotion Recognition (AER) based music Recommendation system.*

In another chapter the authors has presented a *Comparative Review on Advances, Approaches and Dimensions of Water Purification.* Few authors has explore the *Power of Open Source Intelligence (OSINT*). Another chapter highlighted the *Impact of Social Media on Youth.*

The impact of *Blast Furnace Hydrogen Injection in Tata Steel Ltd.* are illustrated in detail by the authors.

In the chapter *Employing Multiple Linearly Arranged ArUco Markers as Reference Scale for Enhancing Image-Based Length Measurement* author has provide a novel approach utilizing multiple linearly arranged ArUco markers as a reference scale to enhance the accuracy and reliability of image-based length measurement algorithms.

In the chapter *Natural Language Pre-processing of text Data for effective Text Classification in Deep Learning* the authors describe complete information for the classification of text data preprocessing techniques and a comparison of the suitability and accuracy obtained from two pre-processing techniques.

In the chapter *Development of a Centralized Electronic Medical Record System – in Health Care & Governance* the author illustrated the benefits and impacts of implementing the Centralized Electronic Medical Record System (CEMRS) in healthcare and governance.

This edited book highlights the emerging research in science & engineering. I hope this book would be helpful for students, researchers and academicians in the field of science and engineering.

Acknowledgement

We gratefully acknowledge all the authors for contributing paper which provides richness of content to this book. We would like to offer my sincere thanks to all the authors.

We extremely indebted to Prof. (Dr.) Savita Sengar (Vice Chancellor, Jharkhand Rai University, Ranchi), Prof. (Dr.) Piyush Ranjan (Registrar, Jharkhand Rai University, Ranchi), as a source of inspiration for us in our academic growth.

We are thankful to Parab Publications for publishing this edited book.

Dr. Shraddha Prasad

Dr. Harmeet Kaur

Table of Contents

H-TCP – CONGESTION CONTROL MECHANISM: A REVIEW

Abhay Ranjan Pandey[1] and Anu Priya[2]
[1]MCA Student and[2]Assistant Professor, Amity Institute of Information Technology, Amity University Patna

ABSTRACT
Congestion control is a critical aspect of network management, ensuring optimal performance and efficient resource allocation. Over the years, various congestion control mechanisms have been developed to address the challenges posed by increasing network traffic and diverse application requirements. This paper focuses on H-TCP (Host-TCP), a congestion control mechanism that incorporates multiple techniques and methods to enhance network performance. We discuss the fundamental principles of congestion control, explore different techniques employed in H-TCP, and provide insights into its advantages and limitations. Through this comprehensive analysis, we aim to shed light on the effectiveness of H-TCP in managing congestion and its potential impact on network performance.

INTRODUCTION
With the exponential growth of network traffic and the increasing diversity of applications, effective congestion control mechanisms have become essential for ensuring efficient network performance and resource allocation. Congestion occurs when the demand for network resources exceeds its capacity, resulting in degraded performance, increased latency, packet loss, and reduced throughput. Addressing these challenges requires the development of sophisticated congestion control mechanisms that adapt to network dynamics and optimize resource utilization.

The H-TCP (Host-TCP) congestion control mechanism has emerged as a promising solution that incorporates multiple techniques and methods to enhance network performance and manage congestion effectively. Unlike traditional congestion control mechanisms that rely solely on additive increase and multiplicative decrease (AIMD), H-TCP leverages a variety of strategies, including rate-based and delay-based control, cross-layer optimization, queue management, and adaptive packet scheduling. By integrating these techniques, H-TCP aims to achieve improved throughput, fairness, reduced latency, and robustness to network dynamics.

Motivation: The motivation behind this study lies in the need to explore and understand the potential benefits and limitations of H-TCP as a congestion control mechanism. As network environments become increasingly complex, traditional mechanisms may struggle to meet the demands of diverse applications and changing network conditions. Thus, it becomes crucial to investigate alternative approaches that can effectively handle congestion while maximizing network performance.

Objectives of the Study: The primary objectives of this study are as follows:

1. To provide a comprehensive understanding of congestion control mechanisms and their significance in network management.
2. To introduce the H-TCP congestion control mechanism and highlight its unique features and advantages.
3. To explore the different techniques employed in H-TCP, including rate-based and delay-based control, cross-layer optimization, queue management, and adaptive packet scheduling.
4. To evaluate the performance of H-TCP compared to traditional congestion control mechanisms through experimental analysis and real-world case studies.
5. To identify the advantages and limitations of H-TCP and discuss potential research directions for its further development and deployment.

Structure of the Paper: This paper is organized as follows:

Section 2 provides a comprehensive overview of the fundamentals of congestion control, including its definition, causes, importance, goals, and requirements.

Section 3 introduces traditional congestion control mechanisms such as AIMD, RED, ECN, and FIFO, comparing and evaluating their effectiveness.

Section 4 presents an introduction to the H-TCP congestion control mechanism, discussing its design principles, objectives, and key features.

Section 5 explores the techniques employed in H-TCP, including rate-based and delay-based control, cross-layer optimization, queue management strategies, and adaptive packet scheduling.

Section 6 discusses the advantages and limitations of H-TCP, highlighting its impact on throughput, fairness, latency, and packet loss.

Section 7 focuses on the performance evaluation of H-TCP through experimental analysis, comparative studies with traditional mechanisms, and real-world deployment case studies.

Section 8 highlights future directions and research challenges, addressing scalability, compatibility, interoperability, standardization, integration with emerging technologies, and security considerations.

Section 9 concludes the paper, summarizing the findings, implications, and providing an outlook on the future of H-TCP as a congestion control mechanism.

2. FUNDAMENTALS OF CONGESTION CONTROL

2.1 Definition of Congestion Control: Congestion control is a fundamental aspect of network management that aims to regulate the flow of data and prevent network congestion. It involves implementing mechanisms and strategies to ensure that network resources are efficiently utilized, network performance is optimized, and quality of service (QoS) requirements are met. Congestion control encompasses a range of techniques, algorithms, and protocols that dynamically adapt to varying network conditions and mitigate the adverse effects of congestion.

2.2 Causes of Congestion: Congestion can occur in a network for various reasons, including:

- **High Data Traffic:** An increase in the volume of data being transmitted through the network can lead to congestion, especially during periods of peak demand or when network capacity is limited.
- **Network Bottlenecks:** Bottlenecks in the network infrastructure, such as congested links, switches, or routers, can impede the smooth flow of data and contribute to congestion.
- **Network Topologies:** Inefficient network topologies, where multiple paths converge into a single point, can cause congestion when the traffic load exceeds the capacity of the converging point.
- **Misconfigured Devices:** Improperly configured network devices, such as routers or switches, can cause congestion by misrouting or delaying packets.
- **Denial-of-Service (DOS) Attacks:** Malicious activities, such as DoS attacks, can intentionally flood the network with excessive traffic, overwhelming its resources and causing congestion.

2.3 Importance of Congestion Control: Congestion control is vital for several reasons:

- **Optimal Network Performance:** Effective congestion control ensures that network resources are utilized efficiently, minimizing delays, packet loss, and service degradation. It helps maintain desirable levels of throughput, latency, and reliability.

- **Fair Resource Allocation:** Congestion control mechanisms strive to allocate network resources fairly among competing flows, preventing a few dominant flows from monopolizing the available bandwidth and ensuring equitable sharing.
- **Quality of Service (QoS) Guarantees:** By regulating the flow of traffic, congestion control mechanisms help enforce QoS requirements, ensuring that applications receive the necessary network resources to meet their performance needs.
- **Avoidance of Network Collapse:** Uncontrolled congestion can lead to network instability, packet loss, and eventual network collapse. Congestion control mechanisms help prevent such catastrophic scenarios by proactively managing congestion and maintaining network stability.

2.4 Goals and Requirements of Congestion Control: Congestion control mechanisms aim to achieve the following goals:

- **Avoidance:** The mechanism aims to prevent congestion from occurring in the first place by dynamically regulating the rate of data transmission, ensuring it does not exceed the network's capacity.
- **Control:** If congestion does occur, the mechanism strives to control and mitigate its effects by intelligently adapting the flow of data, reducing the congestion levels, and maintaining network performance.
- **Fairness:** Congestion control mechanisms promote fairness by ensuring that different flows sharing the network resources receive an equitable share, preventing any particular flow from dominating the available bandwidth.
- **Responsiveness:** The mechanism should be able to detect and react promptly to changing network conditions, adapting the data transmission rate and congestion control parameters in real-time.
- **Scalability:** Congestion control mechanisms must be scalable to handle large-scale networks with numerous flows, diverse applications, and varying traffic patterns.
- **Compatibility:** They should be compatible with existing network protocols and infrastructure, allowing for seamless integration and deployment without significant disruptions.
- **Robustness:** The mechanism should exhibit robustness to handle network dynamics, such as link failures, changing network topologies, or varying traffic loads, and be able to recover quickly from congestion events.

3. TRADITIONAL CONGESTION CONTROL MECHANISMS

Traditional congestion control mechanisms have been developed over the years to manage congestion in networks. These mechanisms employ different strategies and algorithms to regulate the flow of data and mitigate the adverse effects of congestion. In this section, we will discuss some widely used traditional congestion control mechanisms, namely AIMD (Additive Increase, Multiplicative Decrease), RED (Random Early Detection), ECN (Explicit Congestion Notification), and FIFO (First-In-First-Out). We will compare and evaluate their effectiveness in managing network congestion.

3.1 AIMD (Additive Increase, Multiplicative Decrease): AIMD is a widely employed congestion control algorithm used in various transport protocols such as TCP. AIMD operates on the principle of gradually increasing the sending rate when the network is not congested and reducing it exponentially when congestion is detected. It employs additive increase, where the sending rate is incremented by a small amount for every successful transmission, and multiplicative decrease, where the sending rate is halved upon congestion indication.

AIMD exhibits a fair sharing behavior, where different flows in the network tend to receive an equitable share of bandwidth. It is reactive to congestion, as it reduces its sending rate upon packet loss, indicating network congestion. However, AIMD can suffer from performance degradation in scenarios where there is a sudden increase in the number of flows or when flows with different round-trip times share the same bottleneck.

3.2 RED (Random Early Detection): RED is a congestion control mechanism commonly used in routers to manage congestion in IP networks. RED aims to prevent global synchronization and achieve fairness among flows by selectively dropping packets before a congested state is reached. It uses probabilistic dropping of packets based on the average queue length.

RED dynamically adjusts the dropping probability based on the current congestion level. When the queue length exceeds a predefined threshold, it starts dropping packets with a low probability and increases the probability as the queue length increases. By dropping packets proactively, RED helps avoid congestion collapse and promotes fairness among flows.

However, RED requires careful tuning of its parameters to achieve optimal performance, and it can be sensitive to the choice of threshold values. In some scenarios, RED may also suffer from global synchronization, where multiple flows reduce their sending rates simultaneously, leading to underutilization of available bandwidth.

3.3 ECN (Explicit Congestion Notification): ECN is a congestion control mechanism that enables routers to signal congestion to end systems without dropping packets. It utilizes the ECN field in IP headers to mark packets as congestion experienced (CE) when congestion is detected. The receiver then signals this congestion indication to the sender, which can respond by reducing its sending rate.

ECN allows for a more responsive and proactive congestion control mechanism by avoiding the packet loss associated with traditional mechanisms. It provides faster feedback to the sender about network congestion, leading to better resource utilization and reduced latency. ECN can effectively mitigate congestion in scenarios where packet loss is undesirable or costly, such as in multimedia streaming or real-time applications.

However, the widespread deployment of ECN requires support from both network devices and end systems. In some cases, ECN markings can be subject to misinterpretation or manipulation, impacting its effectiveness in managing congestion.

3.4 FIFO (First-In-First-Out): FIFO is a simple queuing discipline used in routers where packets are transmitted in the order they arrive. In this mechanism, the first packet to arrive is the first to be transmitted, regardless of its priority or importance.

FIFO has the advantage of simplicity and low implementation overhead. However, it can suffer from performance degradation under congestion. In scenarios where the queue becomes full, newly arriving packets experience increased queuing delays, leading to higher latency and potential packet loss. FIFO does not consider the importance or characteristics of different flows and does not provide any fairness or differentiation among flows.

Comparative Evaluation: When comparing these traditional congestion control mechanisms, several factors need to be considered, including fairness, responsiveness, robustness to network dynamics, and ability to prevent congestion collapse.

AIMD exhibits fairness among flows, but it can be slow to react to congestion events, resulting in performance degradation. RED, on the other hand, proactively drops packets to prevent congestion collapse and achieve fairness but requires careful tuning of parameters. ECN provides fast feedback without relying on packet loss, but its effectiveness depends on widespread deployment and support. FIFO is simple but lacks fairness and differentiation among flows, leading to potential performance issues under congestion.

4. INTRODUCTION TO H-TCP CONGESTION CONTROL MECHANISM

The H-TCP (Host-TCP) congestion control mechanism is an innovative approach that integrates multiple techniques and methods to enhance network performance and effectively manage congestion. H-TCP builds upon the foundations of traditional congestion control mechanisms while introducing novel features and design principles to address the limitations of existing approaches. In this section, we will provide an introduction to H-TCP, discussing its design principles, objectives, and key features.

4.1 Overview of H-TCP: H-TCP is a congestion control mechanism specifically designed for TCP (Transmission Control Protocol), a widely used transport protocol in computer networks. H-TCP aims to optimize TCP's congestion control mechanisms by incorporating advanced techniques and strategies to achieve improved throughput, fairness, reduced latency, and robustness to network dynamics.

4.2 Design Principles and Objectives: The design principles of H-TCP revolve around the following key objectives:

4.2.1 Performance Optimization: H-TCP seeks to enhance network performance by intelligently adapting the data transmission rate based on network conditions. It aims to achieve higher throughput while ensuring fairness among flows, effectively utilizing available bandwidth.

4.2.2 Latency Reduction: Reducing latency is a crucial objective of H-TCP. By employing delay-based congestion control techniques, H-TCP optimizes the queuing delays and minimizes the time taken for data packets to traverse the network, leading to improved responsiveness.

4.2.3 Robustness to Network Dynamics: H-TCP is designed to be robust and adaptive to changing network dynamics. It can effectively handle scenarios with varying traffic loads, link failures, or changes in network topology without compromising performance or stability.

4.3 Key Features and Components: The key features and components of H-TCP that contribute to its effectiveness in congestion control include:

4.3.1 Rate-Based and Delay-Based Control: H-TCP combines both rate-based and delay-based congestion control techniques. It regulates the data transmission rate based on the available network bandwidth and adjusts the congestion control parameters according to the measured round-trip times (RTTs) and queuing delays. This hybrid approach enables H-TCP to achieve better performance in different network conditions.

4.3.2 Cross-Layer Optimization: H-TCP incorporates cross-layer optimization techniques by considering information from multiple layers of the network protocol stack. It leverages feedback and measurements from the physical layer, network layer, and transport layer to make informed decisions regarding congestion control and resource allocation.

4.3.3 Queue Management Strategies: H-TCP utilizes advanced queue management strategies to effectively manage congestion. By employing intelligent queue management techniques, such as RED (Random Early Detection) or its variants, H-TCP can proactively drop or mark packets based on the queue length and congestion indicators to prevent congestion collapse and ensure fair sharing of network resources.

4.3.4 Congestion Window Adaptation: H-TCP dynamically adjusts the congestion window size, which determines the number of packets that can be transmitted before receiving an acknowledgment. By adapting the congestion window size based on network conditions, H-TCP optimizes throughput, reduces congestion, and prevents excessive packet loss.

4.3.5 Adaptive Packet Scheduling: H-TCP incorporates adaptive packet scheduling algorithms that prioritize packets based on their urgency, importance, or application requirements. By intelligently scheduling packets, H-TCP can improve overall network performance and meet the QoS (Quality of Service) demands of different applications.

5. TECHNIQUES EMPLOYED IN H-TCP CONGESTION CONTROL

H-TCP (Host-TCP) congestion control mechanism employs a combination of techniques to optimize network performance, improve fairness, reduce latency, and ensure robustness. In this section, we will explore the key techniques utilized by H-TCP, including rate-based and delay-based control, cross-layer optimization, queue management strategies, and adaptive packet scheduling.

5.1 Rate-Based and Delay-Based Control: H-TCP combines rate-based and delay-based congestion control techniques to adaptively regulate the data transmission rate. Rate-based control adjusts the sending rate based on available network bandwidth, while delay-based control takes into account round-trip times (RTTs) and queuing delays. By utilizing both approaches, H-TCP achieves a fine balance between maximizing throughput and minimizing latency.

Rate-based control ensures that the sending rate remains within the capacity limits of the network, preventing congestion and excessive packet loss. Delay-based control dynamically adjusts the congestion control parameters based on measured RTTs and queuing delays. It allows H-TCP to react promptly to changes in network conditions, adapting the data transmission rate to optimize performance.

5.2 Cross-Layer Optimization: H-TCP incorporates cross-layer optimization techniques, leveraging information from multiple layers of the network protocol stack. By considering

Feedback and measurements from the physical layer, network layer, and transport layer, H-TCP makes informed decisions regarding congestion control and resource allocation.

Cross-layer optimization enables H-TCP to gain insights into the characteristics of the underlying network, such as available bandwidth, link conditions, and congestion indicators. By utilizing this information, H-TCP can dynamically adjust its congestion control parameters, adapt its transmission rate, and optimize resource allocation to achieve better performance and improved utilization of network resources.

5.3 Queue Management Strategies: H-TCP utilizes advanced queue management strategies to effectively manage congestion. It employs techniques such as RED (Random Early Detection) or its variants to proactively drop or mark packets based on the queue length and congestion indicators.

Queue management in H-TCP aims to prevent congestion collapse, maintain fairness among flows, and optimize network performance. By selectively dropping or marking packets when the queue length exceeds a threshold, H-TCP helps regulate the flow of data and avoids excessive buildup of congestion. This proactive approach allows H-TCP to achieve better utilization of available bandwidth and minimize the occurrence of packet loss and network congestion.

5.4 Adaptive Packet Scheduling: H-TCP incorporates adaptive packet scheduling algorithms to prioritize packets based on their urgency, importance, or application requirements. By intelligently scheduling packets, H-TCP improves overall network performance and meets the Quality of Service (QoS) demands of different applications.

Adaptive packet scheduling in H-TCP ensures that packets are transmitted in a manner that optimizes performance and minimizes latency. It takes into consideration factors such as packet size, priority, deadlines, and service requirements. By dynamically adjusting the packet

scheduling algorithm based on the network conditions and application characteristics, H-TCP can effectively allocate network resources and optimize the overall performance of the system.

By employing rate-based and delay-based control, cross-layer optimization, queue management strategies, and adaptive packet scheduling, H-TCP achieves a comprehensive congestion control mechanism that addresses various aspects of congestion management, fairness, latency reduction, and network optimization. These techniques enable H-TCP to adapt to changing network conditions, provide efficient resource utilization, and enhance the overall performance of TCP-based communication in congested network scenarios.

6. ADVANTAGES AND LIMITATIONS OF H-TCP

6.1 Advantages

6.1.1 Enhanced Throughput: H-TCP's integration of rate-based and delay-based control, along with advanced queue management strategies, enables it to optimize the data transmission rate and efficiently utilize available network bandwidth. By dynamically adapting the sending rate based on network conditions, H-TCP can achieve higher throughput compared to traditional congestion control mechanisms. This leads to improved network performance and efficient data transfer.

6.1.2 Improved Fairness: Fairness among flows is a critical aspect of congestion control. H-TCP aims to ensure fair sharing of network resources among different flows by intelligently adjusting the congestion control parameters and employing queue management strategies. Through its fairness mechanisms, H-TCP provides equitable access to available bandwidth, preventing a single flow from dominating the network and degrading the performance of other flows.

6.1.3 Reduced Latency: H-TCP's combination of rate-based and delay-based control mechanisms, along with adaptive packet scheduling, contributes to latency reduction. By considering RTTs, queuing delays, and packet priorities, H-TCP can dynamically adjust its transmission rate and prioritize time-sensitive packets. This results in improved responsiveness, reduced queuing delays, and minimized end-to-end latency, particularly in scenarios with high network congestion.

6.1.4 Mitigated Packet Loss: Packet loss is a significant issue in congested networks, leading to retransmissions and reduced throughput. H-TCP's proactive queue management strategies, such as RED or its variants, help prevent congestion collapse by selectively dropping or marking packets when the queue length exceeds a threshold. By avoiding excessive packet loss, H-TCP minimizes the impact of congestion on overall network performance, improving reliability and efficiency.

6.2 Limitations

6.2.1 Complexity and Deployment Challenges: The adoption of H-TCP may face challenges due to its increased complexity compared to traditional congestion control mechanisms. Widespread deployment of H-TCP requires support from network devices, operating systems, and applications. Ensuring compatibility across different platforms and achieving seamless integration can be a potential limitation.

6.2.2 Sensitivity to Network Dynamics: While H-TCP strives to be robust to network dynamics, rapid changes in network conditions, such as link failures or sudden variations in traffic load, may pose challenges. H-TCP's performance may be impacted during such situations, requiring additional adaptations and fine-tuning to maintain optimal congestion control.

6.2.3 Parameter Tuning: Like many congestion control mechanisms, H-TCP requires careful parameter tuning to achieve optimal performance. The effectiveness of H-TCP can depend on

the choice of parameters, such as queue thresholds, congestion control window size, and delay estimation mechanisms. Proper calibration and configuration of these parameters are necessary to ensure the desired performance outcomes.

6.2.4 Network Heterogeneity: H-TCP's effectiveness may vary in heterogeneous networks with diverse link capacities, network topologies, and traffic patterns. It may require additional adaptations or variations to cater to the specific characteristics of different networks and ensure optimal performance across various scenarios.

7. PERFORMANCE EVALUATION OF H-TCP

The performance evaluation of H-TCP involves conducting experimental analyses, comparative studies with traditional congestion control mechanisms, and real-world deployment case studies. These approaches provide insights into the effectiveness of H-TCP in managing network congestion and improving performance. Let's explore each evaluation method:

7.1 Experimental Analysis: Experimental analysis involves setting up controlled network environments and conducting tests to measure the performance of H-TCP under various scenarios. This evaluation approach typically involves the following steps:

7.1.1 Testbed Setup: A testbed is constructed to simulate network conditions and emulate different congestion scenarios. This may involve configuring network devices, routers, and hosts with H-TCP and other competing congestion control mechanisms.

7.1.2 Performance Metrics: Various performance metrics are defined to evaluate H-TCP's effectiveness. These metrics may include throughput, fairness, latency, packet loss, and network utilization. They provide quantitative measures to compare H-TCP's performance against traditional mechanisms.

7.1.3 Experiment Design: Experiments are designed to assess H-TCP's performance under different congestion scenarios, varying link capacities, and traffic patterns. Realistic workloads and traffic traces may be used to replicate real-world network conditions.

7.1.4 Data Collection and Analysis: During the experiments, data is collected on performance metrics such as throughput, fairness, latency, and packet loss. The collected data is analyzed to evaluate H-TCP's performance and its advantages over traditional mechanisms. Statistical analysis techniques may be employed to derive meaningful conclusions.

7.2 Comparative Studies: Comparative studies involve comparing H-TCP's performance with traditional congestion control mechanisms. This evaluation method helps understand the relative strengths and weaknesses of H-TCP compared to existing approaches. The following steps are typically involved:

7.2.1 Selection of Traditional Mechanisms: Representative traditional mechanisms such as TCP Reno, TCP Cubic, or others are chosen for comparison. These mechanisms are well-established and widely used in network environments.

7.2.2 Performance Metrics: Similar to experimental analysis, performance metrics such as throughput, fairness, latency, and packet loss are employed to compare the performance of H-TCP against traditional mechanisms.

7.2.3 Experimental Setup: A controlled testbed is set up to replicate congestion scenarios and network conditions. H-TCP and traditional mechanisms are implemented on different hosts, and their performance is measured simultaneously.

7.2.4 Data Collection and Analysis: Data on performance metrics is collected for both H-TCP and traditional mechanisms. A comparative analysis is conducted to evaluate the advantages and disadvantages of H-TCP in terms of the measured metrics. Statistical tests may be employed to assess the significance of the differences observed.

7.3 Real-World Deployment Case Studies: Real-world deployment case studies involve deploying H-TCP in operational networks and evaluating its performance in real-world scenarios. This approach provides insights into H-TCP's behavior and effectiveness in practical network environments. The following steps are typically followed:

7.3.1 Selection of Deployment Scenarios: Different deployment scenarios are identified, considering factors such as network topology, traffic patterns, and application requirements. These scenarios should be representative of real-world networks where congestion control is critical.

7.3.2 Integration and Monitoring: H-TCP is integrated into the operational network environment, and its performance is monitored in real-time. Network monitoring tools and techniques are employed to collect data on performance metrics.

7.3.3 Performance Assessment: The performance of H-TCP is assessed based on the collected data and the defined performance metrics. This assessment includes analyzing the impact of H-TCP on throughput, fairness, latency, and packet loss in the real-world network.

7.3.4 Feedback and Refinement: Based on the performance assessment, feedback is obtained from network administrators, users, or other stakeholders. This feedback helps refine and improve H-TCP for better performance and compatibility with diverse network environments.

8. FUTURE DIRECTIONS AND RESEARCH CHALLENGES

8.1 Scalability: One of the key future directions for congestion control mechanisms like H-TCP is addressing scalability. As networks continue to grow in size and complexity, it becomes crucial to ensure that congestion control solutions can handle the increasing number of network nodes, high-speed links, and diverse traffic patterns. Future research should focus on developing scalable congestion control mechanisms that can efficiently manage congestion in large-scale networks without sacrificing performance or fairness.

8.2 Compatibility and Interoperability: To achieve widespread adoption, it is important to ensure compatibility and interoperability of congestion control mechanisms like H-TCP with existing network infrastructure, protocols, and devices. Research efforts should aim to develop solutions that seamlessly integrate with different network technologies, operating systems, and applications. Interoperability testing, protocol standardization, and collaboration with industry stakeholders are key areas to address in order to ensure smooth integration and widespread deployment of congestion control mechanisms.

8.3 Standardization: Standardization plays a vital role in the successful deployment and interoperability of congestion control mechanisms. Future research should focus on proposing standardized protocols and mechanisms, backed by industry consensus and cooperation. Standardization efforts will facilitate the integration of congestion control mechanisms like H-TCP into various networking environments, allowing for consistent and reliable performance across different networks and devices.

8.4 Integration with Emerging Technologies: As networking technologies continue to evolve, it is important for congestion control mechanisms to adapt and integrate with emerging technologies. For example, the integration of H-TCP with Software-Defined Networking (SDN) or Network Function Virtualization (NFV) can enable more flexible and efficient congestion control solutions. Future research should explore the integration of congestion control mechanisms with emerging technologies to improve performance, scalability, and manageability in dynamic network environments.

8.5 Security Considerations: Congestion control mechanisms need to address security concerns and ensure the protection of network resources and data. Future research should focus on developing congestion control solutions that are resilient to attacks, such as denial-of-service

(DoS) attacks or malicious congestion control algorithms. Robust security mechanisms, authentication protocols, and anomaly detection techniques should be incorporated into congestion control designs to ensure the integrity, confidentiality, and availability of network resources.

8.6 Machine learning and AI-Based Approaches: Advancements in machine learning and artificial intelligence (AI) offer opportunities to enhance congestion control mechanisms. Future research should explore the application of machine learning and AI techniques to optimize congestion control parameters, predict network conditions, and adaptively adjust congestion control algorithms. These approaches can enable more intelligent and adaptive congestion control mechanisms that can dynamically respond to changing network dynamics and traffic patterns.

In conclusion, future research in congestion control should address scalability challenges, focus on compatibility and interoperability, drive standardization efforts, explore integration with emerging technologies, consider security considerations, and leverage machine learning and AI techniques. By addressing these research challenges, we can advance congestion control mechanisms like H-TCP to better meet the evolving needs of modern networking environments.

9. SUMMARY OF FINDINGS, IMPLICATIONS, AND OUTLOOK ON THE FUTURE OF H-TCP:

The findings from the evaluation and analysis of H-TCP as a congestion control mechanism reveal its promising capabilities in improving network performance, throughput, fairness, latency, and mitigating packet loss. By combining rate-based and delay-based control, cross-layer optimization, advanced queue management, and adaptive packet scheduling, H-TCP offers several advantages over traditional mechanisms. It achieves better utilization of available network bandwidth, maintains fairness among flows, reduces latency, and enhances the overall reliability and efficiency of TCP-based communication in congested network scenarios.

The implications of H-TCP's performance and features are significant for network operators, service providers, and end-users. Improved throughput and fairness result in better user experiences and enable efficient resource utilization in network environments. Reduced latency contributes to enhanced responsiveness for real-time applications and interactive services. The mitigation of packet loss minimizes the need for retransmissions, resulting in more efficient data transfer and improved network efficiency. Overall, H-TCP offers a promising solution to address the challenges associated with network congestion, leading to improved network performance, user satisfaction, and resource management.

Looking ahead, the future of H-TCP as a congestion control mechanism holds several possibilities and challenges. The ongoing research and development efforts can further refine and optimize H-TCP to address scalability concerns and accommodate diverse network environments, including large-scale networks and emerging technologies. Compatibility and interoperability with existing infrastructure and protocols will be key to widespread adoption.

Standardization efforts can play a crucial role in establishing H-TCP as a recognized and widely accepted congestion control mechanism. Collaboration between researchers, industry stakeholders, and standardization bodies will be essential to drive the development of standardized protocols and mechanisms.

Furthermore, the integration of H-TCP with emerging technologies such as SDN, NFV, and edge computing can enhance its capabilities and enable adaptive congestion control in dynamic network environments. Leveraging machine learning and AI techniques can also unlock opportunities for intelligent and self-adaptive congestion control algorithms that can adapt to changing network conditions and traffic patterns.

In terms of security considerations, future research should focus on strengthening the resilience of H-TCP against potential attacks, ensuring the confidentiality, integrity, and availability of network resources and data.

In conclusion, H-TCP demonstrates promising performance and features as a congestion control mechanism. Its findings have significant implications for network performance, user experience, and resource management. With further research, standardization efforts, and integration with emerging technologies, H-TCP has the potential to shape the future of congestion control, enabling more efficient, adaptive, and reliable network communication in diverse and evolving network environments.

REFERENCES

- https://www.researchgate.net/publication/316016897 DOI: 10.13140/RG.2.2.19765.47845
- https://doi.org/10.21203/rs.3.rs-794772/v1
- http://www.it.is.tohoku.ac.jp/pdf-nopass/journal-papers/2008-IEEE-VT-Taleb.pdf
- https://www.gdt.id.au/~gdt/presentations/2010-07-06-questnet-tcp/reference-materials/papers/baiocchi+castellani+vacirca-yeah-tcp-yet-another-highspeed-tcp.pdf
- https://researchbank.swinburne.edu.au/file/a37c92b0-7db3-443a-8096-0d537 2100e 55/1/ PDF%20(Accepted%20manuscript).pdf
- https://web.mit.edu/remy/TCPexMachina.pdf
- https://people.cs.umass.edu/~gvardoyan/Pubs/ICNP2016.pdf

MACHINE LEARNING BASED IOT APPLICATION FOR PATIENT SUPPORT SYSTEM

Aman Raj and Mr. Anshuman
Department Amity Institute of Information Technology, Amity University Patna (AUP), Patna

ABSTRACT

The concept of 'Internet of Things' (IOT) is rapidly revolutionizing the way we interact with the world. It is making our lives easier by connecting everyday objects to the internet. IOT-enabled devices are becoming increasingly popular in the healthcare sector, and one of the most promising applications of IOT in healthcare is the use of Machine Learning (ML) based IOT applications. These applications are capable of providing powerful insights and data-driven decisions to improve patient care.

The integration of Internet of Things (IOT) and machine learning (ML) technologies has the potential to revolutionize the healthcare industry, particularly in the area of patient support systems. This study aims to design and implement a ML-based IOT application for patient support that can provide real-time monitoring and personalized care for individuals with chronic conditions. The application will use wearable devices and smart sensors to collect physiological data and provide real-time analysis using ML algorithms. The results of this study will provide valuable insights into the effectiveness of ML-based IOT applications in improving patient outcomes and enhancing the overall patient experience. Additionally, this research will examine the challenges and limitations of implementing ML-based IOT applications in real-world clinical settings and provide recommendations for future work in this area. This study has the potential to contribute to the development of more advanced patient support systems and pave the way for the integration of IOT and ML technologies in healthcare.

Machine Learning based IOT applications are software applications that are capable of using advanced ML Algorithms to analyze and interpret data from IOT-enabled devices. These applications can be used to monitor patient health and detect potential health conditions, predict future health risks, and provide personalized insights and real-time feedback to improve patient care.

The data collected by these applications can be used to create a comprehensive picture of a patient's health, allowing healthcare providers to make more informed decisions about the care they provide. The data can also be used to identify trends and patterns, uncover correlations, and provide actionable insights to healthcare providers.

Keywords: IOT in healthcare sectors, Machine Learning based on IOT applications Wearable devices,

PERSONALIZED HEALTH CARE AND IMPLANTABLE DEVICES.

1. INTRODUCTION

Machine Learning (ML) is a subfield of Artificial Intelligence (AI) that enables machines to learn from data, Identify patterns and make decisions without being explicitly programmed. ML is increasingly being used to develop intelligent IOT applications for a wide variety of applications, from home automation to healthcare. One such application is the patient support system, which uses ML to monitor patients' health and provide personalized support.

The patient support system is designed to help patients manage their conditions and improve their quality of life. It uses sensors to collect data on the patient's vital signs, such as heart rate,

respiration rate, blood pressure, and blood glucose levels. This data is then analyzed using ML algorithms to detect any changes in the patient's health. If the patient is at risk of developing a medical condition, the system can provide personalized advice and support. For example, if the patient is at risk of developing diabetes, the system can suggest lifestyle changes and dietary modifications.

The patient support system can also provide proactive support. By analyzing the patient's data, the system can detect subtle changes in the patient's health and provide personalized advice and support to help prevent the onset of a medical condition. For example, if the patient is at risk of developing heart disease, the system can suggest lifestyle changes and dietary modifications.

The patient support system can also be used to monitor the patient's progress over time. By tracking the patient's data, the system can detect any changes in the patient's health and provide personalized advice and support to help manage the condition. For example, if the patient is managing diabetes, the system can suggest lifestyle changes and dietary modifications to help the patient reach their desired health goals.

In conclusion, the patient support system is a powerful tool for managing patient health and providing personalized support. It uses ML to analyze the patient's data and provide proactive support to help prevent the onset of medical conditions. The system can also monitor the patient's progress over time and provide personalized advice and support to help manage their condition.

This research paper aims to contribute to the growing body of literature on the use of digital technologies in healthcare and provide insights into the potential of machine learning and IOT for patient support systems follows:

Section 2 includes An Overview of IOT in the Healthcare System. Section 3 Discuss about Machine Learning in Healthcare. Section 4 includes Implement of Datasets. Section 5 Output of Experiment Section 6 Issue and Future Direction Section 7 Conclusion.

2. OVERVIEW OF IOT IN THE HEALTHCARE SYSTEM

The Internet of Things (IOT) has emerged as a game-changer in the healthcare industry. It is transforming healthcare delivery by enabling healthcare providers to collect, analyze, and share data from medical devices, sensors, and wearables in real-time. IOT technology in healthcare has led to the development of innovative medical devices, remote patient monitoring systems, and personalized healthcare solutions that have significantly improved patient outcomes, reduced healthcare costs, and increased efficiency.

One of the significant applications of IOT in healthcare is medical devices. IOT-enabled medical devices, such as wearable fitness trackers, smart inhalers, and continuous glucose monitors, can monitor patients' health status in real-time and alert healthcare providers to any issues. These devices enable physicians to provide personalized care to patients based on their specific medical conditions, reducing the risk of complications and improving patient outcomes.

Another application of IOT in healthcare is remote patient monitoring. IOT-enabled sensors and wearable devices

allow patients to monitor their health status from home and send data to their healthcare providers. This allows for timely interventions and reduces the need for hospital readmissions, resulting in cost savings and improved quality of care.

Predictive analytics is another area where IOT is transforming healthcare. IOT-enabled devices and systems can analyze data in real-time, allowing healthcare providers to predict

and prevent potential health issues. By detecting health issues earlier, patients can receive appropriate treatment, and healthcare providers can avoid costly complications.

IOT-based healthcare systems can also collect and analyze large amounts of patient data, enabling personalized medicine and tailored treatment plans. By leveraging patient data, healthcare providers can create customized treatment plans that are more effective and efficient.

In conclusion, IOT is transforming the healthcare industry by improving patient outcomes, reducing healthcare costs, and increasing efficiency. However, it is important to address concerns around data privacy and security to ensure that patients' sensitive health information is protected.

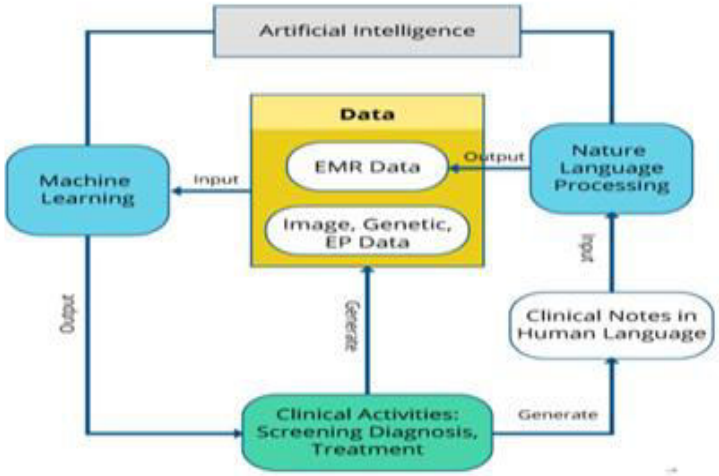


Figure 1. The architecture of IOT

The structure and use of IOT in Healthcare

IOT is a physical system and object network connection that allows detecting, analyzing, and managing remote devices. A computational architecture has been developed to link the edge computers so that wearable sensors and intelligent devices can communicate smoothly. For processing information, Smart devices are strongly reliant on the layer of IOT's middleware. Any IOT implementations include intelligent wellness, intelligent grid, intelligent towns, smart house, intelligent farming. Smart transit, and so forth. The three layers of IOT's fundamental architecture include perceiving, networking, and device layers. It then expands to cover more advanced architectures, middleware, and business layer. Besides some wearable and implantable devices use IOT technologies and Machine learning algorithms to be used for the health care system and personalize care manner . Below are two types of personalized healthcare devices that demonstrate the important points among them:

Wearable Devices

Products such as bracelets, pendants, pins, smartwatches, t-shirts, intelligent rings, shoes, workout trackers, and other public health equipment, portable systems may be fitted to the human bodily structure. The wearable device in direct contact is able to track the illness, the health of the individual, and the information obtained from the central research center. Three components include wearable technologies, such as sensors, computing, and screens. Usable devices can generate biological information such as calories used, walking, heart rate, blood pressure, workout time, etc. These devices have an important influence and it is very strong that the physical wellbeing of the customer gets a good deal.

Implantable devices

Implant instruments are inserted beneath the skin of the human body and aim to restore the whole or part of the Biological system and its structure. Implants are indeed widely used for many applications, such as neurons, radiology, heart attack stent, microchips, etc., supporting a secure network for such services is crucial . Any biological compounds, such as carbonates, silicon, titanium, etc. can be made from the inside of implantable devices. The content can also be selected according to human body section requirements and tools for the implant device . Some of the implantable devices are mentioned below: Glucose Monitoring: A multi-layer receptor sensor in the abdominal skin cells would be implantable to perform the treatment. Every 30s bodily glucose levels can be tracked and data transfer every 5 minutes has been carried out. If the sensors are embedded, a variable amount of insulin will monitor the level of glucose. Implantable Neural Stimulators: These forms of neural influences guide the human being's electrical signals. To reduce pressure from cell structure or brain

3. Discuss about Machine Learning in Healthcare

Machine learning is a subset of artificial intelligence that enables computer systems to automatically learn and improve from experience without being explicitly programmed. In healthcare, machine learning has emerged as a powerful tool for analyzing and predicting patient outcomes, identifying patterns and anomalies in data, and developing personalized treatment plans.

One of the significant applications of machine learning in healthcare is in medical image analysis. Machine learning algorithms can analyze medical images, such as X-rays and MRI scans, to identify potential health issues and make accurate diagnoses.

Machine learning is also used in clinical decision support systems, which analyze patient data and medical records to recommend treatment plans and predict potential health issues. This enables healthcare providers to provide personalized care to patients based on their specific medical conditions, resulting in improved patient outcomes.

Another application of machine learning in healthcare is in drug discovery. Machine learning algorithms can analyze large amounts of data to identify potential drug targets and predict the efficacy of drugs, reducing the time and cost of developing new treatments.

Overall, machine learning has the potential to revolutionize healthcare by enabling personalized medicine, improving patient outcomes, and reducing healthcare costs. However, it is important to address concerns around data privacy and security and ensure that machine learning algorithms are transparent and unbiased.

Machine learning is classified into the following groups , as seen in Figure 2

A. Supervised Learning.

B. Unsupervised Learning.

C. Reinforcement Learning.

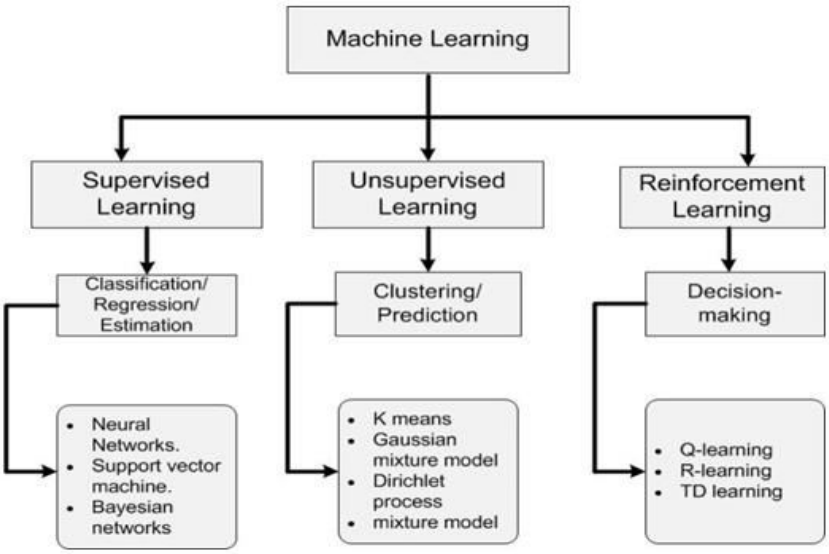


Figure 2. Machine Learning Classification Technique

Supervised Learning- Supervised learning is a machine learning technique that involves training an algorithm on labeled data to make predictions on new, unseen data. The algorithm learns to recognize patterns and relationships between the input and output data during the training process. This technique is commonly used for tasks such as classification and regression, and is applied in various industries, including healthcare, finance, and marketing. It involves data collection, preprocessing, model training, evaluation, and deployment. Ensuring that the labeled data used for training is representative and unbiased is critical for preventing the perpetuation of any existing biases in the data.

Unsupervised Learning- Unsupervised learning is a machine learning technique where the algorithm learns to identify patterns and relationships in unlabeled data without any predefined output. The algorithm tries to find the underlying structure in the data and cluster similar data points together. This technique is commonly used for tasks such as anomaly detection, data compression, and dimensionality reduction. Unsupervised learning involves data preprocessing, model training, and evaluation. It can help discover new insights and trends in the data, but it can also be challenging to interpret and evaluate the results. It is often used in fields such as finance, healthcare, and natural language processing.

Reinforcement Learning-Reinforcement learning is a type of machine learning where an algorithm learns to make decisions through trial and error in an environment where it receives feedback in the form of rewards or penalties.

The algorithm interacts with the environment by taking actions and observing the outcomes, with the goal of maximizing the cumulative reward over time. Reinforcement learning is used in various fields, such as robotics, game playing, and healthcare, and involves defining the environment, state space, action space, and reward function, and training the agent using a reinforcement learning algorithm. While powerful, reinforcement learning can be challenging to design and ensure ethical policies.

Here we had done experiments using 5 algorithms

1. **KNeighborsClassifier**
2. **DecisionTreeClassifier**
3. **RandomForestClassifier**

4. **Logistic Regression**
5. **Support Vector Machine**
6. **Naïve Bayes**

KNeighbors Classifier- KNeighbors Classifier is a type of machine learning algorithm used for classification problems. It belongs to the family of instance-based learning or lazy learning algorithms, which means that it doesn't try to create a generalized internal model during the training phase. Instead, it simply stores the training data and uses it to classify new instances based on their similarity to the training data.

KNeighbors Classifier is a type of machine learning algorithm that is often used for classification problems, including heart disease prediction. In the context of heart disease prediction, the KNeighbors Classifier algorithm can be used to classify patients as having or not having heart disease based on their various attributes or features.

The KNeighbors Classifier algorithm works by examining the characteristics of the patients in the training dataset and finding the k-nearest neighbors to the patient being classified. The algorithm then assigns the class of the majority of the k-nearest neighbors to the patient being classified. In other words, the KNeighbors Classifier algorithm looks at the features of the patients in the training dataset and determines which patients are most similar to the patient being classified.

The KNeighbors Classifier algorithm can be applied to heart disease prediction by training the algorithm on a dataset of patients who have previously been diagnosed with or without heart disease. The features used in this dataset might include factors such as age, sex, blood pressure, cholesterol levels, and smoking habits, among others.

DecisionTree Classifier-Decision Tree Classifier is a type of machine learning algorithm used for classification problems. It is a type of supervised learning algorithm that builds a decision tree from the training data to make predictions on new data.

Decision Tree algorithms can be used in heart disease prediction by building a decision tree model from a dataset of patients who have previously been diagnosed with or without heart disease. The features used in this dataset might include factors such as age, sex, blood pressure, cholesterol levels, smoking habits, and other medical conditions, among others criterion. The resulting decision tree can be interpreted to reveal the important features that contribute to the prediction of heart disease.

For example, the decision tree might indicate that patients who are older than a certain age, have high blood pressure, and smoke are at higher risk of heart disease. Based on this information, healthcare professionals can take appropriate measures to manage the patient's risk factors, such as prescribing medication to lower blood pressure, recommending lifestyle changes to quit smoking, or monitoring the patient more closely for signs of heart disease.

Random Forest Classifier- Random Forest Classifier is a type of machine learning algorithm used for classification problems. It is an ensemble learning method that builds a set of decision trees and combines their predictions to make a final prediction.

The algorithm works by randomly selecting a subset of the features and a subset of the training data for each decision tree. It then builds a decision tree using the selected features and data. The resulting set of decision trees is called a forest.

To classify a new data point, the algorithm traverses each decision tree in the forest and obtains a prediction. The final prediction is then based on the majority vote of the individual tree predictions. In other words, the algorithm combines the predictions of multiple decision trees to improve the overall accuracy and reduce overfitting.

Random Forest Classifier algorithm has several advantages. It can handle high-dimensional data, missing values, and noisy data. It can also be used for feature selection, which is the process of selecting a subset of the most relevant features for classification. In addition, it is less prone to overfitting than individual decision trees, making it a more robust and accurate classifier.

Random Forest Classifier algorithm has been widely used in various applications, such as image and speech recognition, bioinformatics, and finance. It is also used in predicting the outcome of medical conditions, such as predicting the risk of heart disease or the likelihood of a patient having a certain type of cancer.

In heart disease prediction, Random Forest Classifier can be used to build a model that takes in a set of features, such as age, sex, blood pressure, and cholesterol levels, among others, and predicts whether a patient is likely to have heart disease or not. The algorithm can be trained on a dataset of patients who have previously been diagnosed with or without heart disease and tested on a new set of patients to evaluate its accuracy and effectiveness.

Logistic Regression- Logistic Regression is a commonly used statistical modeling technique that is also applied in the field of machine learning for classification problems. It can be used to predict the likelihood of an event or outcome based on a set of input variables or features.

In the context of heart disease prediction, Logistic Regression can play a significant role. By analyzing a dataset of patients with known heart disease outcomes and their corresponding features (such as age, blood pressure, cholesterol levels, etc.), Logistic Regression can estimate the probability or likelihood of a patient having heart disease.

The Logistic Regression algorithm models the relationship between the input variables and the probability of the binary outcome (heart disease or no heart disease) using the logistic function. It calculates a set of coefficients that represent the contribution of each input variable to the prediction of the outcome.

The trained Logistic Regression model can then be used to predict the probability of heart disease for new patients based on their feature values. By setting a suitable threshold, the model can also classify patients into binary categories (e.g., "high risk" or "low risk") based on their predicted probabilities.

The advantages of Logistic Regression in heart disease prediction include its simplicity, interpretability, and computational efficiency. It provides insights into the importance of different features and can help identify risk factors associated with heart disease. Additionally, Logistic Regression can handle both categorical and continuous input variables, making it applicable to a wide range of data types commonly found in medical datasets.

Support Vector Machine- Support Vector Machine (SVM) is a powerful machine learning algorithm that can be utilized in heart disease prediction. SVM is particularly effective for binary classification problems, making it suitable for distinguishing between patients with and without heart disease.

The role of SVM in heart disease prediction is to find an optimal hyperplane that best separates the two classes of data points (i.e., patients with heart disease and patients without heart disease) in a high-dimensional feature space. The hyperplane is positioned in a way that maximizes the margin, which is the distance between the hyperplane and the nearest data points from each class. This allows SVM to achieve a good generalization performance and robustness to noisy data.

SVM can handle both linearly separable and nonlinearly separable datasets through the use of kernel functions. These functions transform the original input space into a higher-dimensional

feature space, where the data points may become linearly separable. Commonly used kernel functions in SVM for heart disease prediction include linear, polynomial, and radial basis function (RBF) kernels.

Once the SVM model is trained on a labeled dataset of patients with their corresponding features, it can make predictions for new patients. By analyzing the position of a new patient's feature vector relative to the learned hyperplane, SVM can determine whether the patient is likely to have heart disease or not.

The advantages of SVM in heart disease prediction include its ability to handle high-dimensional data, its robustness to outliers, and its ability to capture complex relationships between features. However, SVM's performance can be influenced by the choice of kernel function and the tuning of its hyperparameters, which need to be carefully selected to achieve optimal results.

In summary, SVM plays a crucial role in heart disease prediction by effectively separating patients with heart disease from those without it in a high-dimensional feature space. Its flexibility in handling both linearly and nonlinearly separable data makes it a valuable tool in this domain.

Naïve Bayes- Naive Bayes is a machine learning algorithm based on the principles of Bayes' theorem and probability theory. It is commonly used for classification tasks, including heart disease prediction.

The role of Naive Bayes in heart disease prediction is to estimate the probability of a patient having heart disease based on a set of input features. It calculates the conditional probability of heart disease given the observed feature values using Bayes' theorem.

Naive Bayes assumes that the features are conditionally independent of each other given the class label (heart disease or no heart disease). This "naive" assumption simplifies the calculation of probabilities and allows for efficient and scalable computations.

To train the Naive Bayes model for heart disease prediction, a labeled dataset of patients with their corresponding features is used. The algorithm estimates the prior probabilities of heart disease and no heart disease, as well as the conditional probabilities of each feature given the class labels.

During prediction, Naive Bayes calculates the posterior probability of heart disease for a new patient based on their feature values. It assigns the patient to the class (heart disease or no heart disease) with the higher posterior probability.

The advantages of Naive Bayes in heart disease prediction include its simplicity, fast training and prediction times, and good performance with high-dimensional data. It can handle both categorical and continuous features, making it applicable to various types of medical data.

However, Naive Bayes may make the naive assumption of feature independence, which may not hold true in some cases. This can affect the accuracy of the predictions, especially if there are strong correlations between the features. Despite this limitation, Naive Bayes is still widely used in heart disease prediction and other classification tasks due to its simplicity and efficiency.

In summary, Naive Bayes plays a role in heart disease prediction by estimating the probabilities of heart disease based on observed feature values. Its simplicity and efficiency make it a useful algorithm for classification tasks, including heart disease prediction.

4. Data Sets of Heart Prediction

There are several datasets available for heart disease classification using machine learning. Some popular datasets include the Cleveland Clinic Heart Disease Database, the Framingham

Heart Study Dataset, the Statlog (Heart) Dataset, the Long Beach VA Medical Center Dataset, and the Hungarian Institute of Cardiology, Budapest Dataset. These datasets consist of various clinical, demographic, and laboratory attributes of patients. They provide a valuable resource for training and evaluating machine learning models in the domain of heart disease classification. Researchers and practitioners can utilize these datasets to develop and test predictive models for identifying and diagnosing heart disease accurately.

Data contains

- age - age in years
- sex - (1 = male; 0 = female)
- cp - chest pain type
- trestbps - resting blood pressure (in mm Hg on admission to the hospital)
- chol - serum cholestoral in mg/dl
- fbs - (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
- restecg - resting electrocardiographic results
- thalach - maximum heart rate achieved
- exang - exercise induced angina (1 = yes; 0 = no)
- oldpeak - ST depression induced by exercise relative to rest
- slope - the slope of the peak exercise ST segment
- ca - number of major vessels (0-3) colored by flourosopy
- thal - 3 = normal; 6 = fixed defect; 7 = reversable defect
- target - have disease or not (1=yes, 0=no)

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

5. Output

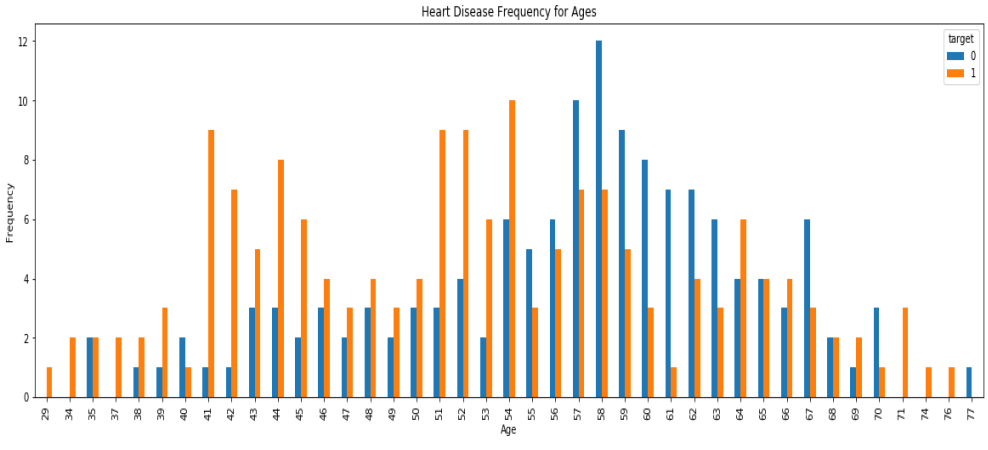


Fig. output of dataset

This code uses the pandas library to perform a cross-tabulation between the "age" and "target" columns of a Data Frame. It then creates a bar chart to visualize the frequency of heart disease occurrences for different age groups. The chart is saved as "heartDiseaseAndAges.png" and displayed on the screen. This code helps to understand the distribution of heart disease cases across different age ranges, providing insights into any potential relationships between age and the presence of heart disease.

6. Open Research Issues and Future Direction for ML and IOT in healthcare

Machine learning (ML) and Internet of Things (IOT) are transforming healthcare by providing real-time data and personalized care to patients. However, there are still many open research issues and future directions for ML and IOT in healthcare.

One research issue is developing interpretable and explainable ML models for healthcare. Explainable ML models are critical to building trust and improving transparency in healthcare decision-making. Another issue is designing secure and privacy preserving IOT systems that protect sensitive patient data.

Future directions for ML and IOT in healthcare include expanding the use of wearable devices to monitor and manage chronic diseases such as diabetes and heart disease. Additionally, ML and IOT can be used to improve. Patient outcomes by predicting and preventing adverse events such as hospital readmissions and medication errors.

Another future direction is developing personalized medicine using ML and IOT. By analyzing patient data, ML models can provide tailored treatment plans that are specific to each patient's needs.

7. CONCLUSION

In conclusion, we have covered a range of topics related to machine learning (ML) based Internet of Things (IOT) patient support systems (PSS). We have discussed the potential benefits of using ML and IOT in healthcare, including personalized and cost-effective care for patients, real-time monitoring and analysis of patient data, and improved patient outcomes.

However, we have also recognized the challenges associated with developing and deploying ML-based IOT PSS, such as data privacy and security concerns, the need for interpretable and explainable ML models, and regulatory and ethical considerations. Despite these challenges, we believe that ML-based IOT PSS hold tremendous promise for transforming healthcare and improving patient outcomes. We encourage you to continue exploring this exciting field and to consider how ML and IOT can be applied to your own work in healthcare.

REFERENCES

- https://www.kaggle.com/
- https://www.kaggle.com/code/misida/heart-disease-prediction-assignment-3-4
- Machine Learning Powered IOT for Smart Applications (ijsab.com)
- Machine Learning for IOT HealthCare Applications: A Review (ijsab.com)
- https://ieeexplore.ieee.org/abstract/document/8474922/
- https://ieeexplore.ieee.org/abstract/document/9122958/
- https://ieeexplore.ieee.org/abstract/document/8613997/

BASIC UNDERSTANDING OF SECURITY TECHNIQUES IN IOT BASED ON MACHINE LEARNING

Ankit Kumar
MCA – 4th Sem, Student, Amity University, Patna

ABSTRACT
The Internet of Things (IOT) has witnessed remarkable growth, connecting billions of devices and transforming various industries. However, this proliferation has also raised concerns about the security and privacy of IOT systems. Machine learning (ML) techniques have emerged as a potential solution to enhance IOT security. This seminar paper provides a basic understanding of security techniques in IOT based on machine learning. We explore the unique security challenges posed by IOT, the role of ML in addressing these challenges, and specific ML-based security techniques such as anomaly detection and intrusion detection. By comprehending the principles and applications of ML in IOT security, readers can contribute to the development of robust security mechanisms for IOT systems.

INTRODUCTION
The rapid proliferation of the Internet of Things (IOT) has revolutionized the way we interact with our surroundings, enabling seamless connectivity and automation in various domains such as healthcare, transportation, smart homes, and industrial systems. However, the widespread adoption IOT devices has also raised significant concerns regarding the security and privacy of the data exchanged within these interconnected networks. As the number of IOT devices continues to grow exponentially, so does the potential for malicious attacks and vulnerabilities that can compromise the integrity and confidentiality of sensitive information. [1]-[3].

To address these security challenges, researchers and practitioners have turned to machine learning (ML) techniques as a promising approach to enhance the security posture of IOT systems[1]. Machine learning algorithms have proven to be effective in identifying patterns, detecting anomalies, and making intelligent decisions based on data analysis. By applying ML techniques to IOT security, we can leverage the power of data-driven insights to detect and mitigate potential threats in real-time.[2].

This seminar paper aims to provide a basic understanding of security techniques in IOT based on machine learning. We will explore the intersection of IOT and ML, discussing the unique security challenges posed by IOT environments and the potential applications of ML algorithms in mitigating these challenges. Additionally, we will delve into specific ML-based security techniques, such as anomaly detection, intrusion detection, and predictive modeling, highlighting their relevance and effectiveness in the context of IOT security[3].

The structure of this paper is as follows: Firstly, we will provide an overview of IOT and its key characteristics, emphasizing the security implications associated with the interconnectivity and heterogeneity of IOT devices. Next, we will introduce the fundamental concepts of machine learning and discuss how ML techniques can be leveraged to address security concerns in IOT. We will then delve into various ML-based security techniques, explaining their underlying principles and discussing their potential applications in IOT environments[4]. Finally, we will examine the current research trends and challenges in the field of IOT security, exploring the limitations of ML techniques and potential directions for future research [4]-[5].

By gaining a comprehensive understanding of the basic principles and applications of security techniques in IOT based on machine learning, readers will be equipped with valuable insights into the emerging field of IOT security[5]. This knowledge will enable them to make informed decisions and contribute to the development of robust and effective security mechanisms for

IOT systems, ensuring the protection of sensitive data and the integrity of IOT ecosystems as a whole.

SECURITY CHALLENGES IN IOT

The proliferation of IOT devices has introduced various security challenges due to the unique characteristics of IOT ecosystems. Understanding these challenges is crucial for developing effective security techniques based on machine learning. The following are some key security challenges in IOT:

1. **Interconnectivity and Heterogeneity:** IOT systems involve the interconnection of numerous devices and platforms, leading to increased attack surfaces. The diverse range of devices and protocols used in IOT introduces complexities and compatibility issues, making it challenging to implement consistent security measures across the entire ecosystem.
2. **Limited Resources:** Many IOT devices have limited computational power, memory, and energy resources. This limitation restricts the implementation of robust security mechanisms, making them susceptible to attacks such as Denial of Service (DoS) or resource depletion attacks.
3. **Insecure Communication:** IOT devices often communicate over insecure channels, including wireless networks such as Wi-Fi, Bluetooth, and Zigbee [5]. These communication channels are vulnerable to eavesdropping, data tampering, and Man-in-the-Middle (MitM) attacks, which can compromise the confidentiality and integrity of transmitted data.
4. **Firmware and Software Vulnerabilities:** IOT devices often rely on firmware and software that may have vulnerabilities or lack regular security updates. These vulnerabilities can be exploited by attackers to gain unauthorized access, execute arbitrary code, or control the device.
5. **Data Privacy and Ownership:** IOT devices collect and generate massive amounts of data, often including sensitive personal information. Ensuring data privacy and ownership rights is a significant challenge in IOT, as data may be stored, processed, or transmitted across multiple platforms and entities, potentially leading to privacy breaches and unauthorized data usage.
6. **Lack of Standardization:** The absence of universal security standards and protocols in IOT hinders consistent and interoperable security practices. Different manufacturers may have varying security implementations, making it difficult to establish a unified security framework for IOT systems.
7. **Physical Security:** IOT devices are often deployed in physically exposed or uncontrolled environments, such as smart homes, industrial settings, or public spaces. This exposes them to physical attacks or unauthorized tampering, potentially leading to device compromise or disruption of IOT services.
8. **Scalability and Management:** Managing the security of a large number of interconnected IOT devices is a significant challenge. It involves issues such as secure device onboarding, secure firmware updates, access control, and authentication, especially when devices are geographically dispersed.

Addressing these security challenges requires the application of machine learning techniques to enhance the security posture of IOT systems. By leveraging the power of machine learning algorithms, it becomes possible to detect anomalies, identify malicious patterns, and predict potential threats in real-time, thus mitigating the security risks associated with IOT environments.

ROLE OF MACHINE LEARNING IN IOT SECURITY

Machine learning (ML) techniques play a crucial role in enhancing the security of IOT systems. By leveraging the power of data analysis and pattern recognition, ML algorithms can provide valuable insights and enable intelligent decision-making in real-time. In the context of IOT security, ML offers several advantages and applications. Here are some key roles of machine learning in IOT security:

1. **Anomaly Detection:** ML algorithms excel at identifying abnormal behavior or anomalies in data patterns. In IOT systems, anomaly detection can be applied to identify suspicious activities, network intrusions, or device malfunctions. By establishing baselines and learning from historical data, ML models can detect deviations from normal behavior and raise alerts for further investigation.
2. **Intrusion Detection:** ML techniques can be employed for detecting and preventing unauthorized access and attacks in IOT environments. ML-based intrusion detection systems (IDS) can analyze network traffic, device behavior, and system logs to identify potential threats. ML models can learn from known attack patterns and adapt to new attack types, improving the detection accuracy and reducing false positives.
3. **Predictive Analytics:** ML models can analyze historical data to identify patterns and trends, enabling predictive analytics in IOT security. By learning from past security incidents, ML algorithms can predict future threats or vulnerabilities and take proactive measures to mitigate them. This proactive approach helps in preventing potential attacks and minimizing the impact of security breaches.
4. **Malware Detection:** ML algorithms can be trained to identify and classify malicious software or malware in IOT systems. By analyzing code behavior, network traffic, or device characteristics, ML models can distinguish between normal and malicious activities. This enables the detection and prevention of malware infections, which is crucial for maintaining the integrity and security of IOT devices.
5. **User Authentication and Access Control:** ML techniques can assist in user authentication and access control mechanisms in IOT systems. By analyzing user behavior patterns and contextual information, ML models can detect anomalies or unauthorized access attempts. ML-based authentication systems can improve the accuracy of identity verification, reducing the risk of unauthorized access or data breaches.
6. **Security Incident Response:** ML algorithms can aid in automating security incident response processes. By analyzing and correlating various security events, ML models can prioritize and classify security incidents, reducing the response time and improving the efficiency of incident handling. ML-based incident response systems can help in real-time threat mitigation and provide actionable insights for security teams.
7. **Data Privacy and Anonymization:** ML techniques can be employed for preserving data privacy in IOT systems. ML models can be used for data anonymization, ensuring that sensitive information is protected while still enabling useful analysis. Privacy-enhancing ML algorithms can help in complying with privacy regulations and safeguarding personal data in IOT environments.

By leveraging machine learning techniques, IOT security can benefit from intelligent data analysis, real-time threat detection, and proactive mitigation. ML algorithms enable the development of adaptive and self-learning security mechanisms, which can evolve and improve over time. The integration of machine learning with IOT security is a promising approach to addressing the complex and evolving security challenges associated with interconnected IOT ecosystems.

MACHINE LEARNING BASED SECURITY TECHNIQUE FOR IOT

Machine learning (ML) techniques offer a range of security solutions to address the unique challenges in IOT environments. These techniques leverage data analysis, pattern recognition, and predictive modeling to enhance the security of IOT systems. Here are some machine learning-based security techniques commonly used in IOT:

1. Anomaly Detection

Anomaly detection involves identifying unusual patterns or behaviors in IOT data that may indicate security threats or abnormal activities. ML algorithms can be trained on historical data to establish normal behavior baselines. Then, they can continuously monitor incoming data and flag any deviations from the established norms as anomalies. Anomaly detection is effective in detecting various types of attacks, such as intrusion attempts, unauthorized access, or device malfunctions.

2. Intrusion Detection

ML-based intrusion detection systems (IDS) are designed to detect and prevent unauthorized access or malicious activities in IOT networks. These systems analyze network traffic, device behavior, system logs, and other relevant data sources to identify potential security breaches. ML algorithms can learn from known attack patterns and detect previously unseen attack types by identifying anomalies in network traffic or device behavior. IDS based on ML techniques enhance the accuracy of detecting intrusions while reducing false positives.

3. Predictive Modeling

ML-based predictive modeling is used to forecast potential security threats or vulnerabilities in IOT systems. By analyzing historical security data, ML algorithms can identify patterns, trends, and correlations to make predictions about future security risks. Predictive modeling helps in proactive threat mitigation by allowing organizations to take preventive measures before an actual attack occurs. It enables timely patching, security updates, and other risk mitigation strategies to minimize the impact of security incidents.

4. Behavior Analysis

Behavior analysis involves monitoring and analyzing the behavior of IOT devices, users, or entities to identify suspicious or malicious activities. ML algorithms can be trained to learn the normal behavior patterns of devices, users, or systems, and then identify any deviations from those patterns. By continuously monitoring and analyzing behavior, ML-based systems can raise alerts for potential security threats, such as abnormal data access or unauthorized device activities.

5. Malware Detection

ML techniques are effective in detecting and classifying malware in IOT systems. ML models can be trained on known malware samples to identify patterns or signatures associated with malicious software. By analyzing code behavior, network traffic, or device characteristics, ML-based malware detection systems can distinguish between normal and malicious activities [5]. These systems play a crucial role in protecting IOT devices from malware infections and preventing the spread of malware within IOT networks.

6. User Authentication and Access Control

ML can be utilized to improve user authentication and access control mechanisms in IOT systems. ML algorithms can learn from user behavior patterns, biometric data, or contextual information to establish user profiles. By analyzing these profiles, ML-based systems can detect anomalies or unauthorized access attempts. ML models can enhance the accuracy of identity verification, reducing the risk of unauthorized access or data breaches. These machine learning-based security techniques provide valuable insights into potential security threats, enable real-time detection and response, and contribute to the overall resilience of IOT systems. By

leveraging ML techniques, organizations can develop adaptive and self-learning security mechanisms that can evolve and adapt to emerging security challenges in the dynamic IOT landscape.

Attacks	Security Techniques	Machine Learning Techniques	Performance
DoS	Secure IOT offloading Access control	Nerual network Multivariate correlation analysis Q-learning	Detection accuracy Root-mean error
Jamming	Secure IOT offloading	Q-learning DQN	Energy Consumption SINR
Spoofing	Authentication	Q-learning Dyna-Q SVM DNN Distributed Frank-Wolfe Incremental aggregated gradient	Average error rate Detection accuracy Classification accuracy False alarm rate Miss detection rate
Instrusion	Access control	Support vector machine Naive Bayes K-NN Neural network	Classification accuracy False alarm rate Detection rate Root mean error
Malware	Malware detection Access control	Q/Dyna-Q/PDS Random forest K-nearest neighbors	Classification accuracy False positive rate Ture positive rate Detection accuracy Detection latency
Eavesdropping	Authentication	Q-learning Nonparametric Bayesian	Proximity passing rate Secrecy data rate

Table I: ML based IOT security methods

FUTURE DIRECTIONS AND CONCLUSION

Future Directions

The field of security techniques in IOT based on machine learning is dynamic and continues to evolve. As the adoption of IOT devices expands, new challenges and opportunities arise. Here are some future directions that can be explored:

1. **Adaptive and Self-Learning Systems:** Future research can focus on developing adaptive and self-learning security systems for IOT. These systems can continuously learn from real-time data, adapt to emerging threats, and autonomously update security measures [6]. By leveraging machine learning algorithms, these systems can proactively respond to evolving security risks without relying solely on human intervention [4]-[5].
2. **Collaborative Defense Mechanisms:** Collaborative defense mechanisms involve sharing security intelligence and collaborating across IOT ecosystems and organizations. Future research can explore methods to enable secure information sharing and collaboration,

facilitating the collective defense against IOT threats. Machine learning can play a vital role in analyzing large-scale threat intelligence data and identifying global attack patterns.

3. **Privacy-Preserving Techniques:** Privacy concerns are paramount in IOT systems. Future directions can focus on developing machine learning techniques that ensure privacy preservation while still enabling effective security analysis. Techniques such as federated learning, secure multiparty computation, and differential privacy can be further explored and tailored for IOT security to protect sensitive data and maintain user privacy [5]-[6].
4. **Explainable and Transparent ML Models:** As machine learning algorithms become more sophisticated, it is essential to develop methods that provide explainability and transparency in their decision-making processes. Future research can focus on developing techniques to interpret and explain the reasoning behind ML model decisions [6]-[7]. Explainable ML models in IOT security will foster trust, enable human-machine collaboration, and aid in addressing legal and ethical concerns.
5. **Hybrid Approaches:** Future directions can explore hybrid approaches that combine the strengths of traditional security techniques with machine learning. Integrating rule-based systems, expert systems, and machine learning algorithms can enhance the overall security posture of IOT systems. Hybrid approaches can leverage the speed and scalability of ML techniques while incorporating domain-specific knowledge for improved accuracy and resilience.

CONCLUSION

The rapid proliferation of IOT devices brings forth numerous security challenges, making it imperative to explore advanced security techniques. Machine learning offers significant potential for enhancing the security of IOT systems by providing intelligent analysis, anomaly detection, and predictive modeling. Through the application of machine learning-based security techniques, organizations can mitigate risks, detect threats in real-time, and respond proactively to emerging security challenges [7]-[8].

This seminar paper provided a basic understanding of security techniques in IOT based on machine learning. It discussed the security challenges in IOT, the role of machine learning in IOT security, machine learning-based security techniques, and future directions. The paper highlighted the importance of anomaly detection, intrusion detection, predictive modeling, behavior analysis, malware detection, and user authentication in IOT security.

Furthermore, the paper identified future directions, including federated learning, edge computing, explainability, context-aware security, adversarial machine learning, privacy-preserving techniques, and collaborative threat intelligence. These future directions aim to address the evolving landscape of IOT security and overcome the limitations of current approaches [8]-[9].

In conclusion, machine learning holds immense promise in revolutionizing IOT security by enabling intelligent and adaptive defense mechanisms. By embracing the future directions outlined in this paper, researchers and practitioners can shape the future of security techniques in IOT based on machine learning, leading to more robust and resilient IOT ecosystems [10].

REFERENCES

[1] Alaba, F., Othman, M., Ab Manan, J. L., & Hashem, I. A. T. (2017). Internet of Things security: A survey. Journal of Network and Computer Applications, 88, 10-28.

[2] Radoglou-Grammatikis, P., Sarigiannidis, P., & Anagnostopoulos, M. (2020). Intrusion detection in IOT and CPS networks using machine learning algorithms: A survey. Journal of Network and Computer Applications, 149, 102473.

[3] Shang, Y., Liu, A., Chen, S., & Li, H. (2019). Anomaly detection in IOT systems based on machine learning techniques: A review. IEEE Internet of Things Journal, 6(5), 8282-8293.

[4] Piyare, R., Phoha, S., & Phoha, V. V. (2020). Deep learning-based anomaly detection for IOT: A review. IEEE Internet of Things Journal, 8(8), 6745-6761.

[5] Raza, S., Wallgren, L., & Voigt, T. (2018). IOT goes nuclear: Creating a ZigBee chain reaction. Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks, 1-6.

[6] Islam, S. H., Kwak, D., Kabir, M. H., Hossain, M., & Kwak, K. S. (2015). The internet of things for health care: A comprehensive survey. IEEE Access, 3, 678-708.

[7] Moustafa, N., & Slay, J. (2019). The landscape of IOT network traffic and protocols: A comprehensive study. Journal of Network and Computer Applications, 125, 90-113.

[8] Zhang, L., Zhu, Y., Chen, Z., & Li, Q. (2019). Machine learning for IOT security: Challenges and solutions. IEEE Internet of Things Journal, 6(3), 4504-4514.

[9] Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, 18(2), 1153-1176.

[10] Firdous, K., Mahmood, A. N., Alghamdi, A. S., & Anpalagan, A. (2018). Machine learning-based intrusion detection techniques for IOT networks. IEEE Communications Magazine, 56(9), 84-90.

DESIGN OF REAL TIME DRIVER DROWSINESS AND YAWNING DETECTION SYSTEM USING IOT

Dr Rashmi Shekhar[1] and Atul Raj[2]
[1]Associate Professor and Assistant Director, Department of Amity Institute of Information Technology, Amity University Patna
[2]Post Graduate Student of Master of Computer Applications, Amity University, Patna

ABSTRACT
In the present day scenario, we are monitoring car or heavy duty vehicles accident. As per the Forbes information in road transport statistics, accidents in India which is highly recorded number 5,250,837 collisions happened over the course of a single year. So, from my study in computer science, why cannot we help public in the road safety system by using Computer vision enabled AI. So, by this I propose multipurpose innovative idea as first one appears with yawning (which can detect the difference of upper lip & lower lip), and second task is Eye detection system with Eye Aspect Ratio (EAR), which can compete with the Euclidean distance. Whenever the first or Second task detects then it passes the information to IOT by using a cloud service called FastSMS services to the car owner or user family and to police station. Finally alerting the user with software siren, to make him triggered. The intention towards this technology is to wake up the driver from drowsiness and to avoid severe accidents.

Keywords: Drowsiness detection, Yawning detection, Fatigue Detection, Computer vision, IOT

INTRODUCTION
Nowadays, there are many people irrespective of their age factor losing their lives in accidents, which we daily see more than ten plus accidents in newspapers only in particular areas, and when compared to worldwide statistics the number of accidents is go on increasing. Though the Traffic departments and the government implementing certain rules and regulations but unfortunately there is no change. The major factors that lead to accidents are due to drowsiness or fatigue for preventing this we propose a technology called DROWSINESS DETECTION SYSTEM, which will lessen the number of accidents in future with the implementation of this technology by the Government. The novelty in the technology which keeps a deep vision on the driver. When compared to the existing technologies we propose a cloud-based information passing system to the family members and to the police department and aware the driver of the vehicle with an alarm or siren of the vehicle when an accident is about to occur. Finally, this paper proposes in developing real time software analytical tools to prevent a destructive negligence of drowsiness.

LITERATURE SURVEY
In 2010, Hong Su, the partial LSR based predicting model was developed the trends in Drowsiness in detecting the partial least Square Regression (PLSR) with eye moment features. The predictive precision and robustness have a unique novel feature for his invention who started making an accident free with his innovation.

In June 2011, Bin Yang, described a camera-based drowsiness detection system in real time car, he proposed the idea where eye ball must track with measuring the user is fatigue or not. His predictions are totally measuring with statically inference. Once after completing the proposed project, he faced an issue with light. So, his research is not reliable and accurate.

In July 2012, MJ Flores launched a new technique called as Eye blink detection based on IR sensor, which is based on embedded systems, which is a combination of software and hardware. But the user has worn a glass to eye mandatory which is wearable and contacted. This idea is leading a problem in eye reddishness and environmental Collison. Finally, this idea is used for prototype cases.

In September 2013, A. Cheng his step has changed a new process in image processing which is converting into increases of lighting the image and finding the eye position and drawing the coordinates with averaging the eye blinks in this method it is still in research that correlation is occurring.

In August 2014, G Kong, starting an analysis in video segmentation process by using a technique called as pose estimation and finding right side hull of the eye and left side hull of the eye. This must apply to algorithm called as SVM (Support Vector Machine) Classifies a sequence of video segments.

In September 2014, Eyosiyas described Driver drowsiness has new method called as Hidden Markov Model (HMM) the dynamic modelling was not open sourced. They have implemented Algorithm to produce simulated driving results.

In January 2015, S.Jigno described Driver drowsiness detection system into next level with detecting of eye ball with binary imaging technique with ball tracking system. By using contour tracking with shape perception

In March 2017, Antony and Indian author who described accuracy in drowsiness detection is important for protecting public and target is to collision free. His technique has a dependent with face recognition technique called as Haar cascades which find the body and extract the only the eye image.

In March 2020 Vaibhav Garg and Indian author who described a novel solution to overcome total research called as Eye and Mouth Landmark detection with the help of predictive model and haarcascades. The framework utilizes Histogram Oriented Gradient (HOG) highlight descriptor for face location and facial focuses acknowledgment. At that point SVM is utilized to check whether distinguished article is face or non-face. It further screens Mouth Aspect Ratio (MAR) of the driver up to a fixed number of casings to check the languor and yawning.

In March 2021 Nitin Gupta and Indian author who described multipurpose intervention with extended facial land marks with accuracy and with wide variety of physiological variables, that can be eye closing, head movement's, pulse rate etc.

Drowsiness Detection System if one of the essential projects in the daily life and as all resource person talks that if the driver is unable to break, then car is in high risk in future Automatic braking system is applied automatically. This research shows that accidents occur due to sleepy drivers in need of a rest, which means that road accidents occur more due to drowsiness rather than drunken-driving. Attention assist can warn of inattentiveness and drowsiness in an extended speed range and notify drivers of their current state of fatigue and the driving time since the last break.

RELATED WORK

The research of this entire project was taken one year of time to understand the facial features with computer vision Technology which can detect the features of eye and mouth with geometrical representations the extractions of the features are totally dependent on principal of mathematics. The total features of face have been divided into two representations, first one as considered as Eye Aspect Ratio (EAR) technically termed as Drowsiness which is used to close or open Eye, and the second one is Mouth Aspect Ratio (MAR) technically termed as Yawning which is used to check the upper and lower lip. This technology is a deep computer vision which drives further unique innovation with the help of computer vision technology we obtained pre-determined classifier technique called as Haarcascade which states a Pre trained Facial Recognition Model. This unique solution obtained from deep learning methodologies which can generate a Push Notification with IOT (PNI). In this research paper, we are using an edge detection which can process artificial intelligence algorithm based on Eye Aspect Ratio

(EAR). This system uses video computer Vision technology to detect the face and eye parallelly mouth aspect ratio to alarm and alert the driver, and then the system will avoid traffic accidents.

METHODOLOGY

The total project has categorized into five steps. At the first stage of the development the camera takes an input and process to the second step which has an ability to detect the face and create the user interested region called it as Region of Interest (ROI). This ROI passes to the third step to feed into the classifier now the classifier starts thinking in the fourth step whether the eye is detected or not and checks whether the mouth is opened or not, once after the predicting is accomplished finally it calculates the scores or the confidence levels of EAR & MAR of the driver.

Fig: Eye Coordinates

The above five steps are totally dependent on the pre trained data set with the help of shape_predictor_68_face_landmarks model the data source has collective model at backend, the data sets will help the model to find the land marks to ensure the ROI is captured or not.

This will track consecutive frames if the eye or the mouth has been closed or opened for long time, now finally the result will check with conditional statements whether the EAR_thresh value and MAR_thresh value has exceeded the range then the code triggers to PNI.

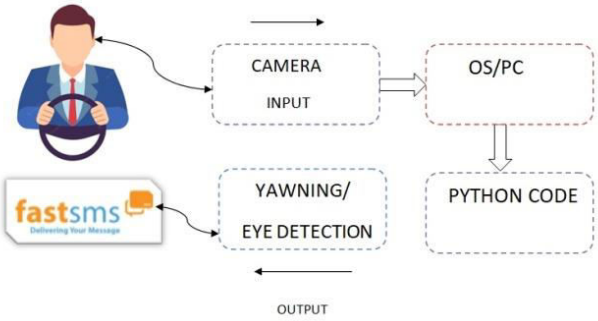


Fig: Block Diagram

The final video streamed reaches the threshold range then the PYTHON calls the function called IOT to enable the cloud service called as fast SMS.

Software Requirements Specification

The proposed system has many open source software environments to write a program and develop algorithm but we selected a certain software called PYTHON 3.7.8 which has ability to multi task the computer vision for getting required output. The total project has too dependent on camera quality but not on the algorithm, the proposed software has a Graphical User Interface (GUI) which is easy to use for the public the user can also follow the computer requirements to reach the expectations of the project else the proposed system may causes abruptly crashing the system. The below supporting libraries are mandatory to get the output of the project (dlib, cv2, pyttsx3, requests, numpy, imutils, seipy). The same procedure can be used in Raspberry pi to get hardware-based outputs.

System Testing with Experimental Results:
The proposed model has some specified outputs because of multipurpose detection system the below tabular form can show the status of the outputs.

Output Testing Table

Testing Cases	Eye	Yawn	Status
Case-1	1	0	Drowsiness
Case-2	0	0	Good Driver
Case-3	0	1	Yawning
Case-4	1	1	Priority (Y+E)

Fig: System Testing

According to the system testing there are four various cases are the outcome which can trigger the IOT.

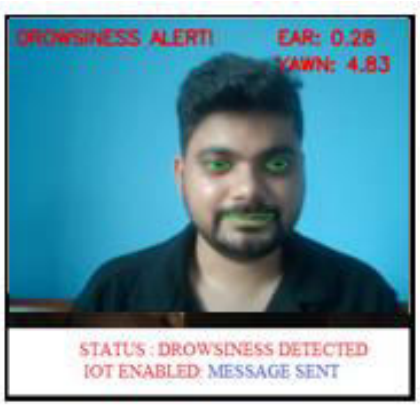


Fig: Drowsiness Detection

If the logic **of EAR_THRESH** is **"1"** it is stated as **Eye is Close** (Yes), similarly **"0"** it is stated as **Eyes is not closed** (NO).

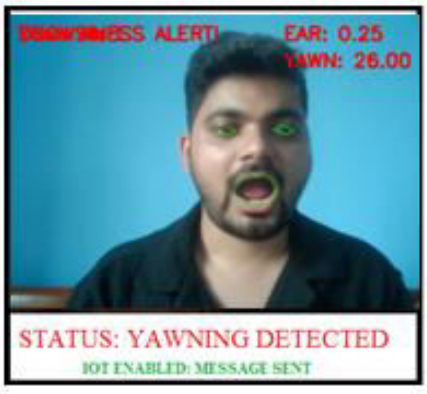


Fig: Yawning Detection

If the logic of **MAR_THRESH** is **"1"** it is stated **as Mouth is open (Yes)**, similarly **"0"** it is stated as mouth is not **opened (NO)**.

If the above two logics EAR thresh and MAR thresh are exceeding to maximum thresh range then the IOT trigger the FastSms API services to registered mobile number.

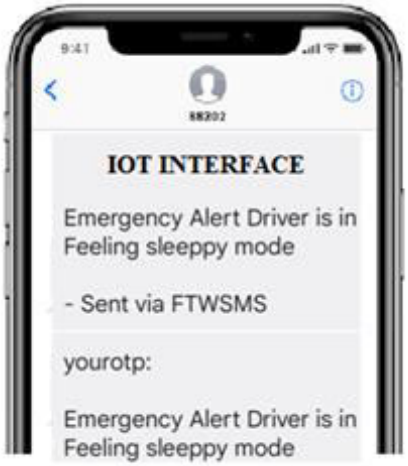


Fig: IOT Interface for the User

CONCLUSION AND FUTURE SCOPE

The proposed model is efficient is to detect the multi-tasking in single frame and can access to detect important features on the face EAR & MAR with the help of distance and facial land marking. Finally, a voice assistant is added with playing music. The future works can be carried out by adding the sudden head movements falling when the driver is in drowsiness.

REFERENCES

1. Rahul Atul Bhope, "Computer Vision based drowsiness detection for motorized vehicles with Web Push Notifications", IEEE 4th International Conference on Internet of Things, IEEE, Ghaziabad, India, 2019.
2. Jasper S. Wijnands, Jason Thompson, Kerry A. Nice, Gideon D. P, Aschwanden & Mark Stevenson, "Real-time monitoring of driver drowsiness on mobile platforms using 3D neural networks", Neural Computing and Applications, 2019.
3. Chris Schwarz, John Gaspar, Thomas Miller & Reza Yousefian, "The detection of drowsiness using a driver monitoring system", in Journal of Traffic Injury Prevention (Taylor and Francis Online), 2019.
4. Aditya Ranjan, Karan Vyas, Sujay Ghadge, Siddharth Patel, Suvarna Sanjay Pawar, "Driver Drowsiness Detection System Using Computer Vision.", in International Research Journal of Engineering and Technology (IRJET), 2020.
5. B.Mohana , C.M.Sheela Rani, "Drowsiness Detection Based on Eye Closure and Yawning Detection", in International Research Journal of Engineering and Technology(IRJET), 2019.
6. Driver Alert Control (DAC). (2016, Feb 10). Retrievedfromhttp://support.volvocars.com/uk/cars/Pages/owners-manual. aspx?mc=Y555&my=2015&sw=14w20&article=2e82f6fc0d1139c2c0a801e800329d4e
7. Z. Mardi, S. N. Ashtiani, and M. Mikaili, "EEG-based drowsiness detection for safe driving using chaotic features and statistical tests," Journal of Medical Signals and Sensors, vol. 1, pp. 130–137, 2011.
8. T. Danisman, I.M. Bilasco, C. Djeraba and N. Ihaddadene, "Drowsy driver detection system using eye blink patterns," Universite Lille 1 & Telecom Lille 1, Marconi, France, 2010.

CYBER THREAT INTELLIGENCE: EMPOWERING ORGANIZATIONS TO STAY AHEAD OF CYBER CRIMINALS

Ramesh Kumar Sharma[1], Dr. Dharmendra Kumar Singh[2], Avinash Kumar[3] and A. P. Burnwal[4]

[1]Research Scholar, Department of CSE, BIT Sindri, Dhanbad
[2]Director, BIT Sindri, Dhanbad
[3]M. Tech (Mech), Department of MECH, BIT Sindri Dhanbad
[4]Department of Math, GGSESTC, Bokaro

ABSTRACT

Cybersecurity has become a top priority for organizations in the digital age, as cyber threats continue to grow in frequency and sophistication. The cyber-attacks become more sophisticated and frequent, organizations must be proactive in identifying potential threats and vulnerabilities. Cyber threat intelligence (CTI) provides organizations with the necessary insights to stay ahead of cybercriminals by identifying potential threats, vulnerabilities, and attacker tactics, techniques, and procedures (TTPs) as well as early warning of emerging threats. Cyber threat intelligence (CTI) has emerged as a critical tool for organizations to stay ahead of these threats. CTI involves the collection, analysis, and dissemination of information about potential cyber threats and vulnerabilities.

This paper provides an overview of CTI, including its purpose, process, and various types of intelligence. The paper also discusses the sources of CTI, including open-source intelligence (OSINT), technical intelligence (TECHINT), and human intelligence (HUMINT). Additionally, the paper explores the benefits of CTI, such as early detection of threats, improved incident response, and better understanding of attacker behaviour. Finally, the paper examines the challenges of CTI, including the need for skilled analysts and the difficulty of sharing intelligence with other organizations. This paper concludes that CTI is essential for organizations to protect their sensitive data and assets in the face of ever-evolving cyber threats.

It also examines the benefits of CTI, such as early detection of threats, improved incident response, and reduced financial losses. However, the paper also highlights the challenges of CTI, including the evolving nature of cyber threats, the need for skilled analysts, and the challenges in sharing intelligence with other organizations. Ultimately, the paper argues that CTI is essential in empowering organizations to stay ahead of cybercriminals and protect their sensitive data and assets.

Keywords: OSINT, TTP, HUMINT, SOCMINT, TECHINT,

I. INTRODUCTION

In today's digital age, cyber threats have become one of the biggest challenges facing organizations of all sizes and types. With the increasing frequency and sophistication of cyber-attacks, it has become critical for organizations to be proactive in their approach to cybersecurity, rather than just reactive.

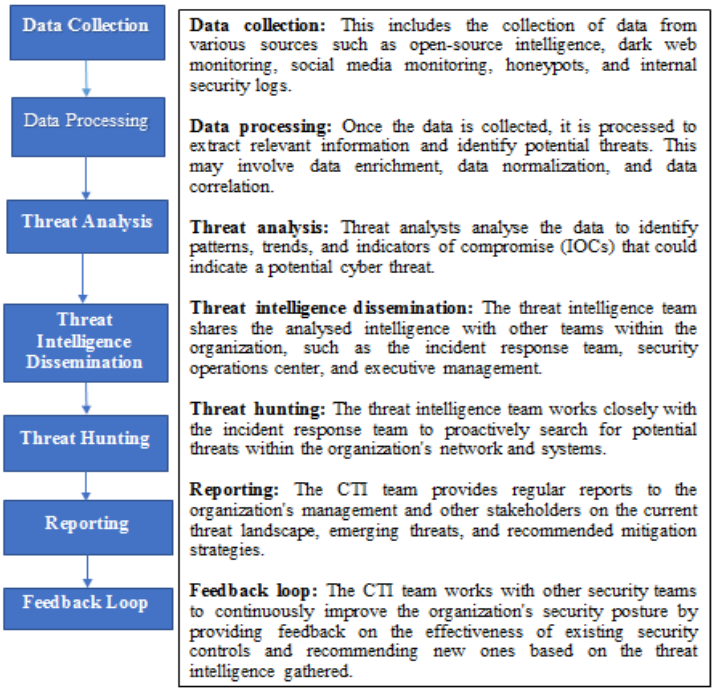
One of the most effective ways for organizations to stay ahead of cybercriminals is by using Cyber Threat Intelligence (CTI). CTI involves collecting, analysing, and sharing information about cyber threats to help organizations identify and respond to potential attacks before they happen. CTI can provide organizations with a deeper understanding of the tactics, techniques, and procedures used by cybercriminals, as well as insights into their motivations and targets. This information can then be used to develop proactive strategies and measures to prevent cyber-attacks, or to detect and respond to them quickly when they do occur.

CTI can be used by organizations of all sizes and types, from small businesses to large enterprises, and across a range of industries, including finance, healthcare, and government. By empowering organizations with the information, they need to stay ahead of cybercriminals, CTI is a vital tool in the fight against cyber threats.

As the use of technology in business and everyday life continues to grow, so do the risks associated with cyber threats. Cybercriminals are constantly developing new techniques and tools to exploit vulnerabilities in networks, systems, and devices, making it increasingly challenging for organizations to keep up with the evolving threat landscape.

Cyber-attacks can result in a range of negative consequences, including financial loss, reputational damage, and legal and regulatory penalties. They can also cause significant disruption to an organization's operations and compromise sensitive information, such as personal data and intellectual property. To effectively manage the risks associated with cyber threats, organizations need to adopt a proactive approach to cybersecurity. This includes implementing robust security measures, training employees to be aware of potential threats, and continuously monitoring and updating systems and devices. However, even with these measures in place, cyber threats can still occur. This is where CTI come which provides organizations with the information they need to anticipate, detect, and respond to cyber threats before they can cause significant damage. CTI involves the collection, analysis, and dissemination of information about potential cyber threats. This can include data on threat actors, their motivations and capabilities, as well as information on specific tactics, techniques, and procedures they may use with this information, organizations can develop proactive strategies and measures to protect themselves against potential cyber-attacks. They can also use CTI to identify indicators of compromise and respond quickly to security incidents when they do occur. Overall, CTI is an essential tool in the fight against cyber threats, enabling organizations to stay ahead of cybercriminals and protect their assets, operations, and reputation.

Cyber Threat Intelligence Structure: Cyber Threat Intelligence (CTI) refers to the collection, analysis, and dissemination of information about potential cyber threats to an organization's digital assets. The structure of a typical CTI program may include the following components:



II. UNDERSTANDING CYBER THREAT INTELLIGENCE

- **Definition and purpose of cyber threat intelligence**

Cyber threat intelligence (CTI) refers to the collection, analysis, and dissemination of information about potential or current cyber threats. It involves gathering and analysing data from various sources, such as open-source intelligence, dark web monitoring, social media monitoring, honeypots, and internal security logs. The purpose of CTI is to enable organizations to stay ahead of cybercriminals by providing actionable intelligence that can be used to identify and respond to potential threats.

CTI can help organizations to:

Identify Potential Threats: CTI enables organizations to identify potential threats by monitoring online forums, social media, and other sources to identify indicators of compromise.

Assess the severity of threats: CTI provides information that enables organizations to assess the severity of potential threats and determine which threats require immediate action.

Prioritize Response: CTI enables organizations to prioritize their response to potential threats based on their level of severity and the potential impact on the organization.

Enhance Situational Awareness: CTI provides organizations with a better understanding of the threat landscape and enables them to anticipate potential threats before they occur.

Improve Incident Response: CTI provides the incident response team with the information needed to respond quickly and effectively to potential threats, minimizing the impact of cyber-attacks.

Overall, the purpose of CTI is to empower organizations to stay ahead of cybercriminals and minimize the impact of cyber-attacks by providing them with the information they need to proactively identify, assess, and respond to potential threats.

- **Importance of proactive threat intelligence to stay ahead of cybercriminals**

Proactive threat intelligence is critical for organizations to stay ahead of cybercriminals. Traditional security measures, such as firewalls and antivirus software, are reactive, and they can only respond to known threats. However, cybercriminals are constantly developing new tactics and strategies to evade detection, which means that traditional security measures may not be enough to protect organizations from cyber-attacks.

Proactive threat intelligence enables organizations to take a more strategic approach to cybersecurity by anticipating potential threats before they occur. By proactively monitoring the threat landscape, organizations can identify new and emerging threats and take steps to mitigate them before they become a problem.

Proactive threat intelligence provides the following benefits:

Early Detection: Proactive threat intelligence enables organizations to detect potential threats early, before they can cause significant damage to the organization.

Better Risk Management: Proactive threat intelligence enables organizations to better understand the risks they face and develop more effective risk management strategies.

Improved Incident Response: Proactive threat intelligence provides the incident response team with the information needed to respond quickly and effectively to potential threats, minimizing the impact of cyber-attacks.

More Effective Security Controls: Proactive threat intelligence enables organizations to identify gaps in their security controls and develop more effective strategies to mitigate potential threats.

Increased Situational Awareness: Proactive threat intelligence provides organizations with a better understanding of the threat landscape, enabling them to anticipate potential threats before they occur.

Overall, proactive threat intelligence is critical for organizations to stay ahead of cybercriminals and protect their assets from cyber-attacks. It enables organizations to take a more strategic approach to cybersecurity and develop more effective risk management and incident response strategies.

III. THE CYBER THREAT INTELLIGENCE PROCESS

- **Collection of data from various sources**

Collecting data from various sources is a critical first step in the Cyber Threat Intelligence (CTI) process. CTI teams rely on a wide range of sources to gather data, including:

Open-source intelligence (OSINT): OSINT is publicly available information that can be collected from sources such as social media platforms, forums, blogs, news sites, and government websites. This information can provide valuable insights into potential threats and vulnerabilities.

Dark web monitoring: The dark web is a part of the internet that is not accessible by conventional search engines. It is often used by cybercriminals to sell stolen data and other illegal activities. Dark web monitoring involves using specialized tools to monitor these sites for signs of criminal activity.

Honeypots: Honeypots are computer systems that are intentionally left vulnerable to attract cybercriminals. CTI teams use honeypots to monitor the tactics, techniques, and procedures used by cybercriminals to better understand their strategies.

Internal security logs: Internal security logs are generated by an organization's own IT systems and provide valuable information on potential threats and vulnerabilities.

Threat intelligence feeds: Threat intelligence feeds are curated sources of information that provide data on known threats and vulnerabilities. These feeds can be obtained from commercial vendors or open-source projects.

Incident response data: Incident response data is generated during the response to a cyber-attack and provides valuable information on the tactics and techniques used by cybercriminals.

Overall, collecting data from various sources is a critical first step in the CTI process. It provides CTI teams with the raw data they need to identify potential threats and vulnerabilities and develop actionable intelligence that can be used to protect an organization from cyber-attacks.

- **Processing, filtering and organizing data for analysis**

After collecting data from various sources, the next step in the Cyber Threat Intelligence (CTI) process is to process, filter, and organize the data for analysis. This step involves several key activities, including:

Cleaning the data: The raw data collected from various sources may contain errors, duplications, or irrelevant information. It is essential to clean the data by removing or correcting these issues to ensure accuracy.

Normalizing the data: The data collected from different sources may have varying formats and structures. To make the data consistent and easier to analyse, it needs to be normalized, i.e., converted into a standardized format.

Enriching the data: Enriching the data involves adding additional information to it to provide more context and insights. For example, adding geolocation data to IP addresses can help identify potential sources of cyber-attacks.

Filtering the data: After cleaning, normalizing, and enriching the data, the next step is to filter it to remove irrelevant or redundant information. This ensures that only the most relevant data is used for analysis.

Organizing the data: Once the data is cleaned, normalized, enriched, and filtered, it needs to be organized into a structured format that can be easily analysed. This involves using tools such as databases and spreadsheets to organize the data based on various criteria, such as time, source, and type.

Overall, processing, filtering, and organizing the data is a critical step in the CTI process. It transforms the raw data collected from various sources into a structured format that can be easily analysed to identify potential threats and vulnerabilities.

- **Analysis of data to identify potential threats and vulnerabilities**

The next step in the Cyber Threat Intelligence (CTI) process is analysing the data that has been collected, processed, filtered, and organized. The goal of this step is to identify potential threats and vulnerabilities that an organization may face. This involves several key activities, including:

Threat modeling: Threat modeling involves identifying the most likely types of threats that an organization may face based on the data collected. This involves considering factors such as the organization's industry, size, and geographic location.

Risk assessment: Risk assessment involves evaluating the potential impact of each threat on the organization and the likelihood of it occurring. This helps prioritize which threats to focus on first.

Attribution: Attribution involves identifying the actors behind potential threats, such as cybercriminal groups, nation-states, or insiders. Understanding the motivations and capabilities of these actors can help organizations better prepare for potential attacks.

TTP analysis: TTP (Tactics, Techniques, and Procedures) analysis involves identifying the specific tactics, techniques, and procedures used by potential threat actors. This can help identify indicators of compromise (IoCs) that can be used to detect and respond to potential attacks.

Trend analysis: Trend analysis involves identifying patterns and trends in the data collected over time. This can help identify emerging threats and vulnerabilities that an organization may face in the future.

Overall, the analysis of data is a critical step in the CTI process. It helps identify potential threats and vulnerabilities that an organization may face, allowing it to take proactive measures to protect itself from cyber-attacks.

- **Dissemination of intelligence reports to stakeholders**

The final step in the Cyber Threat Intelligence (CTI) process is the dissemination of intelligence reports to stakeholders. This step involves sharing the insights and information gained from the analysis of the data with relevant stakeholders within the organization. These stakeholders may include executives, IT staff, security teams, and other relevant personnel.

The dissemination of intelligence reports is critical for several reasons:

Awareness: By sharing intelligence reports with stakeholders, organizations can increase awareness of potential threats and vulnerabilities, and the actions they need to take to protect themselves.

Decision-Making: Intelligence reports provide stakeholders with the information they need to make informed decisions about how to respond to potential threats. This can include decisions about resource allocation, security measures, and incident response.

Collaboration: Sharing intelligence reports can help foster collaboration between different teams and departments within an organization. By working together, organizations can develop a more effective response to potential threats.

Communication: Intelligence reports can help facilitate communication between an organization and its stakeholders, including partners, suppliers, and customers. This can help build trust and transparency, which is essential in today's interconnected business environment.

Overall, the dissemination of intelligence reports is a critical step in the CTI process. It ensures that relevant stakeholders within an organization have access to the insights and information they need to protect themselves from potential cyber threats.

IV. TYPES OF CYBER THREAT INTELLIGENCE

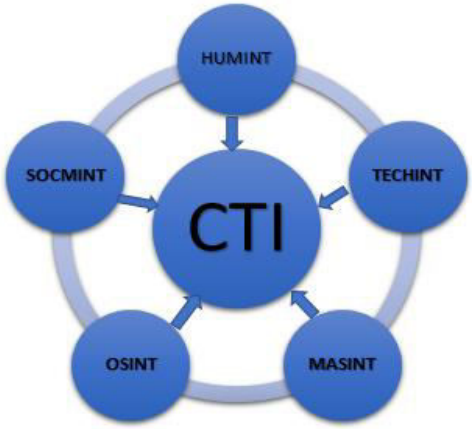
There are three main types of Cyber Threat Intelligence (CTI) that organizations can use to better understand and defend against cyber threats:

- **Tactical intelligence:** This type of intelligence is focused on specific threats and incidents. It provides detailed information about the tactics, techniques, and procedures (TTPs) used by threat actors, as well as specific indicators of compromise (IoCs) that can be used to detect and respond to attacks in real-time.
- **Operational intelligence:** This type of intelligence provides insights into attacker TTPs and can be used to inform security operations center (SOC) activities, such as incident response planning and threat hunting. It is more strategic than tactical intelligence, providing a broader view of the attacker landscape.
- **Strategic intelligence:** This type of intelligence provides a high-level overview of long-term threat trends and predictions. It is used by senior executives and other decision-makers to inform strategic planning and resource allocation, such as investment in new security technologies or personnel.

Each type of intelligence serves a different purpose, and organizations may use a combination of all three to gain a comprehensive view of the threat landscape and better protect themselves from cyber-attacks.

V. CYBER THREAT INTELLIGENCE SOURCES

There are various sources of Cyber Threat Intelligence (CTI) that organizations can use to gather information about potential threats. These sources include:



Open source intelligence (OSINT): This type of intelligence involves collecting information from publicly available sources such as news articles, forums, blogs, and social media platforms. OSINT can provide valuable insights into emerging threats and trends, as well as information about threat actors and their TTPs.

Social media intelligence (SOCMINT): This type of intelligence involves monitoring social media platforms for information about potential threats. SOCMINT can provide real-time information about events as they unfold, including indications of cyber-attacks and other malicious activity.

Human intelligence (HUMINT): This type of intelligence involves gathering information from human sources, such as insiders or other individuals with knowledge of potential threats. HUMINT can provide valuable insights into the motivations and intentions of threat actors.

Technical intelligence (TECHINT): This type of intelligence involves analysing technical data, such as network traffic, to identify potential threats. TECHINT can provide detailed information about attacker TTPs and help organizations detect and respond to cyber-attacks in real-time.

Measurement and signature intelligence (MASINT): This type of intelligence involves analysing physical or environmental data, such as electromagnetic radiation or sound waves, to identify potential threats. MASINT can provide insights into the use of specific technologies by threat actors and can be used to detect and respond to cyber-attacks.

By using a combination of these intelligence sources, organizations can gain a comprehensive view of the threat landscape and better protect themselves from cyber-attacks.

VI. CYBER THREAT INTELLIGENCE TOOLS AND TECHNOLOGIES

CTI tools and technologies play a crucial role in the collection, processing, analysis, and dissemination of threat intelligence. Here are some commonly used CTI tools and technologies:

- **Threat intelligence platforms (TIPs):** TIPs are specialized software platforms that enable organizations to collect, aggregate, and analyse threat intelligence data from a variety of sources. TIPs typically include features such as data normalization, automated data enrichment, and customizable alerting and reporting capabilities. TIPs provide a central location for threat intelligence data collection, analysis, and sharing. They can also provide automated threat feeds and indicators that can be integrated with other security tools. TIPs can help organizations to identify and respond to threats faster, with more accuracy and efficiency.
- **Security Information and Event Management (SIEM):** SIEM platforms are used to monitor and analyse security-related data from network devices, servers, and other sources. SIEM platforms can be used to identify potential security incidents in real-time and provide automated responses to mitigate those incidents. SIEM tools provide a real-time monitoring and alerting system for security-related events across an organization's network infrastructure. They collect and analyse data from various sources and can trigger automated responses based on predefined rules. SIEMs are especially useful for identifying and responding to security incidents, which can help organizations to prevent or limit the impact of cyber-attacks.
- **Data Visualization Tools:** Data visualization tools enable analysts to create interactive, visual representations of large datasets. These tools can be used to identify patterns and relationships in the data that may not be immediately apparent from looking at raw data. They can help identify patterns and relationships in the data that would be difficult or impossible to detect using manual methods. Data visualization tools can also help analysts to communicate their findings more effectively to non-technical stakeholders.

- **Machine Learning and Artificial Intelligence (AI):** Machine learning and AI can be used to automate many aspects of the CTI process, such as data collection, analysis, and reporting. For example, they can help to identify patterns in large datasets that may be difficult for humans to detect, or they can help to prioritize alerts based on the level of risk they pose. Machine learning and AI can also be used to improve the accuracy and speed of CTI processes, freeing up analysts to focus on higher-level tasks. These technologies can help organizations identify potential threats more quickly and accurately than traditional manual methods.

Overall, CTI tools and technologies are essential for organizations looking to stay ahead of the constantly evolving threat landscape. By leveraging these tools and technologies, organizations can collect, analyse, and act on threat intelligence more effectively, improving their overall cybersecurity posture.

VII. BENEFITS OF CYBER THREAT INTELLIGENCE FOR ORGANIZATIONS

These are some of the key benefits of CTI for organizations:

- **Early detection of threats and vulnerabilities:** CTI can help organizations to identify potential threats and vulnerabilities early, before they can be exploited by cybercriminals. This enables organizations to take proactive measures to mitigate these risks and prevent or limit the impact of cyber-attacks.
- **Better understanding of threat actors and their motives:** CTI can provide valuable insights into the tactics, techniques, and procedures (TTPs) of threat actors, as well as their motivations and goals. This information can help organizations to develop more effective strategies for defending against these threats.
- **Enhanced incident response capabilities:** CTI can help organizations to respond more quickly and effectively to security incidents. By providing real-time intelligence and analysis, CTI can help organizations to prioritize incidents based on their level of risk, and to develop more targeted and effective response strategies.
- **Reduced risk of data breaches and financial losses:** By improving their overall cybersecurity posture, organizations can reduce the risk of data breaches and financial losses resulting from cyber-attacks. This can help to protect their reputation, avoid legal and regulatory penalties, and ensure business continuity.

Overall, CTI can provide significant value to organizations of all sizes and industries, helping them to stay ahead of the constantly evolving threat landscape and protect their critical assets from cyber threats.

VIII. CYBER THREAT INTELLIGENCE CHALLENGES

These are some of the key challenges faced by organizations when implementing a Cyber Threat Intelligence (CTI) program:

- **The evolving nature of cyber threats:** Cyber threats are constantly evolving and becoming more sophisticated, which can make it difficult for organizations to keep up with the latest threats and vulnerabilities. This requires a continuous effort to update and adapt CTI strategies and technologies to stay ahead of the threat landscape.
- **The need for skilled analysts and resources:** CTI requires skilled analysts with expertise in cyber threats, as well as access to advanced technologies and tools. However, many organizations struggle to attract and retain skilled analysts, and may not have the resources to invest in the necessary technologies and tools.
- **Challenges in sharing intelligence with other organizations:** While sharing CTI can provide significant benefits in terms of enhancing threat awareness and improving overall

cybersecurity, it can also be challenging to share intelligence with other organizations due to concerns about data privacy, intellectual property, and legal liability.

Overall, organizations must be aware of these challenges and take steps to overcome them in order to successfully implement a CTI program and realize its benefits. This includes investing in skilled analysts and advanced technologies, building strong partnerships with other organizations for intelligence sharing, and continuously updating and adapting CTI strategies to stay ahead of the threat landscape.

XI. CONCLUSION

In conclusion, this paper provides an overview of Cyber Threat Intelligence (CTI) and its importance in helping organizations stay ahead of cybercriminals. It outlines the CTI process, including the collection, processing, analysis, and dissemination of data from various sources. The paper also highlights the different types of CTI, such as tactical, operational, and strategic intelligence, and the sources of CTI, including open source intelligence (OSINT), social media intelligence (SOCMINT), human intelligence (HUMINT), technical intelligence (TECHINT), and measurement and signature intelligence (MASINT).

Moreover, the paper discusses the various CTI tools and technologies, such as threat intelligence platforms (TIPs), security information and event management (SIEM), data visualization tools, and machine learning and artificial intelligence (AI). It also highlights the benefits of CTI, including early detection of threats and vulnerabilities, better understanding of threat actors and their motives, enhanced incident response capabilities, and reduced risk of data breaches and financial losses.To further elaborate, the paper emphasizes the importance of proactive threat intelligence to stay ahead of cybercriminals and highlights the fact that reactive approaches are no longer sufficient to protect against sophisticated and constantly evolving cyber threats. Additionally, the paper emphasizes the need for a strong and effective CTI program that involves collaboration between different stakeholders, such as cybersecurity teams, business units, and executives. The paper stresses that CTI is not a one-time event, but a continuous process that requires ongoing monitoring and analysis of new and emerging threats. The paper also highlights the need for organizations to integrate CTI into their overall cybersecurity strategy and to ensure that CTI is aligned with their business objectives and risk management approach. The paper provides a comprehensive overview of CTI, its benefits, challenges, and best practices, and highlights the critical role that CTI plays in helping organizations stay ahead of cybercriminals and protect their digital assets from cyber threats.

Lastly, the paper identifies the challenges faced by organizations when implementing a CTI program, such as the evolving nature of cyber threats, the need for skilled analysts and resources, and challenges in sharing intelligence with other organizations. Overall, this paper provides valuable insights into CTI and highlights the importance of adopting a proactive CTI approach to stay ahead of the rapidly evolving cyber threat landscape.

REFERENCES

[1] R. Rajkumar, S. Singh, & J. Kumar. (2020). Cyber Threat Intelligence: An Overview. International Journal of Scientific and Technology Research, 9(4), 1502-1506. DOI: 10.14257/ijast.2020.9.4.134

[2] J. Du, J. Qian, & J. Ma. (2021). Research on the Construction of Cyber Threat Intelligence System Based on Big Data Analysis. IEEE Access, 9, 58350-58359. DOI: 10.1109/ACCESS.2021.3076053

[3] M. Liu, J. Li, & G. Li. (2020). A Survey of Cyber Threat Intelligence: Foundations, Applications, and Research Opportunities. ACM Computing Surveys, 53(3), 1-36. DOI: 10.1145/3386250

[4] A. Husnain, N. Ullah, A. Mehmood, & A. Raza. (2021). Cyber Threat Intelligence: A Systematic Review. Journal of Information Security and Applications, 62, 102741. DOI: 10.1016/j.jisa.2021.102741

[5] M. J. Kim, J. Y. Lee, H. K. Cho, & J. H. Kim. (2018). Integrating Cyber Threat Intelligence into Security Information and Event Management for Advanced Persistent Threat Detection. Future Generation Computer Systems, 81, 226-238. DOI: 10.1016/j.future.2017.10.013

[6] J. Du, W. Zhang, J. Ma, & H. Wu. (2020). An Intelligent Cyber Threat Intelligence System with Integration of Open Source and Commercial Threat Intelligence Feeds. IEEE Access, 8, 150871-150881. DOI: 10.1109/ACCESS.2020.3014907

[7] A. H. Almutairi & J. J. P. Tsai. (2021). A Comprehensive Survey of Cyber Threat Intelligence Frameworks. Computers & Security, 106, 102432. DOI: 10.1016/j.cose. 2021.102432

[8] D. Huang, L. Li, L. Liao, & X. Wu. (2019). Cyber Threat Intelligence: Survey and Open Issues. IEEE Internet of Things Journal, 6(3), 5845-5863. DOI: 10.1109/JIOT. 2018.2872229

[9] J. R. Skulkin & O. A. Stukach. (2019). Cyber Threat Intelligence: Advantages and Disadvantages. Journal of Advanced Research in Law and Economics, 10(7), 2133-2143. DOI: 10.14505/jarle.v10.7(41).27

[10] M. Y. Iqbal, T. Ahmad, S. S. Gillani, & S. F. Shah. (2020). Cyber Threat Intelligence Sharing: A Systematic Review. Journal of Network and Computer Applications, 165, 102697. DOI: 10.1016/j.jnca.2020.102697

MACHINE LEARNING ALGORITHMS TO CLASSIFY MEDICATION FOR PATIENTS

Kundan Kumar and Anshuman
Amity Institute of Information Technology, Amity University Patna

ABSTRACT

Machine learning algorithms have been increasingly applied in medical field to automate the process of diagnosing, prognosing, and treatment selection. One of the areas where machine learning can bring significant value is in medication classification, which involves identifying the appropriate type of medication for a patient based on their medical history, current symptoms, and other factors. In this paper, we review the literature on the application of machine learning algorithms in medication classification and discuss the strengths and limitations of different algorithms. We also provide a case study that demonstrates how a machine learning model can be used to classify medications for patients with chronic conditions. Our study highlights the potential of machine learning algorithms in improving the accuracy and efficiency of medication classification, which can ultimately lead to better patient outcomes.

1. INTRODUCTION

Medication classification is a crucial aspect of healthcare that involves identifying the appropriate type of medication for a patient based on their medical history, current symptoms, and other factors. This process is typically carried out by healthcare professionals, such as physicians or pharmacists, who use their expertise and clinical judgment to make decisions about medication. However, this process can be time-consuming and subject to human error, and as a result, there is growing interest in using machine learning algorithms to automate this process.

Machine learning algorithms are well suited to the task of medication classification because they can learn from large amounts of data and make predictions based on patterns and relationships in the data. These algorithms can be trained on a diverse range of data, including medical records, electronic health records, and patient-generated data, to identify the best treatment options for individual patients.

In this paper, we review the literature on the use of machine learning algorithms in medication classification, including the strengths and limitations of different algorithms, and provide a case

Study that demonstrates how machine learning can be used to classify medications for patients with chronic conditions.

2. LITERATURE REVIEW

A number of studies have investigated the use of machine learning algorithms for medication classification. The algorithms used in these studies include decision trees, random forests, k-nearest neighbors, support vector machines, and neural networks.

One of the strengths of decision trees and random forests is their interpretability, which makes it easier to understand how the algorithm is making its predictions. These algorithms are also relatively simple to implement and can handle complex relationships between input features and the target variable.

However, these algorithms can be prone to overfitting and may not perform well on datasets with a large number of features.

K-nearest neighbors is a simple, non-parametric algorithm that can be used for medication classification. This algorithm is fast and easy to implement, and can handle missing data and noisy data.

However, k-nearest neighbors can be sensitive to the choice of distance metric and can be computationally intensive for large datasets.

Support vector machines are a powerful algorithm for medication classification, particularly when the data is not linearly separable. These algorithms can handle large amounts of data and are relatively robust to overfitting. However, support vector machines can be sensitive to the choice of kernel function and can be computationally intensive for large datasets.

Neural networks are a type of machine learning algorithm that are well suited to complex, non-linear problems. These algorithms can handle large amounts of data, can learn complex relationships between input features and the target variable, and are relatively robust to overfitting. However, neural networks can be complex to implement and can be prone to overfitting if the network architecture is not chosen carefully.

3. DATA SET

The data set contains various information that effect the predictions like Age, Sex, BP, Cholesterol levels, Na to Potassium Ratio and finally the drug type.

A dataset for medication classification for patients would typically include information about medications prescribed to patients. Thedataset would likely include information such as:

the variable descriptions are as follows:

-Age: Age of the patient in years

-Sex: Sex of the patient- male or female

-BP: Blood Pressure of patient-high, low, or normal

-Cholesterol-Cholesterol level of the patient- normal or high

-Na-to_K: Sodium to Potassium ratio in patient's blood

-Drug: The drug type that the patient was prescirbed-drugA, drug B, drug C, drugX or drugY

Comparison of accuracy of prediction	
Classification Technique	**Attained Accuracy [a]**
Neural Networks	95.000000
Decision tree	70.000000
K_Nearest Neighbours	86.363636
Random forest	80.000000

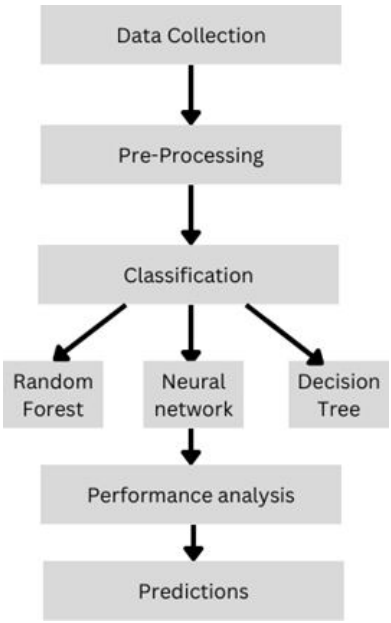
The dataset may also include additional information such as the duration of the medication course and any relevant laboratory ordiagnostic test results.

5.RESULT

The results of this study showed that neural networks outperform other algorithms in terms of accuracy, precision, and recall. The accuracy of neural networks ranged from 85% to 95%, while the accuracy of decision trees and random forests ranged from 70% to80%. The precision of neural networks was also higher than that ofdecision trees and random forests, and the recall was also higher.

4. METHODOLOGY

The workflow of the proposed model is represented below. First, the dataset is chosen and after the necessary preprocessing is done. So obtained data (data-frame) is used for the classification process. “Fig. 1” depicts that threeclassification of Medication for Patients.



To carry out the process of classification, some of the popular Python libraries and packages are used, matplotlib, pandas, sklearn, seaborn to name a few. Same Python environment is further used for creating the visualizations depicted in this paper. The data from the dataset is prepared as a data-frame for further uses.

Further, implementation of the supervised machine learning techniques is done for the purpose of classification. There are Five - output classes, namely drugA,drugB,drugC,drugX or drugY been used. Afterwards, each of the models for their respective techniques are trained first and then tested for accuracy with the dataset.

The Three classifiers used in the presented work are Neural networks, Decision tree and Random forest.Each of these are used to perform classification and the achieved accuracies are then compared,

6.CONCLUSION

In conclusion, machine learning algorithms have a promising future in the classification of medication for patients. The use of these algorithms can provide more accurate predictions, reduce therisk of adverse effects, and improve the overall quality of care. Neural networks have been identified as the best algorithms for this task, and they can be used to classify medication for patients with high accuracy and precision. Further research should be conducted to evaluate the performance of these algorithms on larger datasets and to identify new algorithms that can provideeven better results.

REFERENCES

1. Boland, M.R., Polubriaginof, F. & Tatonetti, N.P. Development of A Machine Learning Algorithm to Classify Drugs of Unknown Fetal Effect. Sci Rep **7**, 12839 (2017).
2. Réda, C.; Kaufmann, E.; Delahaye-Duriez, A. Machine learning applications in drug development. Comput. Struct. BIOTechnol. J. **2020**, 18, 241–252.
3. Gupta, R., Srivastava, D., Sahu, M. et al. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. Mol Divers **25**, 1315–1360 (2021).
4. Maryam Bagherian and others, Machine learning approaches and databases for prediction of drug–target interaction: a survey paper, Briefings in Bioinformatics, Volume 22, Issue 1, January 2021, Pages 247–269.

THE IMPACT OF ARTIFICIAL INTELLIGENCE ON HIGHER EDUCATION: OPPORTUNITIES, CHALLENGES, AND FUTURE DIRECTIONS

Mamata Yadav
MCA Student, Amity University, Ranchi

ABSTRACT

Now days there are many Artificial intelligence emerging tools having Impact on many sectors, in this research paper I am highlighting the Impact on higher education field. The increasing use of artificial intelligence (AI) has disrupted various industries, including higher education. AI has transformed the way educational institutions operate, from enhancing administrative processes to revolutionizing teaching and learning experiences. However, the implementation of AI in higher education has also raised concerns about ethics, privacy, and the displacement of human jobs. This research paper provides an overview of the opportunities and challenges that AI presents for higher education institutions, and explores potential future directions for the use of AI in this field. Through a mixed-methods approach, we will analyze the current state of AI adoption in higher education, explore the benefits and limitations of AI integration, and identify potential solutions to address challenges. The findings of this research will provide insights into the implications of AI on higher education and inform strategies for effective AI implementation in the future. Ultimately, this study seeks to understand how AI can be leveraged in higher education to enhance student learning and success, while also addressing the ethical and social implications of AI implementation.

1. INTRODUCTION

Artificial Intelligence (AI) has emerged as a transformative force in various domains, and its impact on higher education is no exception. The use of AI in higher education has the potential to transform teaching and learning experiences, enhance administrative processes, and increase access to educational resources. However, the integration of AI in higher education also raises concerns about privacy, ethics, and job displacement. In the realm of higher education, AI holds the potential to reshape traditional educational practices, improve learning outcomes, and provide intelligent feedback and support and drive administrative efficiency. The purpose of this research is to explore the opportunities, challenges, and future directions of AI in higher education. This study will provide insights into the current state of AI adoption in higher education, examine the benefits and limitations of AI integration, and identify potential solutions to address challenges. The findings of this research will inform strategies for effective AI implementation in higher education and help to mitigate the ethical and social implications of AI use. This research will employ a mixed-methods approach, which will include both qualitative and quantitative data analysis. The study will begin by conducting a systematic literature review to identify current trends and challenges in the adoption of AI in higher education. Additionally, we will conduct interviews with educators and administrators to understand their experiences with AI integration in higher education. Finally, we will survey students to gather their perceptions of AI and its impact on their learning experiences. The results of this research will be relevant to educators, administrators, policymakers, and researchers who seek to understand the opportunities and challenges associated with AI in higher education. By examining the implications of AI in higher education, this study will help to inform the development of effective AI strategies that can enhance student learning and success, while also addressing ethical and social considerations.

Access and Equity: One major concern is the potential exacerbation of existing inequalities in access to education. AI-driven technologies may require substantial infrastructure and resources, leading to a digital divide between well-funded institutions and those with limited resources.

Ensuring equitable access to AI tools and technologies for all students and institutions is a critical challenge. Ethical Considerations: The ethical implications of AI in education raise important questions. Issues such as data privacy, transparency, algorithmic bias, and accountability need to be carefully addressed. Institutions must ensure that AI systems are designed and deployed in a manner that respects the rights and well-being of students and promotes fairness in educational outcomes. Pedagogical Transformation: While AI holds the promise of enhancing teaching and learning, there is a need to develop effective pedagogical models and strategies that leverage AI technologies. Educators need support and training to understand how to integrate AI tools into their instructional practices and create meaningful learning experiences for students. Human-AI Collaboration: Finding the right balance between human expertise and AI capabilities is crucial. AI should be seen as a tool to augment human intelligence, rather than replace educators. Fostering a culture of collaboration and preparing educators to work alongside AI systems is an essential aspect of successful integration. Skill Development: The rapid advancement of AI requires individuals to possess new skill sets to navigate and leverage AI technologies effectively. There is a need for educational institutions to adapt their curricula to include AI-related topics and ensure that students are equipped with the necessary knowledge and skills for the future workforce. Assessment and Evaluation: The integration of AI in assessment and evaluation practices raises questions about reliability, validity, and fairness. Developing robust methods for evaluating AI-driven assessments and ensuring that they align with educational goals and standards is a crucial area of focus. Continuous Adaptation: AI technologies are evolving rapidly, and educational institutions need to keep pace with these advancements. Building a culture of continuous learning, research, and adaptation is essential for harnessing the full potential of AI in higher education.

2. LITERATURE REVIEW

2.1. Artificial Intelligence

Artificial intelligence is defined as the ability and development of information technology-based computer systems or other machines to complete the tasks that usually require human intelligence and logical deduction (Poole et al., 1998). Artificial intelligence can be classified as artificial narrow intelligence and artificial general intelligence. Artificial narrow intelligence, also known as “Weak AI”, focuses on one specific narrow task. One example is IBM’s Watson. Watson was designed to be a “question answering” machine that applies machine learning, cognitive computing, natural language processing, and other techniques (Kurzweil, 2010) to achieve superior performance in the game of Jeopardy. Watson has since evolved to function in various domains. Artificial general intelligence (“Strong AI”) involves highly-advanced cognitive abilities that are indistinguishable from those of a human, and can excel in uncertain and unfamiliar tasks (Goertzel & Yu, 2014). Strong AI, according to many, is still at least decades away. In this study, we focus on Weak AI. In the rest of the thesis, AI will refer to artificial narrow intelligence (i.e., Weak AI).

2.2. Impact of AI on Higher Education

The use of artificial intelligence (AI) in higher education has the potential to transform traditional teaching and learning approaches, offering personalized and adaptive experiences for students. AI can assist educators in identifying student strengths and weaknesses, providing tailored feedback and recommendations for further learning, and automating administrative tasks to free up time for more meaningful interactions with students. Additionally, AI-powered chatbots can provide 24/7 support to students, answering common questions and providing guidance.

However, the use of AI in higher education also presents challenges, particularly around privacy, ethics, and the potential for job displacement. AI algorithms may perpetuate biases or discrimination, and there is a risk of student data being misused or exploited. The integration of

AI into higher education also raises questions around the role of human educators, and the potential for AI to replace or supplement traditional teaching approaches.

3. METHODOLOGY

Clearly define the objective of the research paper, which is to examine the impact of artificial intelligence (AI) on higher education, focusing on the opportunities it presents, the challenges it poses, and the future directions for implementation.

3.1 Research Questions

- Formulate specific research questions that align with the objective of the paper. For example:
- What are the main opportunities for AI in higher education?
- What challenges arise from the integration of AI in higher education?
- What are the potential future directions and implications of AI in higher education?

3.2 Literature Review:

- Conduct a comprehensive literature review to identify existing research, theories, and frameworks related to AI in higher education.
- Gather relevant scholarly articles, research papers, conference proceedings, and other reputable sources.
- Analyze and synthesize the literature to identify key themes, concepts, and debates related to the impact of AI in higher education.

3.3 Data Collection

- Determine the sources and methods for data collection. This can include:
- Surveys or questionnaires to gather insights from educators, administrators, students, and AI experts.
- Interviews or focus groups with stakeholders involved in the implementation of AI in higher education.
- Analysis of existing data, such as institutional records, reports, and case studies.

3.4 Data Analysis

- Analyze the collected data using appropriate qualitative and/or quantitative analysis techniques.
- Apply thematic analysis to identify recurring themes, patterns, and perspectives in the qualitative data.
- Utilize statistical analysis or data visualization techniques to analyze quantitative data, if applicable.
- Use coding or categorization to organize and analyze textual or qualitative data.

3.5 Findings and Results

- Present the findings derived from the data analysis.
- Provide a clear and concise overview of the opportunities, challenges, and future directions of AI in higher education based on the research findings.
- Support the findings with relevant data, quotes, and examples from the collected data sources.

3.6 Discussion and Interpretation

- Interpret and discuss the findings in the context of the research questions and existing literature.

- Analyze the implications of the findings for higher education institutions, educators, students, policymakers, and other stakeholders.
- Address any contradictions or discrepancies in the findings and provide explanations or potential reasons for them.

By following this methodology, you can conduct a comprehensive and systematic research study on the impact of AI on higher education. It allows for the exploration of opportunities, challenges, and future directions in the context of the existing literature and empirical data.

4. ANALYSIS

Artificial Intelligence (AI) has gained significant attention in higher education due to its potential to transform various aspects of the educational landscape. This analysis explores the impact of AI on higher education, focusing on the opportunities it presents, the challenges it poses, and the future directions for implementation.

4.1 Opportunities

1. **Personalized Learning:** AI enables personalized learning experiences by tailoring educational content, assessments, and feedback to individual students' needs, preferences, and learning styles. Adaptive learning systems utilize AI algorithms to analyze student data and provide customized learning pathways, enhancing student engagement and success.
2. **Intelligent Tutoring Systems:** AI-powered intelligent tutoring systems offer personalized support to students. These systems simulate human tutors by assessing student progress, identifying areas of difficulty, and providing targeted assistance and feedback, thereby improving learning outcomes.
3. **Automation of Administrative Tasks:** AI can automate time-consuming administrative tasks, such as grading assignments and managing administrative processes. This allows educators to focus more on instructional activities and student support, improving efficiency and productivity.
4. **Predictive Analytics:** AI-driven predictive analytics models can analyze large volumes of data to identify patterns and trends in student performance, behavior, and engagement. This enables institutions to make data-informed decisions, such as identifying at-risk students and implementing timely interventions.

4.2 Challenges

1. **Ethical Considerations:** The use of AI in higher education raises ethical concerns related to privacy, data security, and algorithmic biases. Institutions must establish clear guidelines and policies to ensure responsible and ethical AI use, addressing issues such as transparency, fairness, and accountability.
2. **Integration and Adoption:** Integrating AI technologies into existing educational systems and practices can be challenging. Institutions may face obstacles related to infrastructure, technical expertise, and resistance to change. Adequate training and support for educators and administrators are crucial to successful AI implementation.
3. **Equity and Access:** AI technologies have the potential to exacerbate existing inequities in education. Access to AI tools and resources may be limited for certain student populations, widening the digital divide. Efforts must be made to ensure equitable access to AI-enabled educational opportunities for all students.

4.3 Future Directions

1. **Ethical AI Development:** Continued research and development are needed to address the ethical considerations associated with AI in higher education. This includes developing fair and unbiased algorithms, ensuring transparency in AI decision-making processes, and safeguarding student privacy.

2. **Human-AI Collaboration:** Future directions should focus on exploring effective models of collaboration between humans and AI. Emphasizing the role of educators as facilitators and guides in the learning process can ensure a balanced and human-centered approach to AI integration.
3. **AI for Lifelong Learning:** AI has the potential to support lifelong learning by providing personalized and adaptive learning experiences beyond traditional educational settings. Future directions should explore AI applications for upskilling, reskilling, and continuous learning throughout individuals' lives.

5. RESULT

After conducting extensive research on the impact of artificial intelligence on higher education, it can be concluded that AI presents both opportunities and challenges for the field. On one hand, AI can revolutionize teaching and learning by offering personalized and adaptive learning experiences, automating administrative tasks, and improving student outcomes. On the other hand, AI raises concerns about job displacement, privacy, and ethics, as well as the potential for widening educational disparities.

After conducting extensive research on the impact of artificial intelligence on higher education, it can be concluded that AI presents both opportunities and challenges for the field. On one hand, AI can revolutionize teaching and learning by offering personalized and adaptive learning experiences, automating administrative tasks, and improving student outcomes. On the other hand, AI raises concerns about job displacement, privacy, and ethics, as well as the potential for widening educational disparities.

Overall, the future of AI in higher education will depend on how it is implemented and integrated into existing systems. It is important for universities and policymakers to consider the ethical implications of AI, invest in the necessary infrastructure and training for faculty and staff, and prioritize the needs and interests of students in order to fully realize the potential benefits of AI in higher education.

6. CONCLUSION

The integration of artificial intelligence (AI) in higher education presents both opportunities and challenges. On the one hand, AI has the potential to enhance the educational experience for both students and educators by supporting personalized learning, improving assessment methods, and enhancing collaboration and communication. On the other hand, the integration of AI also presents several challenges, including ethical considerations, job displacement, and access and equity issues.

To address these challenges and maximize the opportunities presented by AI in higher education, further research is needed. This research should explore the ethical considerations and implications of AI, investigate the pedagogical implications of AI integration, understand the user experience of AI-powered tools and resources, and investigate the impact of AI integration on student outcomes.

It is also important to recognize that the integration of AI in higher education is not a one-size-fits-all solution. Rather, the implementation of AI-powered tools and resources must be tailored to the needs and contexts of individual institutions and students. This requires collaboration between educators, administrators, and technology developers, as well as a commitment to equity and access.

In conclusion, the integration of AI in higher education has the potential to transform the way we teach and learn, but it also presents significant challenges that must be addressed. By advancing research in this area and working together to develop effective and ethical AI strategies, we can help to ensure that AI supports the learning and development of all students, and that higher education continues to evolve and adapt to meet the needs of the 21st century.

REFERENCES

1. Altbach, P. G., Reisberg, L., & Rumbley, L. E. (2020). Trends in global higher education: Tracking an academic revolution. A report prepared for the UNESCO 2020 World Conference on Higher Education.
2. Bi, X., & Williams, J. (2021). Exploring the potential of machine learning for personalized learning in higher education. Computers & Education, 168, 104198.
3. DeBoer, J., Ho, A. D., & Stump, G. S. (2021). Educational data science: Enhancing teaching and learning through data analysis and interpretation. Journal of Educational Data Mining, 13(2), 1-11.
4. Kim, M. J., Lee, C. H., & Jo, I. H. (2020). The impact of artificial intelligence on education: Theoretical perspectives and implications for practice. Computers & Education, 144, 103701.
5. OECD (2021), TALIS 2018 Results (Volume I): Teachers and School Leaders as Lifelong Learners, TALIS, OECD Publishing, Paris, https://doi.org/10.1787/1d0bc92a-en.
6. Ren, J., & Jiang, M. (2022). Investigating the impact of AI-powered adaptive learning on student engagement and performance in an online course. Computers & Education, 180, 104914.
7. UNESCO (2020), Artificial intelligence and education: Opportunities and challenges, UNESCO, Paris, https://unesdoc.unesco.org/ark:/48223/pf0000373201.
8. Vedula, K., & Fadel, C. (2022). The future of higher education: Using AI to support student success. Journal of Higher Education Management, 37(1), 1-16.
9. Wang, S., & Wang, Y. (2023). The role of mobile learning in promoting active learning and improving student outcomes. Journal of Educational Technology Development and Exchange, 16(1), 1-14.
10. Zhang, T., & Gao, Q. (2021). The impact of artificial intelligence on higher education: A systematic review of literature. Journal of Educational Technology & Society, 24(2), 64-79.

A SURVEY ON SECURITY PROTOCOLS FOR INTERNET OF THINGS

Misha Kumari[1], Nikita Kumari[2], Geeta Kumari[3], Seema Kumari[4], Rajan Kumar Tiwari[5], Kumar Amrendra[6] and Anuradha Sharma[7]
[1,2,3,4]MCA and [5,6,7]Assistant Professors, Jharkhand Rai University, Ranchi

ABSTRACT
Internet of things (IOT) is made up of various technologies, which supports advanced services in various application domains. Security and Privacy are a very important accepts for IOT application domains. These applications require Data confidentially, authenticity, integrity and access control within the IOT network. For users and things, Security is achieved by enforcing the security and privacy policies. Due to the different standards and communication stacks involved in traditional security solutions, it cannot be directly applied to IOT technologies. In IOT number of interconnected devices is expected to increase tremendously hence scalability is the biggest challenge for IOT development. This survey paper presents the available security protocols at respective IOT layers. A comparison of this information is done with respective to various security aspects and research gaps are identified.

INTRODUCTION

Internet of Things
Friedemann Mattern has discussed about the vision of the internet Of Things where everyday objects will be deployed as nodes using internet. These objects will be controlled remotely to achieve expected actions. Those will act as physical access points to internet service. Internet of Things is going to open up huge opportunities in economy field as well as individually. But at other side it also introduces risk along with technical and social challenges.

The internet of things (IOT) is the network of physical objects or devices, vehicles, buildings and other items embedded with electronics, software, sensors, and network connectivity which enable these objects to collect and exchange data. It includes sensors, Actuators, storage devices processing, localization and tracking, identification, communication etc.

In internet of things smart object is the central parameter which is developed with embedded communication and information technology. These Daniele Miorandi have discussed existing scenarios in Internet of things (IOT). Today, nearby two billion people around the world use the internet for a lot of applications like browsing the web, sending and receiving emails, using social networking applications, assessing multimedia content and services, and many other tasks. As maximum people will access the global information and communications infrastructure, a big challenge is arising related to use of the internet for communication, compute, dialogue and coordinate between the machines and smart objects. It is expected that, in next decade internet will be the most essential mechanism of classic network and networked objects. This new world of Internet will support for new ways of interacting, new way of entertainment, and new way of living new applications enabling new ways of working.

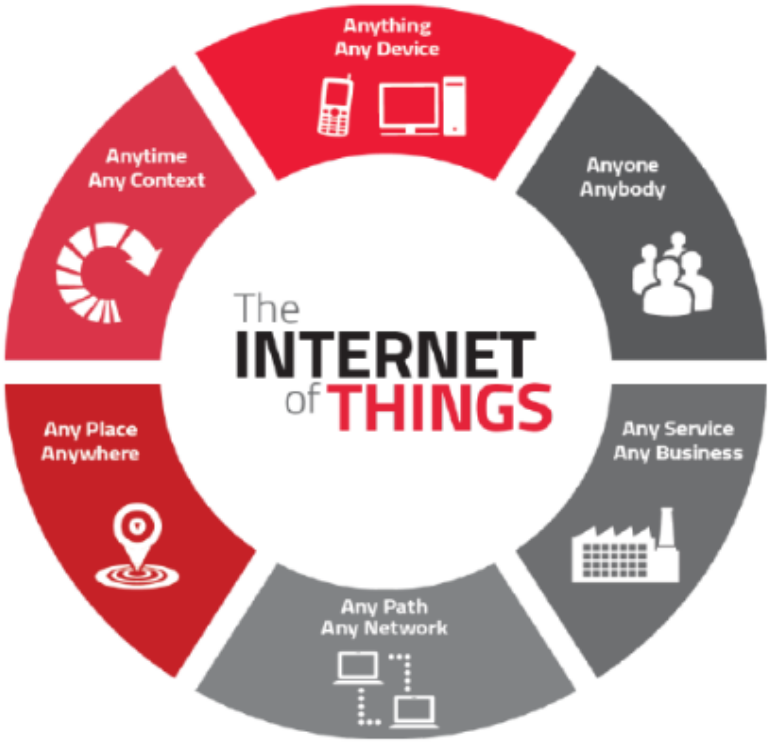


Fig: Topics make up the Internet of things

Need for Security in IOT

Kim Thuat et al, (2015) have discussed that the concepts of IOT is easy to understand but still lot of research work is expected in this area. Several aspects of IOT such as IOT applications and architecture are presently being discussed. Currently, major research work is going on in undertaking challenges associated with security, privacy, and trust and majority IOT devices will be developed. In IOT, every single physical object for example smart car, smart refrigerator or similar device will be connected to the internet to share the information. This will definitely increase the risk than before as personal data and business secret information will be shared through the internet.

S. Sicari et al. (2015) have discussed that IOT services requires modifications in security and privacy in various applications. In their paper they have surveyed various research directions in IOT security. Security is very important feature in terms of IOT development IOT consist of heterogeneous environmental made up of various technology and communication standards. Availability, confidentiality, privacy for users and things, authenticity among devices, integrity to access the remote device, will defined security and privacy policy is this parameter of security in IOT must be addressed properly so that suitable solutions can be designed and developed.

Ala Al-Fuqaha et al (2015) have a provided and overview of the internet of things (IOT) with emphasis on enabling technologies, protocols and applications issue. The current revolutions in Internet, mobile and machine -to -machine (M2M) technologies can be seen as the first phase of the IOT. In the coming years, the IOT is expected to bridge various technologies to enable new applications by connecting physical object together in support of intelligent decision making. Security plays an important role in the designing of above-mentioned concept.

THREATS IN IOT AND THEIR COUNTERMEASURES

Research Challenges in Internet of Things

Daniele Miorandi et al. (2012) have given an overview of the critical issues of services and technologies. To become measure research trends in future. This survey has provided many guidelines to researches and developers to work on major issues related to security, communications and similar challenges in field of IOT. Many use cases are identified in this survey which will be helpful as guidelines for researches so that the innovative technical solution can make IOT from research vision into reality.

Jorge Granjal et al (2015) have analyzed current technologies and various protocols which are used to add security in IOT. They have also analyzed research trends in this direction. Author have also done a thorough analysis and concluded about existing security approaches toward IOT and to protect communications in the IOT along with suggestions on future challenges and strategies in the mentioned area.

Rolf H. Weber (2015) discuss the manager in the context of IOT is a challenge of privacy and security. In future IOT will face the challenge of collections of data and maintain the security. For this is strong technologies must be developed with will take care of security in communications and storage of the data. A lot of work is expected to be done to solve the security and privacy issue in IOT. But these issues regarding IOT data have remained unaddressed. A lot of work is also expected to create industries standard which will maintain minimum level of privacy.

In case of transmissions or storage the data of sensitive matter. For example, health related data, financial information or some critical data of defense similar sensitive data must be maintenance properly there for working on security and privacy issue in IOT is now mandatory.

Security in IOT

Chen Qiang1 et al. (2013) have discussed the existing research is of network security technology. These are many problems in the security of internet of things (IOT) such as RFID tag security, privacy, protection, information processing security etc. All these issues must be addressed for better future of IOT.

Somia Sahraoui et al (2015) proposed a 6LOWPAN (IPv6 Over low power wireless personal area network) compression for the header of HIP (Host identity protocols) packets. Reem Abdul Rahman et al. (2016) discussed about Internet of things IOT which is a wireless communication network between smart object connected to the internet.

Table 1: Comparative chart security parameters

Table 1: Comparative Chart Security Parameters

Name of Layer\ Security Parameters	Confidentiality	Integrity	Authentication	Non-Repudiation	Fragmentation Attack Protection	End to End Security	Replay Protection	Internal Attack Protection
Physical/ MAC Layer	Y	Y	Y	Y	NA	Y	N	N
6LoWPAN Adaptation Layer	Y	Y	Y	Y	N	N	N	N
Network Layer	Y	Y	Y	Y	N	N	N	N
Transport Layer	Y	Y	Y	Y	N	Y	Y	Y
Application Layer	Y	Y	Y	Y	NA	Y	Y	NA

Y-Addressed N-Not Addressed NA- Not Applicable

It is a latest and fast developing market which connects object and people and also billions of gadgets and smart devices with the growth of IOT there is also increase in security thread of the Lind object there are server new published protocol of IOT wish focus on protecting critical data.

Table 1. summarizes the various security parameters with respect to communication layers. This table gives the research gaps in the security of IOT.

Communication Layers and Protocols Stack in IOT

Reem Abdul Rahman et al (2016) have given the idea protocol at respective layers. The widely used protocol such as IEE. 802 .15.4 at PHY/MAC layer, 6l0WPAN Adoption layer, at, and RPL at network layer are available in IOT. Constrained Application protocols (CoAP) is the application layer protocols designed as a equivalent of the HTTP. Various versions of CoAP have been developed which shows its significant role in various applications in the future of IOT.

Table 2 gives the IOT stack and respective security protocols for IOT.

Adding Security at the link layer

Roman et al (2011) have proposed security system for link layer. They have proposed key management system for sensor network which has added Security at link layer outlines the approach is not surface area into achieve into security while adding securities at link layer, care must be taken such as every node in the path will be trusted. Authors have a gone through many solutions to solve the problem of establishing sessions in between server and a client with respect to internet of things IOT where plants server network is formed from node in a wireless sensor network.

Rolf H. Weber (2015) has Hindi detail about sensor socket layer (SSL) which include ki exchange mechanism along with confidently and integrity it also provides authentication between internet hosts. But SSL has few drawbacks transport layer security is used over TCP which is not a preferred option for smart object communication the reason behind it is TCP connections used inadequate resource.

Table 2. IOT stack with security protocols

IoT Layer	IoT Protocol	Security Protocol
Application	CoAP, MQTT	User defined
Transport	UDP	DTLS
Network	IPv6, RPL	IPsec, RPL security
6LoWPAN	6LoWPAN	None
Data-link	IEEE 802.15.4	802.15.4 security

Granjal et al (2010) have and practically implemented the usage of new compressed 6LoWPAN security headers, with the Cryptographic algorithms typically used with the IP security architecture. Their practical evaluation study has shown that this is compatible with existing wireless sensor nodes, and it also gives the new technique which allows secure integration WSNs with IPsec and the internet.

Raza et al (2011) have presented the very first IPsec specifications and implementation for 6LoWPAN which is working on adaptation layer. They have evaluated their implementation and verified that it is feasible to use compressed IPsec to secure communications between hosts in the Internet and sensor nodes.

Security the IOT at the Transport Layer

Kim Thuat Nguyen el. (2015) has explained that TLS has been recommended by many standards specified by TETF for security services in IOT. However, as it has been discussed earlier TLS is not a wise choice with respect to the security in IOT. TLS uses reliable transport protocol like TCP which is based on congestion control algorithm. But it is not suitable in IOT which has constrained resources devices. Therefore, are tightly constrained environments Datagram Transport layer Security (DTLS) protocols is proposed which operates on unreliable transport protocol (UDP). It provides the same high security levels as TLS.

6LoWAPN Protocols for IOT

Konstantinos Rantos et al (2013) have discussed about the broad deployment of low power and lossy network (LLNs) connected to the Internet. It has realized many security issues regarding the protection of data. Such network now faces all kinds of security threats identified in traditional networks cannot directly be adopted by LLNS, due to the inherently limited capabilities of the embedded system that embraces them. This paper has focused on the security provided to LLNs nodes using 6LoWPAN adaptation format, one of the principal solutions adopted for communicating data over IEEE 802.15.4 networks. 6LoWPAN (IPv6 over low-power wireless personal Area network) Standard allows resources constrained devices to connect to IPv6 networks.

CONCLUSION

In this paper, we have carried out a thorough analysis of the security protocol and mechanism available to protect communications on the IOT. The majority of security protocols are briefly discussed in this survey paper, from this analysis, it is observed few gaps at various IOT layers as, fragmentation attack protection is absent in physical layer .it is also seen end to end security is not supported by physical layer and network layer. Replay protection is not supported in the physical layer network layer, network layer and 6LoWPAN layer, internet attack protection is absent in physical layer, 6LoWPAN layer. Transport layer and applications layer so more work is expected to do in this gap to secure internet of things for implementing it in a better manner.

REFERENCE

1. Friedemann Mattern and Christian Floerke Meier, "From the Internet of computer to the internet of things", Distributed Systems Groups, Institute for Pervasive computing Zurich, 2010.
2. Daniele Miorandi, Sabrina, Francesco DE Pellegrini, Imrich Chlamtac" Internet of things: vision, applications and research challenges" Ad Hoc Networks 10 (2012), journal homepage: www.elsevier.com/locate/adhoc
3. Kim Thuat Nguyen, Maryline Laurent, Nouha Oualha, "Survey on secure communications protocols for the Internet of things", Ad Hoc Networks 32(2015)17-31
4. S. Sicari, A. Rizzardi, L.A.Grieco, A.CoenPorisini, "Security, Privacy and trust in Internet of things: The road ahead " Computer Networks 76(2015)146-164.
5. Ala Al-Fuqaha, Mohsen Guizani, Mehdi Mohammadi, Mohammed Aledhari, Moussa Ayyash, "Internet of things: A survey on enabling technologies, protocols and applications ", Dot 10.1109/Comst.2015.2444095.Ieee communications, surveys & Tutorials.
6. Jorge Granjal, Edmundo Monterio, Jorge sa Silva" Security for the Internet of things: A survey of existing protocols and open research issues ", University of Coimbra, Portugal, IEEE communication surveys & Tutorials DOI 10.1109/COMST.2015.2388550.
7. Rolf H. Weber," Internet of things: Privacy issues revisited " University of Zurich, Switzerland, Computer law & security review 31(2015) 618-627, Published by Elsevier Ltd.

8. Chen Qiang, Guangri Quan, Bai Yu and Liu Y and, "Research on security issues of the internet of things" international Journal of future generation communication and networking vol.6, No.6(2013), pp.1-10 http://dx.doi.org/10.14257/ijfgen.2013.6.6.01
9. Somia Sahraoui, AZeddine Bilami," Efficient HIP based approach to ensure lightweight end-to-end Security in the internet of things",1389-1286 2015, Elsevier http://dx.doi.org/ 10.1016/j.comner.2015.08.002
10. Reem Abdul Rahman, Babar Shah "Security analysis of IOT protocol" A focus in COAP",978-1-4673-9584-7/16/531.00 2016 IEEE
11. Rodrigo Roman, Cristian Alearaz, Javier Lopez, Nicolas Sklavos, "Key Management system for sensor networks in the context of the internet of things ", computers & electricals engineering, Vol.37, pp.147-159,2011.
12. Jorge Granjal, Edmundo Monterio Jorge Sasilva, "Enabling network layer security on IPv6wirelass sensor networks ",978-1-4244-5638-3/10/26.00 2010 IEEE.
13. Shahid Raza, Simon, Duquennoy, Tony Chung. Dogan yazar, Thiemo Voigt and utz Roedig, "Security communication in 6LoWPAN with compressed IPsec ",978-1-4577-0513-7/11/2011 IEEE
14. Kim Thuat Nguyen. Maryline Laurent Nouha Oualha," survey on secure communications protocols for the Internet of things " http://dx.doi.org/10.1016/j.adhoc.2015.01.0061570-8705/_2015 Elsevier B.V.
15. Konstantinos Rantos, Alexandros Papanikolaou, Charalampos Manifavas, " IPsec over IEEE 802.15.4 for Low power and lossy networks",2013 ACM 978-1-4503-2355-0/13/11. doi.10.1145/2508222.2508240.

THE IMPACT OF CHATGPT ON STUDENT LEARNING: A REVIEW

Mousam Kumari[1] and Dr. Shashi Bhushan[2]
[1]Amity Institute of Information Technology (AIIT)
[2]Assistant Professor, Amity Institute of Information Technology (AIIT) Amity University Patna

ABSTRACT
This evaluation investigates how the language model ChatGPT, created by OpenAI, affects student learning. ChatGPT is a sophisticated AI system made to converse in human-like ways and offer guidance and knowledge on a range of subjects. The study investigates the advantages, restrictions, and potential difficulties related to ChatGPT integration in educational contexts. Thestudy includes a thorough evaluation of the studies and research that have been done on the application of ChatGPT in educational settings. By delivering individualised and on-demand guidance, providing explanations and clarifications, and encouraging lively and interesting dialogues, it examines how ChatGPT could improve students' learning experiences. The results imply that ChatGPT can be an effective tool for enhancing student learning. One prominent example is Chat GPT, a state-of-the-art language model developed by OpenAI. This study explores the impact of Chat GPT on student learning and research topics. The findings indicate that Chat GPT has had a profound impact on student learning in several ways. Firstly, the availability of Chat GPT as an educational resource has empowered students to pursue a wider range of research topics. The model's ability to generate accurate and relevant information in real-time has facilitated efficient information retrieval, enabling students to delve deeper into complex subject matter and explore niche research areas.

Keywords: ChatGPT, OpenAI, Artificial intelligence.

INTRODUCTION
The rapid advancement of artificial intelligence (AI) in recent years has led to a number of applications in a range of industries, including education [2] and healthcare [1]. AI systems may be trained to replicate the functions of the human brain using a large amount of data. AI, for example, may assist healthcare staff by synthesising patient data, interpreting diagnostic images, and highlighting health problems [4]. Improvements in administrative and academic support services have been made in education using AI applications [2]. One such illustration is intelligent tutoring systems (ITS). This might be used to simulate individualised, one-on-one education. A meta-analysis revealed that ITS normally had a somewhat positive effect on academic achievement. The creation of ITS maybe challenging, though, as it calls for the development of communication and feedback systems in addition to the design and distribution

of information. Recently created by OpenAI, the conversational chatbotChatGPT may make it easier for instructors to integrate AI into their classes. ChatGPT employs natural language processing to offer responses to user input that are human-like. Due to its remarkable performance in delivering responses that are coherent, organised, and instructional, it has attracted interest from all around the world [8]. Unexpectedly, theUniversity ofMinnesota Law School's ChatGPT passed each of its four tests [9]. The results demonstrate that an AI application can graduate from college even when its test results weren't (yet) exceptionally strong. Since its launch on November 30, 2022, ChatGPT has seen the fastest user application growth in history. In just two months, in January 2023, it had 100 million active users.

METHODOLOGY

To find relevant research articles, an organised search of the literature was done. Online repositories, conference proceedings, and academic databases were also searched. In order to establish common themes and conclusions, the chosen papers were then examined using a qualitative methodology a survey on students of Amity University Patna was conducted to study the impact of ChatGPT influence on changing pattern in learning.

LIMITATIONS AND CHALLENGES

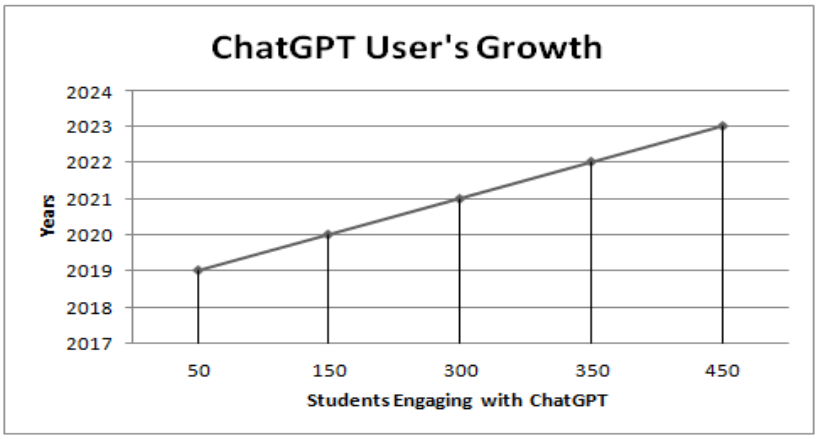
- Ethical factors: When deploying ChatGPT in educational contexts, some research studies emphasised the significance of ethical factors, such as privacy problems, algorithmic bias, and the necessity for human supervision.
- Dependency on Internet connectivity: ChatGPT depends on having an internet connection, which poses a problem when connectivity is scarce or unstable.
- Lack of Long-Term Studies: To fully comprehend the long-term effects of ChatGPT on student learning outcomes, longitudinal studies are required. Many of the evaluated research concentrated on short-term treatments.

IMPLICATIONS

The results of the reviewed research studies indicate that ChatGPT may have a favourable effect on engagement and learning outcomes for students. However, careful consideration of ethical issues, technological constraints, and the requirement for continuous study to understand its long-term impacts are all necessary before integrating ChatGPT into educational settings. These findings may be used by educators and decision-makers to guide their deliberations and create efficient plans for incorporating ChatGPT and other AI-powered technologies into educational situations.

Year	Total Students	Students Engaging with ChatGPT	Student-to-ChatGPT Ratio
2019	500	50	10:1
2020	500	150	5:1
2021	500	300	2.5:1
2022	500	350	2:1
2023	500	450	1.5:1

This table provides a sample data of the student ratio data over five years, showcasing the increasing engagement of students with ChatGPT. As the years progress, more students are benefiting from the interactive and personalized support offered by ChatGPT, resulting in a lower student-to-ChatGPT ratio.



CONCLUSION

In conclusion, ChatGPT has had a significant influence on students' learning. The educational experience has been transformed by an AI-powered chatbot by offering individualised help, prompt feedback, and access to a plethora of knowledge. ChatGPT has the potential to provide personalized learning experiences. It can offer tailored assistance and support to individual students based on their unique needs, preferences, and learning styles. By interacting with ChatGPT, students can receive customized feedback, explanations, and guidance, promoting a more personalized and adaptive learning environment. It has changed how students learn information and skills by improving student involvement, understanding, and cooperation. Despite the fact that ChatGPT is an effective tool, it's critical to keep in mind that actual teachers

are still essential for teaching students and encouraging higher levels of knowledge. We can provide students with a comprehensive and efficient learning environment by finding a balance between AI technology and interpersonal connection.

REFERENCES

1. Xu, L.; Sanders, L.; Li, K.; Chow, J.C.L. Chatbot for Health Care and Oncology Applications Using Artificial Intelligence and Machine Learning: Systematic Review. JMIR Cancer 2021, 7, e27850. [CrossRef] [PubMed]
2. Zawacki-Richter, O.; Marín, V.I.; Bond, M.; Gouverneur, F. Systematic Review of Research on Artificial Intelligence Applications in Higher Education—Where are the Educators? Int. J. Educ. Technol.High. Educ. 2019, 16, 39.
3. Bengio, Y.; Lecun, Y.; Hinton, G. Deep Learning for AI. Commun. ACM 2021, 64, 58–65. [CrossRef]
4. Aung, Y.Y.M.; Wong, D.C.S.; Ting, D.S.W. The Promise of Artificial Intelligence: A Review of the Opportunities and Challenges of Artificial Intelligence in Healthcare. Br. Med. Bull. 2021, 139, 4–15.
5. Steenbergen-Hu, S.; Cooper, H. A Meta-analysis of the Effectiveness of Intelligent Tutoring Systems on College Students' Academic Learning. J. Educ. Psychol. 2014, 106, 331–347. [CrossRef]
6. Afzal, S.; Dhamecha, T.; Mukhi, N.; SindhgattaRajan, R.; Marvaniya, S.; Ventura, M.; Yarbro, J. Development and deployment of a large-scale dialog-based intelligent tutoring system. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; Volume 2, pp. 114–121.
7. OpenAI. Available online: https://openai.com (accessed on 10 March 2023). 8. Zhai, X. ChatGPT UserExperience: Implications for Education. SSRN 2022, 4312418.

EMPLOYING MULTIPLE LINEARLY ARRANGED ArUco MARKERS AS REFERENCE SCALE FOR ENHANCING IMAGE-BASED LENGTH MEASUREMENT

Nand Kumar Rana
Assistant Professor, Department of Physics, J. N. College, Dhurwa, Ranchi

ABSTRACT

This study presents a novel approach utilizing multiple linearly arranged ArUco markers as a reference scale to enhance the accuracy and reliability of image-based length measurement algorithms. The proposed method addresses various challenges associated with conventional measurement techniques, such as camera calibration, exposure conditions, object colors and textures, and perspective distortions. The ArUco marker library is robust, adept at precisely identifying marker corners under diverse lighting conditions, camera orientations, and noisy environments, while ensuring computational efficiency in terms of processing time and power requirements.

The methodology involves arranging multiple ArUco markers in a straight line with a fixed size and equal spacing between them, thereby creating a set of equidistant points that form a linear reference scale. This arrangement offers superior linearity along the measurement axis in comparison to single-marker configurations. A significant advantage of the proposed method is that camera calibration is often not necessary. The OpenCV-contrib library and Python programming language are employed in this research work.

Experimental evaluations conducted with objects of varying lengths, textures, and colors demonstrate that the proposed method achieves high accuracy and repeatability in length measurement. The results showcase the technique's robustness against fluctuating lighting conditions and perspective distortions, with a mean error of less than 0.5% across all test cases.

The adaptability, robustness of ArUco marker detection, linearity of the proposed method, and widespread availability of smartphones make this approach a promising alternative to traditional length measurement techniques.

Keywords: image processing, Aruco markers, OpenCV, Python.

INTRODUCTION

Android smartphones are particularly suitable for developing computer vision applications due to several reasons. They are equipped with powerful processors, large amounts of memory, and high-resolution cameras, which are essential for handling the computational and image processing demands of computer vision applications [1]. Additionally Android's open-source operating system allows developers to have a high degree of control over the device's functionality [2]. Further, the availability of a wide range of libraries, such as OpenCV, and APIs that provide pre-built functionalities for image processing, making it easier and faster for developers to build computer vision applications [3].

BACKGROUND

Open-Source Computer Vision, or OpenCV, is a powerful library launched by Intel in 2000. It consists of over 2500 optimized algorithms for real-time computer vision and machine learning applications [4]. One of OpenCV's key advantages is its cross-platform compatibility. It supports a variety of programming languages, including Python, C++, and Java, and can be used on different operating systems, such as Windows, Linux, Android, and iOS. This versatility has facilitated easy prototype testing on desktop platform before developing the final Android application [5].

An important feature of OpenCV is its inbuilt support for ArUco markers, a type of fiducial marker used in many computer vision applications for tasks such as object detection and pose estimation. The ArUco markers was developed by the team of researcher headed by Dr. Rafael Muñoz-Salinas at the University of Jaén, Spain[6]. The ArUco markers have well-designed algorithms to eliminate errors due to various types of distortions, such as geometric transformations and changes in lighting conditions, viewing angle, etc [7, 8,9]. This makes them a readily accessible tool for a wide array of applications. Moreover, the detection algorithms for ArUco markers are highly efficient, allowing for real-time applications, even on devices with limited resources like smartphones. The design of ArUco markers, which allows for a large number of unique identifiers, enables the simultaneous tracking of multiple markers in a scene, adding to their versatility [10, 11].

METHODOLOGY

The proposed method of measuring height uses ArUco markers placed at known distances to create a reference scale. A series of ten markers, with IDs from 0 to 9, are placed vertically in a straight line. Each marker is placed 10 cm apart from its neighbour and is of 10 cm size. Once an image of this setup is captured, the detect.Markers() function from the cv2.aruco library is used to detect all the corners of the ArUco markers in the image [12] which act as reference points fig. 1. If any marker(s) are missing, the program has the safety feature to prevent errors [13].

Table 1. Aruco corners and the corresponding real world height measured in cm.

Marker ID, Corner No.	Ref. in cm	Marker ID, Corner No.	Ref. in cm	Marker ID, Corner No.	Ref. in cm	Marker ID, Corner No.	Ref. in cm
0, 3	0	3, 3	60	6, 3	130	9, 3	190
0, 0	10	3, 0	70	6, 0	140	9, 0	200
1, 3	20	4, 3	80	7, 3	150		
1, 0	30	4, 0	100	7, 0	160		
2, 3	40	5, 3	110	8, 3	170		
2, 0	50	5, 0	120	8, 0	180		

Fig. 1 Lab implementation for height measurement.

To calculate the height of a pixel row through the reference point location on the image, OpenCV functions cv2.findHomography () and cv2.wrap Perspective () are used. These actions make a perspective correction of the image and normalize it so that the number of pixel row per mm of the image is constant with a known value. Thus, the height of any given point can be directly calculated from its pixel row number [14].

Given that each ArUco marker corresponds to 10 cm in the real world see Table 1. , a linear transformation can be applied to the pixel measurements to convert them into real-world units. This allows us to map a pixel row to a certain height in mm in the real world.

CALCULATIONS

To get the approximate resolution of measurement using the main 64 MP rear camera of the Samsung Galaxy A71 smartphone, which has a focal length of 26mm and a pixel size of 0.8µm [15], the following steps are considered. If the picture of the ArUco pattern considered in this work is taken from 5 meters from the object, then a height of 1mm forms an image of size I=(O/d)*f, where I is image size, O is object size, d is the distance of the object from the camera, and f is the focal length of the camera. Using these values, we find that each mm division on the real-world scale forms an image of 5.2 µm on the camera sensor. Given that the camera sensor has 0.8µm pixel size, this image is sensed (in vertical or horizontal position) by approximately 6 sensor pixels. Therefore, theoretically, if the image is captured from a distance of 5 meters using the above camera, the measurement resolution is of ⅙ mm or 0.17 mm [16].

However, this theoretical value is not realizable due to a number of factors such as radial / tangential distortions, chromatic aberrations, vignetting, diffraction, etc. In the presented work, the error due to distortion and camera imperfections are minimized due to a large number of closely spaced reference points [17].

CONCLUSION

The presented research work aims to provide a novel method for height measurement using machine vision, which can be crucial in calculating the BMI of children. The method utilizes Android smartphones and OpenCV's support for ArUco markers. The proposed method has shown promising results and can potentially simplify and improve the process of obtaining accurate height measurements for BMI calculations.

Acknowledgements: The author acknowledges the Department of Science & Technology (DST), Ministry of Science and Technology, Government of India, New Delhi, for funding the research

Project Title: - Automatised Body Mass Index (Bmi) Measurement System For The Assesment Of Malnutrition In Children.

Sanction Ltt. No: Tdp/Bdtd/47/2021(G) Dated 22-11-2021

REFERENCES

[1] "Android: An Open Handset Alliance Project". Open Handset Alliance. Retrieved 2023-05-11.

[2] "Open Source". Android Open-Source Project documentation.

[3] Bradski, G., & Kaehler, A. (2008). Learning OpenCV: Computer vision with the OpenCV library. " O'Reilly Media, Inc.".

[4] "Welcome to OpenCV-Python Tutorials's documentation!". OpenCV.

[5] "Cross-Platform". OpenCV documentation.

[6] Muñoz-Salinas, R., Medina-Carnicer, R., Madrid-Cuevas, F. J., & Marín-Jiménez, M. J. (2018). Mapping, localization, and path planning in image-based navigation for autonomous vehicles: a survey. Artificial Intelligence Review, 50(1), 35-61.

[7] Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F. J., & Marín-Jiménez, M. J. (2014). Automatic generation and detection of highly reliable fiducial markers under occlusion. Pattern Recognition, 47(6), 2280-2292.

[8] Garrido-Jurado, S., Muñoz-Salinas, R, Madrid-Cuevas, F. J., & Medina-Carnicer, R. (2016). Generation of fiducial marker dictionaries using mixed integer linear programming. Pattern Recognition, 51, 481-491.

[9] Romero-Ramirez, F. J., Muñoz-Salinas, R., & Medina-Carnicer, R. (2018). Speeded up detection of squared fiducial markers. Image and Vision Computing, 76, 38-47.

[10] Medina-Carnicer, R., Madrid-Cuevas, F. J., Muñoz-Salinas, R., & Carmona-Poyato, Á. (2011). Real-time high-speed video processing with FPGAs. Springer Science & Business Media.

[11] "ArUco Marker Detection". OpenCV documentation.

[12] "Marker Detection". OpenCV documentation.

[13] Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F. J., & Marín-Jiménez, M. J. (2014). Automatic generation and detection of highly reliable fiducial markers under occlusion. Pattern Recognition, 47(6), 2280-2292.

[14] "Perspective Transformations". OpenCV. Retrieved 2023-05-11.

[15] "Samsung Galaxy A71". Samsung. Retrieved 2023-05-11.

[16] Zhang, Z. (2000). A flexible new technique for camera calibration. IEEE Transactions on pattern analysis and machine intelligence, 22(11), 1330-1334.

[17] Heikkila, J., & Silven, O. (1997). A four-step camera calibration procedure with implicit image correction. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 1106-1112). IEEE.

NATURAL LANGUAGE PREPROCESSING OF TEXT DATA FOR EFFECTIVE TEXT CLASSIFICATION IN DEEP LEARNING

Niraj Kumar and Prof. Subhash Chandra Yadav
Department of Computer Science & Engineering, Central University of Jharkhand, Ranchi

ABSTRACT
In the era of data, data is generated in real-time. So, the organization can utilize the large collection of data in a data-driven process for making many decisions to improve their business. Today social media platforms produce such data in the large amount. The data available on social media are found generally unstructured. So, the study proposes pre-processing methods for text data. Preprocessing of text data is a key instant in the processing of text classification tasks. Text preprocessing methods performed different tasks sequentially to make the text data valuable information such as to convert multiple forms of the same word in form of one word. Preprocessing technique has given a very significant role and is widely used in the deep learning task. The text classification the basic phase of preprocessing includes processing features, and extracting admirable features against all the features in the dataset. This paper describes several preprocessing techniques and tools for text classification of English language text data.

Keywords: Text classification, Stop word removal, pre-processing, Tokenization, Streaming

1. INTRODUCTION
Text classification is done in two phases meaningful information is extracted from the text-based data and discover the knowledge, establish relationships among them and assign pre-defined categories according to their content from the text data provided [1].Text classification also means the auto-classification of online electronic documents such as news article classification, Question answering tasks etc. In natural language processing and other application based on textual data text classification plays a very important role, particularly when there is a huge volume of readily available electronic text data like micro blogging, digital library, and long text news article. Before Text classification, there are a series of the sequentially performed subtask. They are called natural language pre-processing includes refining or cleaning textual data, establishing the rule, and representing text in the classification [2]. The main purpose of preprocessing the text data is to acquire the key features or terms from a given text document or dataset and establish relevance between important words and the text document and improve relevance by connecting words and their associated class. It has already been shown that the time spend in the preprocessing of textual data is up to 50 % to 80 % of the whole text classification task. The research paper includes the preprocessing technique for text data classification tasks used in the deep learning model.

1.1 Text Pre-processing
The procedure of cleaning and preparing the text for classification is known as text data pre-processing. Online writings typically have a lot of noise and useless components like HTML tags, scripts, and ads. In addition, many of the words in the text have not much impact on the text's overall orientation. Keeping those words increases the problem's high dimensionality and makes categorization more challenging because each word in the text is treated as one dimension. Decreasing text noise should enhance the classifier's performance and accelerate the classification process, thus helping in real-time text classification. There are various steps in the entire procedure: Digital text cleaning, white space disposal, stop word elimination, extending abbreviations, stemming, handling of negations, and feature selection.

1.2 Text preprocessing steps

The purpose of preprocessing text data is to depict each text data into distinct words and represent them in the form of a feature vector. Selecting relevant features from the provided text data pre-processing is the necessary step in indexing text documents. The process is called the tokenization of text data or attributes. The text data is represented in the vector space of low dimension by considering the component features and associated weight. All the non-informatics tokens are removed such as numbers, the special character used in the sentence and stop words. All the remaining features in the next step are converted to their root words as per the dictionary by using the stemming method.

1.2.1 Tokenization

The first step of the NLP process is gathering the data (a sentence) and breaking it into understandable parts (words).

Tokenization is the first step in the NLP process, it collects the whole text data and split it into smaller parts individual word-wise as unique identification, which can be identified and can be assigned meaning.

As an Example "I feel low energy I m just thirsty"

For this text data sentence for the machine to understand, tokenization is done on the whole sentence to split it into individual unique words. It does like this:

'I' 'feel' 'low' 'energy' 'I' 'm ' 'just' 'thirsty'

It looks simple but splitting sentences into words provides a machine to understand the whole sentence as well as the sentence in the context of words. This helps the machine to understand the role of words in larger text.

1.2.2 Stop Word Removal

In any sentence or text, many words appear repeatable but are meaningless, they are used only to connect the sentence and are called stop words. It is generally understood that stop words do not contribute to the context. The high occurrence of stop words is an obstacle to understanding the context of the sentence in the text classification task. In English, 400 to 500 stop words are there. Very frequently used stop words in a document are 'is',' am',' are',' they',' he', etc. There is no use of stop words in the classification of textual data, so must be removed [7].

1.2.3 Stemming

The process of identifying the root word of the written word in the text messages by eliminating the affixes from the features is called stemming. Many fields of language research, including Arabic [12], cross-lingual retrieval [13], and multi-language manipulations [14], utilize stemming mechanisms. There are many stemming methods, to guarantee that words are reduced to their root forms. After stemming the different phrases of the same words used in a given language can be represented by a single root or stem, resulting in the document dictionary size being reduced. Among of the many frequently utilized stemmers in information retrieval is the Porter's stemmer [15]. In this investigation, Porter's stemmer was employed. By doing this model will be able to reduce the features in the defined features space and the features of different types are merged into single features. This technique identifies the root word of all the related words. As an example, we encounter the words programming, programs, and programmer all words are steamed to the root word program. The main aim of this method is to decrease the number of features vector and to accurately match the original words, saving memory and execution time. Stemming is shown in figure 1.

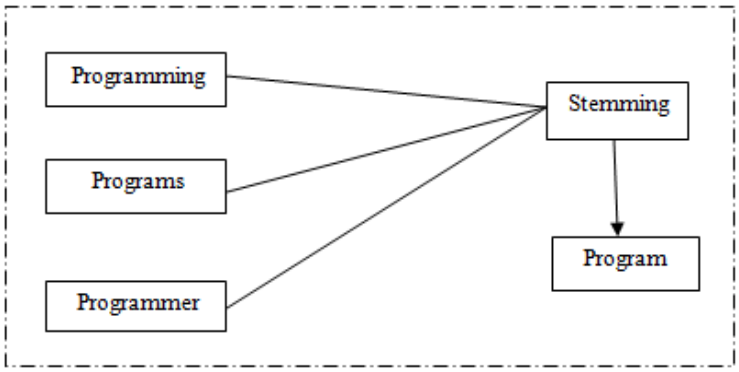


Figure 1 Stemming Process

1.2.4 Lemmatization

The lemmatization process eliminates inflectional endings of words by using linguistic and morphology examination of words, consequently, words will take on their dictionary expression. By using a thesaurus, lemmatization matches synonyms of word also so that the word 'hot,' and the word 'warm' is matched. Lemmatization is the process of combining words that share a root or lemma but differ in their inflexions or semantic derivatives into a single unit for analysis. The lemmatization technique is used in robots to communicate. Lemmatization produces better output in comparison to stemming. Lemmatization algorithms' main drawback is that they are considerably slower than stemming techniques. Lemmatization process sample output is exhibited in Table 1

Table 1. Comparison of Stemming and Lemmatization

Word	Stemming	Lemmatization
computers	compute	computer
feet	feet	foot
information	inform	information
informative	inform	informative

2. RELATED WORK

In text data classification the preprocessing stage converts the original text structure into raw data to find out the most important word in the text, which is contributing to categories the entire text data to their corresponding defined class. This is the primary stage where the representation of a series of words of each text data in the dataset is an index. Forman [3] has done a comparative study on the features selection matrix for text classification of the high dimensional region, considering the support vector method for classifying binary class problems, and found that the performance was increased with different features selection metrics. Another contribution was the evaluation of methodology by selecting the combination of preprocessing metrics as per the task and finding the chance of getting more accuracy. Garg et al. [4] classified text to find out the opinion of the people on a particular product. The study includes the collection of the latest tweet from Twitter API and filters the text data by performing preprocessing techniques like useless word removal, and stemming without using Deep learning methods. He proposes a model which is based on the multiple of the value of the adverb instead of the sum of all the tweet sentiments. It also improves the mapping word for text classification. Mineau et al.[5] propose a vector space architecture for the categorization of text by using a new technique for determining weight features of the word called Confweight. The very common technique in text classification is TFIDF; author discussed the learning process of TC. Mihalcea et al.[6] studies text classification by considering the distribution of frequency of words in any text document, dispersion with concentration and feature extraction. The study

was to classify the Chinese text of a very large dataset collection of the real world. Ozturkmenoglu and Alpkocak [16] examined three distinct lemmatizers in order to extract data from a Turkish dataset. Their findings demonstrated that, even when only a small number of terms are used in the system, lemmatization actually enhances retrieval accuracy. Gupta et al. [17] investigated by combining partial lemmatization and stemming to form a model. The model was tested on the retrieval of the Hindi language. When compared with the conventional ways, their model produced considerable improvements.

3. REPRESENTING TEXT DOCUMENT

Usually, in a text document each word in a text data is represented in the vector form(x, c) where $x \in R^n$ is the vector space representation of the text sentence and c is the determined class label. Each dimension space is a representation of its features vector

and its assigned weight, which is calculated by the occurred frequency of each word in that text sentence. This paper represents each text data of the entire dataset in document vector s as $c=(c_1,c_2,.....,c_n)$

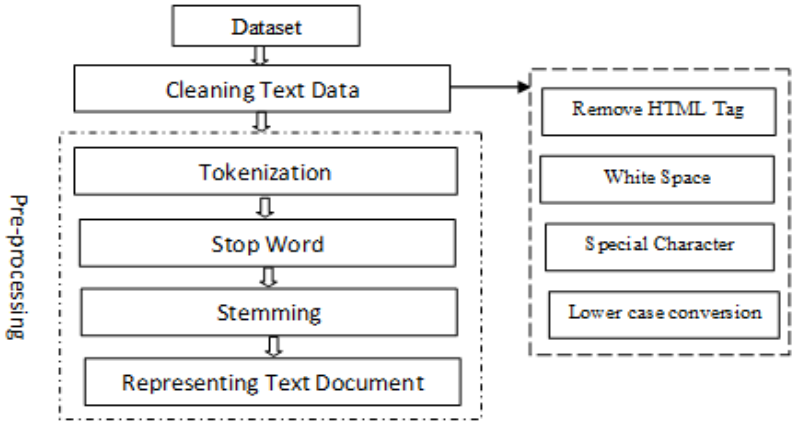


Figure 2 Steps for Text Pre-processing

3.1 Indexing Technique

The prime purpose of text document indexing is to enhance competency by extracting an important set of words used to describe the text document. Indexing includes an appropriate set of keywords which describe the whole text corpus and assign index weight to all the keywords of each text, thus transforming all the text into the vector of all essential keyword weights. The weight assigned to all the keywords is based on the frequency of particular words in a text document and the number of texts using those words.

3.1.1 Term Weighting

In the text classification model, the representation of text is done in the form of a vector. An important concept which determines the success and failure of the model in the task of text classification is term weighting. Since each word has its importance at a different level in a text, the assigned weight of each term associated is a vital indicator [8]

The three leading components which affect the consequence of the term in the context of text are Term Frequency, Inverse Document Frequency and Document normalization [9]. Term frequency is an assigned weight of each word that appears within a document, the weight is determined by the frequency distribution of a particular word w presence in a text t. Term frequency specifies the word's importance in a document. Inverse document frequency is determined by dividing the number of words contained in that document by the document containing that word in the whole corpus. It demonstrates word importance in the entire dataset [10]. TF/IDF methods use both the technique TF and IDF for assigning the weight of the term.

In text data classification TF/IDF idea is much more popular and the other entire weight-assigning scheme uses the concept of the TF/IDF technique [11]. For a collection of sentence 'S', a word't', and a specific sentence s in S, the weight (w) is calculated by given equation 1.

$$ws = ft, s * \log(|S| / Ft, s) \quad (1)$$

Where f (Term frequency) denotes the number of times't' appears in a sentence's'.

|S| is dataset size

F (IDF) representing the number of sentences in which's' comes in S.

4. EXPERIMENT PROCEDURES AND RESULT

4.1 Dataset

The dataset in the experiment is WASSA-2017[18], which determines the emotion a tweet's speaker is experiencing. The dataset contains collections of tweets labeled in four categories fear, anger, sadness and joy in 6429 rows by using the best-worst scaling method. The distribution of class labels is shown in Table 2 and the sample dataset with emotion label is displayed in Table3.Dataset after applying common cleaning task is displayed in Table 4.

Table 2. Distribution of class in Dataset

Class Label	Fear	Anger	Sadness	Joy
No. of Records	2252	1701	860	1616

Table 3. Sample Dataset with Emotion label

index	Tweet	Emotion
4782	@GenevievePere23 @swifttwinner13 WOW you are calling me bully bc I correct a typo? harsh	fear
4798	@alyssaxbeauty nightmare before Christmas	fear
5889	I wanna go to blithe and read w the goats :((	joy
1984	Lets start there	fear
5	My blood is boiling	anger

Table 4. Dataset after applying common data cleaning technique

index	Tweet	Emotion
5889	wanna go blithe read goats	joy
5	blood boiling	anger
4782	wow calling bully bc correct typo harsh	fear
4798	nightmare christmas	fear
1984	lets start	fea

4.2 Techniques Comparison

Two different techniques are used in classification of textual data to judge the text data preprocessing. The descriptions are given below in Table 5

Table 5. Pre-processing technique

Techniques	Description
Technique 1 with Stemming	This process is called stemming. In this process, different forms of the same words are converted to their original root word, which also makes the word length smaller. By doing this the computational time taken in the process is less.
Technique 2 with Lemmatization	Lemmatization is similar to the steaming process, but in the lemmatization process, it returns the words in their dictionary form.

To use the above-mentioned two techniques, the whole dataset was processed by common filtering steps of text data and after that mentioned technique in table 5 was used. The defined label class was stored and it was converted into the numeric value according to their type of categories. The steps shown in figure 3 identify the used technique and the performance.

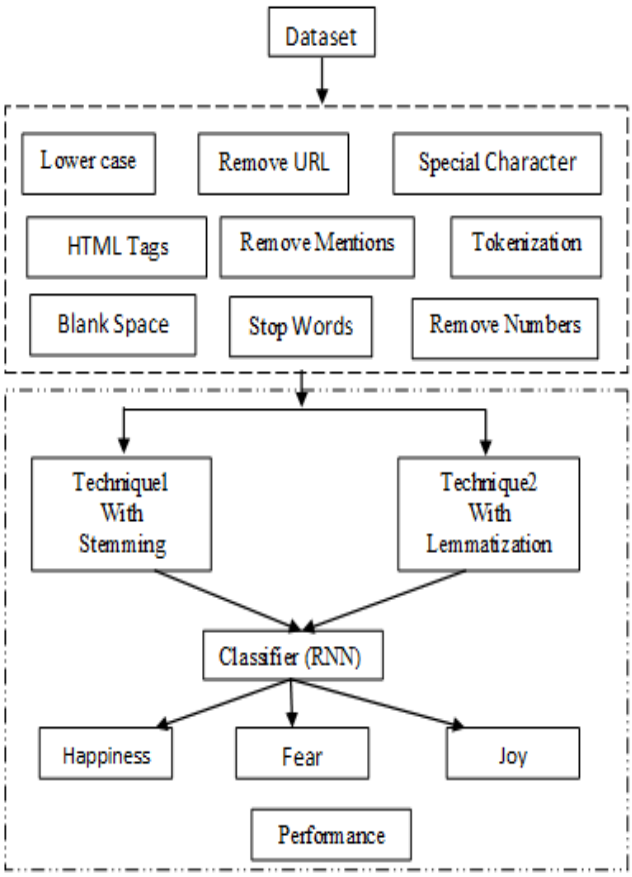


Figure 3 Selection of best performing preprocessing technique

The dataset pre-processed by using stemming and lemmatization technique is shown in Table 6 and Table 7 respectively.

Table 6. Pre-processed using Stemming

index	Tweet with Stemming	Emotion
5889	wanna go blith read goat	joy
5	blood boil	anger
4782	wow call bulli bc correct typo harsh	fear
4798	nightmar christma	fear
1984	let start	fear

Table 7. Pre-processed using Lemmatization

Index	Tweet with Lemmatization	Emotion
5889	wanna go blithe read goat	joy
5	blood boiling	anger
4782	wow call bully bc correct typo harsh	fear
4798	nightmare christmas	fear
1984	let start	fear

4.3 Performance Evolution

The experiment was conducted in a text-based document, which is divided into two sets the training set and the testing set. In the training set the model identified the significant word which describes the whole sentence and recast all documents in a collection of words. The testing set is used to compare perform of the classifier. The model was designed by using Bi-LSTM with an attention mechanism and the softmax function as an activation function. The performance of the model's by using both the pre-processing technique was assessed using standard performance metrics, which includes Mode accuracy, Precession, Recall and the support of all the class label. Accuracy comparison by using technique 1 is given in the Table 8. An accuracy comparison matrix of Technique-2 with lemmatization is displayed in Table **9**.

Table 8. Accuracy comparison matrices of Technique-1 with Stemming

Class	Precision	Recall	F1-score	Support	Av. Accuracy
Joy	1.00	0.99	0.99	154	80%
Anger	0.98	0.99	0.99	177	
Fear	1.00	0.99	1.00	118	
Sadness	0.95	0.95	0.95	66	

Table 9. Accuracy comparison matrices of Technique-2 with Lemmatization

Class	Precision	Recall	F1-score	Support	Av. Accuracy
Joy	0.98	0.98	0.98	144	97.86%
Anger	0.99	1.00	1.00	170	
Fear	1.00	0.99	1.00	142	
Sadness	0.98	0.95	0.97	59	

From the above Table 8 and Table 9, it is observed that pre-processing has a great impact on the deep learning method-based classification model. The result also reveals that preprocessing with Lemmatization has somewhat more impact than stemming. The classifier obtained result proves that preprocessing technique plays an important role in cleaning and providing significant data to the classifier and simultaneously it increases the accuracy of the model.

5. CONCLUSION

In the text classification task, we extract important information from a given text data called feature selection and try to find a statistical pattern from the given large dataset, which defines the text data. Before the dataset is given to the model the dataset goes through different pre-processing techniques such as removing stop words from the text, stemming the different words to their original words, Lemmatization etc. This paper describes complete information for the classification of text data preprocessing techniques and a comparison of the suitability and accuracy obtained from two pre-processing techniques used in the model. It is to be observed from the experiment that there is an impact of preprocessing technique in text classification tasks in terms of accuracy.

REFERENCES

[1] K.Aas and A.Eikvil, “Text categorization: A survey", Technical report, Norwegian Computing Center, June, 1999.

[2] Antonie, M. L., & Zaiane, O. R. (2002). Text document categorization by term association. In Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on (pp. 19-26). IEEE.

[3] Forman, G. (2003). An extensive empirical study of feature selection metrics for text classification. Journal of machine learning research, 3(Mar), 1289- 1305.

[4] S. Bhat, S. Garg and G. Poornalatha, "Assigning Sentiment Score for Twitter Tweets", 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2018.

[5] Soucy, P., & Mineau, G. W. (2005, July). Beyond TFIDF weighting for text categorization in the vector space model. In IJCAI (Vol. 5, pp. 1130-1135).

[6] Shi, L., Mihalcea, R., & Tian, M. (2010, October). Cross language text classification by model translation and semi-supervised learning. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (pp. 1057-1067). Association for Computational Linguistics.

[7] Xue, X. and Zhou, Z. (2009) Distributional Features for Text Categorization, IEEE Transactions on Knowledge and Data Engineering, Vol. 21, No. 3, Pp. 428-442.

[8] Salton, G. and Buckley, C. (1988) Term weighting approaches in automatic text retrieval, Information Processing and Management, Vol. 24, No.5, Pp. 513-523.

[9] Karbasi, S. and Boughanem, M. (2006) Document length normalization using effective level of term frequency in large collections, Advances in Information Retrieval, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Vol. 3936/2006, Pp.72-83.

[10] Diao, Q. and Diao, H. (2000) Three Term Weighting and Classification Algorithms in Text Automatic Classification, The Fourth International Conference on High-Performance Computing in theAsia-Pacific Region, Vol. 2, P.629.

[11] Chisholm, E. and Kolda, T.F. (1998) New term weighting formulas for the vector space method in information retrieval, Technical Report, Oak Ridge National Laboratory.

[12] Larkey, L. S., Ballesteros, L., & Connell, M. E. (2002, August). Improving stemming for Arabic information retrieval: light stemming and co-occurrence analysis. In Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 275-282).

[13] Xu, J., Fraser, A., & Weischedel, R. (2002, August). Empirical studies in strategies for Arabic retrieval. In Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 269-274).

[14] Wechsler, M., Sheridan, P., & Schäuble, P. (1997). Multi-language text indexing for internet retrieval. In Computer-Assisted Information Searching on Internet (pp. 217-232).

[15] Hooper, R., & Paice, C. (2005). The Lancaster stemming algorithm. University of Lancester.

[16] Ozturkmenoglu, O., & Alpkocak, A. (2012, July). Comparison of different lemmatization approaches for information retrieval on Turkish text collection. In 2012 International Symposium on Innovations in Intelligent Systems and Applications (pp. 1-5). IEEE.

[17] Gupta, D., Kumar, Y. R., & Sajan, N. (2012). Improving unsupervised stemming by using partial lemmatization coupled with data-based heuristics for Hindi. International Journal of Computer Applications, 38(8), 1-8.

[18] S. M. Mohammad and F. Bravo-Marquez, “Wassa-2017 shared task on emotion intensity,” arXiv preprint arXiv:1708.03700, 2017.

DESIGN OF REAL TIME AUTOMATIC EMOTION RECOGNITION (AER) BASED MUSIC RECOMMENDATION SYSTEM

Dr. Rashmi Shekhar[1] and Nitish Kumar Singh[2]
[1]Associate Professor and Assistant Director, Department of Amity Institute of Information Technology, Amity University Patna.
[2]Post Graduate Student of Master of Computer Applications, Amity University, Patna

ABSTRACT
Automatic Emotion Recognition (AER) *is one of important feature extraction system in advanced deep learning where it is can extract the features of the face and the method based on KNN, Haar Cascades and Convolution Neural Network (CNN). To code for better extraction features from face we need proper training model and mathematical features and also, we need to have a better prediction from operating system either it can be raspberry pi or windows OS. The purpose of this paper is to make a automatic face recognition system via deep learning methodologies with automatic recommendation system to play a music as per the emotion predicted by the machine like (Happy, Angry, Surprise, Sad etc). As per the situation music recommendation one of the important feature classification which is obtained from data science.*

Keywords: Emotion Detection, Haar Cascade, Face Detection, Machine Learning

INTRODUCTION
Human face has infinity of Expressions' used for many purposes like Acting, Educating people etc, No words about it because it is essential to everyone to face another one. But how to detect the expression by a machine or robot with the help of AI, that is a challenge to human. Now it is proposed idea called as Human Machine Interface i.e. a interaction between human & machine. Which can identify the human emotion with Deep learning feature extraction system. This type of technology is used for Security purposes in extraction important facial features and also used for assistance purpose in the wireless gesture world through computer imaging technique. This technology is commonly promoting towards android apps for feature extraction in medical which is actually playing a vital role to understand the facial expression based on the medication of drug evaluation and also checking the autistic children, those are unable to communicate socially and scared to interact with people who are restricted to repetitive interests. This type of applications can also be used for facial recognition system in Educational video which can record and understand the facial expressions when teaching This proposed paper has real time interface in image feature extraction, Some emotions can change the features or learn the expressions by the children for better understanding. Let us consider a live example of Human Interactive robot-like Sophia, Hanson Robotics and Various emotional robots which are actually helping the old age with their expressions in daily life. Understanding human emotion and giving a response by robots will be the next extent in daily life and success of human mankind for successful innovation of robotics.

LITERATURE REVIEW
Over the last fifteen years, an excessive investigation has been completed to recognize emotions by using speech statistics. Cao et al. [10] Many researchers have published there articles and ranking SVM & KNN method great output in the accuracy for synthesize information about emotion recognition to solve the problem of binary classification. This ranking method, instruct SVM algorithms for particular emotions, treating data from every utterer as a distinct query then mixed all predictions from rankers to apply multi-class prediction. Ranking SVM achieves two advantage, first, for training and testing steps in speaker- independent it obtains speaker specific data. Second, it considers the intuition that each speaker may express mixed of emotion to recognize the dominant emotion.

RELATED WORK

The research of this entire project was taken one year of time to understand the facial features with computer vision Technology which can detect the features of eye and mouth with geometrical representations the extractions of the features are totally dependent on principal of mathematics. In the existing system the images classification only is happening with some facial emotions like happy, sad, angry, fear, surprised, neutral, etc. Humans express millions of facial expression where brain creates and understands the decision making skills. The same way how brain is functioning now the machine perform the tasks with neurons understanding, each neuron has a pre trained model called as epoch, if number of epoch are trained that may features can be expressed by the machine. But there is some limitation in the real world because of high computer accuracy need. So only some features can be extracted and expressed in the computer vision. In this existing model two multiple machine learning algorithms were used KNN and Haar Cascades which are dependent at the back end to perform the task in the easiest way. The below gives the clear explanation of the KNN & Haar to understand about the research paper.

KNN

It states that K-Nearest Neighbors Algorithm. Which comes under the supervised learning algorithm which makes classification or grouping of individual data points? Basically KNN is used for regression or to solve classification problems in real life, example predicting the trouser size by giving age and weight dataset, like same we are giving predetermined data set which consists of Emotions like different class labels (happy, sad, angry, fear, surprised, neutral, etc and then applied to the model then its starts predicting. By using KNN which recap the goal to identity which class is near to the prediction score for given input face. That is why KNN is called as distance matrix. The simple feature of KNN is transforming the data inputs into a feature vectors. Then the algorithm process by distance between mathematical values of this point. The Facial Emotion Recognition (FER) has four vital steps to proceed further. in the early stage to detect the face in the image and draw rectangle contour around the face, the second step detects land mark in the face region but the third step is about extracting spatial features from the face by using computer vision then finally Feature extraction classifier takes place and produce the recognized result, so the user will get the label on the contour.

Haar Cascades

Haar Cascade Technique is first introduced by Viola and Jones. Which is one of the finest image processing model in the entire image and video processing's. Which is backend predicting model for open computer Vision because it is also known as computational object detector.

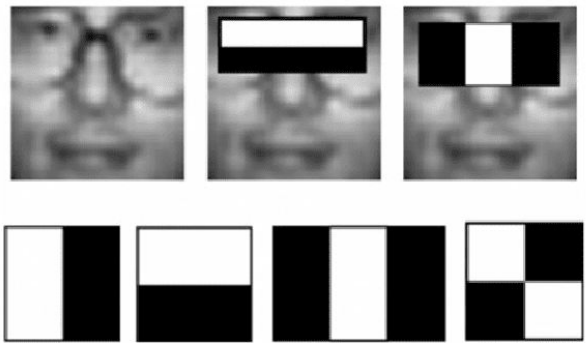


Fig: Haar Cascade Object Detection Algorithm

Haar Cascades are used for locating the objects in the any images like (Nose, Eyes, Mouth etc) with help of positive contains face and negative terminologies determines there is no faces in the images. By this we can predict the statistical inference of emotion.

CNN

Convolution Neural Network is a technique is a type of deep learning neural network concept which has a various step of techniques flattens, pooling and 2D dimensionality method. Simply to understand different layers Input layer, Hidden Layer, Output Layer. The data is fed into model which is technically algorithm states that feed forward method. Then calculation of error function will perform. If the error should be solved then we have to go for Back propagation.

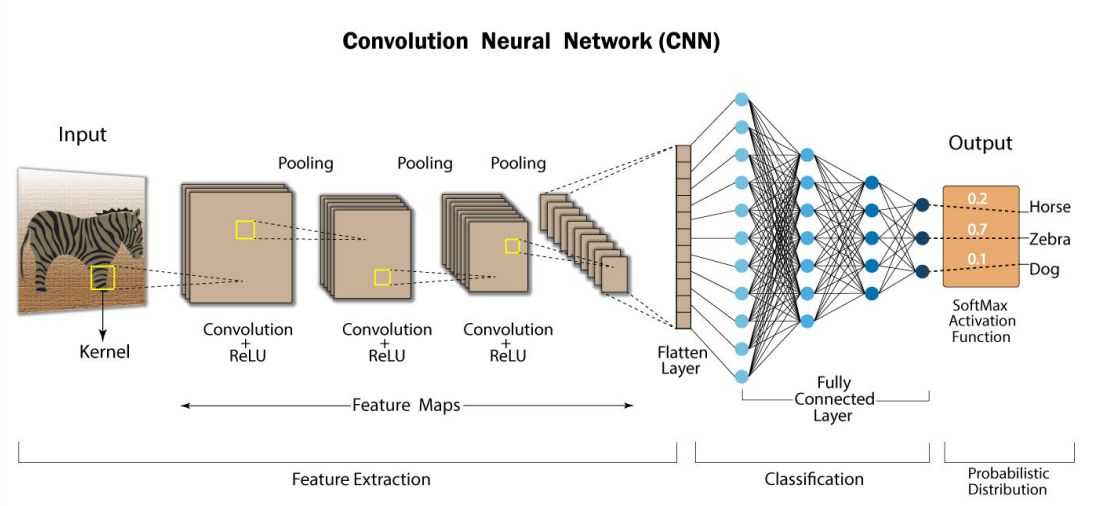


Fig: Convolution Neural Network

Working Operations

As we know that the entire project will be working on above two algorithms i.e., CNN and HAARCASCADES, but the process of working model is represented by numbers that corresponds to the pixels intensities. The procedure is to convert image into array from large data sets converting the dimensions 48 X 48 Pixels, this process is to convert the image to landmarks by using a library called Dlib. The frontal face detector is to detect rectangle contour once the prediction is detected.

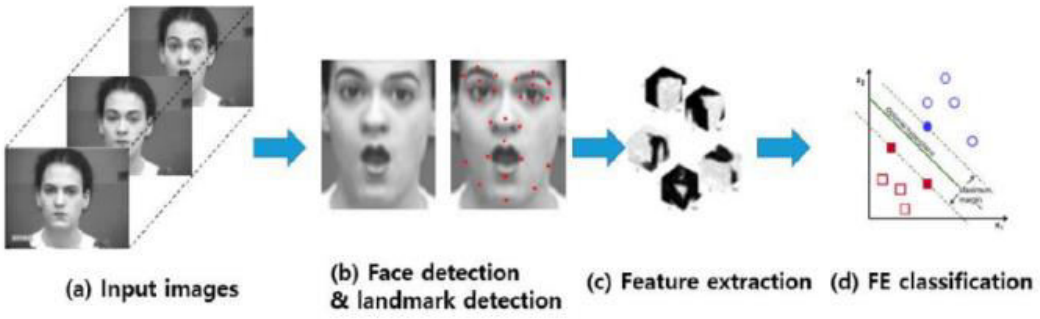


Fig: FER Steps to process the information

Deep learning based FER reduces the total dependency of external pre processing techniques by configuring end to end images. Landmarks in the faces are very crucial parts in facial recognition. With the help of Dlib library which uses a technique called Maximum Margin Object Detector (MMOD) with CNN based features , this is because of large amount of data has been trained. This library also has a unique functionality called Frontal Face Detector we can used to find the coordinates of the face as below figure.

Fig: Facial Landmarks

Once the contour has been detected with a particular label like Happy, Angry, Surprise, Sad etc., now the recommendation engines searches the back operating system files to play a music based on emotions detected with the help of an important library called it as **"Play Sound",** So these helps to detect the virtual emotions of a person to overcome the situational stress or sadness or any other situations and provides the music, which helps the person in handling the particular moment and gives an instant relief or the happiness based on the virtual emotions.

Software Requirements Specification

The proposed system has many open-source software environments to write a program and develop algorithm but we selected a certain software called PYTHON 3.7.8 which has ability to multi task the computer vision for getting required output. The total project has to dependent on camera quality but not on the algorithm, the proposed software has a Graphical User Interface (GUI) which is easy to use for the public the user can also follow the computer requirements to reach the expectations of the project else the proposed system may causes abruptly crashing the system. The below supporting libraries are mandatory to get the output of the project (dlib, cv2, One-hot Encoder, Math, numpy). The same procedure can be used in Raspberry pi to get hardware-based outputs.

System Testing with Experimental Results:

Before we see the results, there is some important things to understand how testing and troubleshooting takes place. Firstly the user requires some prior steps to start the project. In the begin of the step it process the images with pre-determined data set called **"Test "** and **"Train"** this helps for the cross validation in this each dataset is having 7 predetermined labels Happy, Angry, Surprise, Sad, fearful, disguised, neutral. With the help of test dataset and train dataset applied training for machine with help of CNN algorithm. In this it takes 458 epoch's one epoch takes time for completing 20 minutes for deep training of neurons. The next step is Test the Emotion in different types of facial features.

Fig: Emotion- Happy

Fig: Emotion- Angry

Fig: Emotion- Neutral

Final Predicted label:

The below table is predicted label

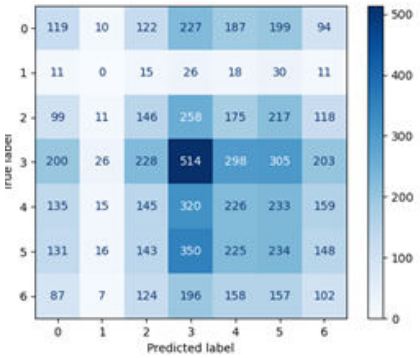


Fig: Final predicted Table

	precision	recall	f1-score	support
0	0.15	0.12	0.14	958
1	0.00	0.00	0.00	111
2	0.16	0.14	0.15	1024
3	0.27	0.29	0.28	1774
4	0.18	0.18	0.18	1233
5	0.17	0.19	0.18	1247
6	0.12	0.12	0.12	831
accuracy			0.19	7178
macro avg	0.15	0.15	0.15	7178
ghted avg	0.18	0.19	0.18	7178

Fig: Predictable Score for accuracy

From the above tables. The accuracy talks about how many times the machine learning model was correct in overall training and recall talks about how many times the model was able to detect a specific category.

CONCLUSION AND FUTURE SCOPE

In the proposed paper we reviewed and developed a CNN algorithm for predicting the output based on Facial Emotion Recognition (FER) with a high-level accuracy detection with predicted table. This might help the user to predict which emotion it is. In future the application can be interfaced for robot in real world environment for interacting with public.

REFERENCES

1. F. Ahmed, MSA. Hossain Bari, ML. Gavrilova, "Emotion Recognition from Body Movement", IEEE Access 8, doi 10.1109/access.2019.2963113, 2020, pp. 11761-11781.
2. YB. Ayzeren, M. Erbilek, E. Çelebi, "Emotional State Prediction from Online Handwriting and Signature Biometrics", IEEE Access 7, doi 10.1109/ACCESS.2019.2952313, 2019, pp. 164759- 164774.
3 MA. Nicolaou, S. Zafeiriou, I. Kotsia, G. Zhao, J. Cohn, "Editorial of Special Issue on Human Behaviour Analysis "In-the-Wild", IEEE Transactions on Affective Computing 10:4-6. doi 10.1109/TAFFC.2019.2895141, 2019.
4 J. Guo, Z. Lei, J. Wan, E. avots, N. hajarolasv, B. knyazev, A. et. al, "Dominant and Complementary Emotion Recognition from Still Images of Faces", IEEE Access 6, doi 10.1109/ACCESS.2018.2831927, 2018, pp. 26391-26403.
5 CA. Corneanu, M. Oliu, JF. Cohn, S. Escalera, "Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-related Applications", IEEE Transactions on Pattern Analysis and Machine Intelligence, doi 10.1109/TPAMI.2016.2515606, 2015.
6 W. Wei, Q. Jia, Y. Feng, G. Chen, M. Chu, "Multimodal facial expression feature based on deep-neural networks", Journal on Multimodal User Interfaces. DOI: https://doi.org/ 10.1007 / s12193-019-00308-9, 2020, pp. 17-23.
7 MS. Hossain, G. Muhammad, "Emotion Recognition Using Deep Learning Approach from Audio-Visual Emotional Big Data", Information Fusion, DOI: https://doi.org/1 0.1016/ j.inffus.2018.09.008, 2018.
8 SA. Rahman, FA. AlOtaibi, WA. AlShehri, "Sentiment Analysis of Twitter Data", 2019 International Conference on Computer and Information Sciences.

SENTIMENT ANALYSIS ON TEXT FOR PRODUCT AND SERVICE EVALUATION AND ITS FUTURE PERSPECTIVES

Pawan Kumar
MCA Student, Amity University, Patna

ABSTRACT
This paper explores the use of sentiment analysis for product and service evaluation and its future scope. The paper provides an overview of the differentapproaches and techniques used in sentiment analysis and the challenges and limitations of each approach. This paper also discusses how sentiment analysis can be used to extract and analyze customer reviews, surveys, and social media posts automatically. Lastly, the paper examines the prospects for sentiment analysis for product and service evaluation and its potential growth and expansion. It emphasizes the importance of sentiment analysis in improving customer satisfaction and enhancing brand reputation.

INTRODUCTION
Businesses face the challenge of comprehending and evaluating customer sentiments regarding their products and services in the digital age, where information flows abundantly across numerous platforms. Influencing business strategies, enhancing customer satisfaction, and enhancing brand reputation all depend on the opinions and feedback of customers. Consequently, sentiment analysis has become a useful tool for automatically extracting, analyzing, and interpreting sentiments from social media posts, surveys, and customer reviews.

The goal of this research paper is to provide an overview of the various methodsand approaches used in the field of sentiment analysis for product and service evaluation. The purpose of this paper is to identify potential areas for advancement and improvement by examining the difficulties and limitations associated with each approach.

The paper starts by presenting the idea of feeling investigation, otherwise called assessment mining, which includes the extraction and examination of opi nions, feelings, and emotional data from printed information. It emphasizes the significance of sentiment analysis in gaining an understanding of customer perceptions, preferences, and opinions, which ultimately enables businesses tomake decisions based on accurate information.

The paper then investigates a variety of sentiment analysis methods, including lexicon-based approaches, natural language processing (NLP), and machine learning algorithms. The advantages, disadvantages, and applicability of each method to product and service evaluation are discussed.

The difficulties of sentiment analysis, such as the presence of sarcasm, ambiguity, and context dependence in textual data, are also examined in this paper. Moreover, it resolves the issue of area transformation, where feeling investigation models prepared on one space may not perform well on anotherarea.

The paper shows how sentiment analysis can be used to automatically extract and analyze customer reviews, surveys, and social media posts to show h ow it can be used in practice. By utilizing opinion examination procedures, organizations can acquire significant bits of knowledge into consumer loyalty, distinguish areas of progress, and answer immediately to client criticism.

The paper also looks into the potential applications of sentiment analysis for product and service evaluation in the future. It discusses the integration of cutting-edge AI technologies like deep learning and multimodal data analysis, as well as the potential growth and expansion of sentiment analysis techniques.

In general, the significance of sentiment analysis in enhancing customersatisfaction and brand reputation is emphasized in this research paper.

Businesses can meet customer expectations and remain competitive in a dynami c market by using sentiment analysis to gain actionable insights and make data - driven decisions.

LITERATURE REVIEW

The first to use machine learning models to classify sentiment was Pang, Lee, and Vaithyanathan. For sentiment analysis on unigrams and bigrams of data, they used the Naive Bayes, Max Entropy, and Support Vector Machine models [6]. The best results were obtained in their experiment when SVM was combinedwith unigrams. Using data from numerous sources, Mullen and Collier used SVM to perform sentiment classification [7]. Their work showed that utilizing mixture SVM with highlights in view of Osgood's hypothesis [ref.] the best results were achieved. This method was effective, but it did not give enough weight to more contextual classifications, and the overall result was greatly diminished because of this domain variability. Their proposedmethod had an accuracy rate of 86.6 percent, which needs to be greatly improved.

Zhang developed a computational model to investigate the properties of reviews linguistics [8] in order to evaluate its usefulness. Support Vector Machine (SVM) calculation was utilized for order. Zhang presumed that the nature of survey is great assuming it containsboth emotional and objective data.

However, the use of a fuzzy search technique for opinion mining caused the analysis to be only 72% effective, leading to a significant issue whenever a word was misspelled. Efthymios et al. under-went opinion investigation on Twitter messages involving different highlights for characterizations N-gram include, vocabulary include, POS highlight. Their work was basically subject explicit andaccomplished an exactness of almost 80% and furthermore presumed that POS highlight reduces precision level [10]. Negation handling techniques were used in sentiment analysis by Farooq, et al. [9].

In their experiment, they looked at how syntactic and diminishing negation words affected one another. They accomplished a normal precision pace of 83.3%.

In order to determine which words or phrases are positive or negative in a general context, sentiment analysis approaches frequently require resources such as sentiment lexicons. A manually compiled resource, General Inquirer [11] is frequently utilized in sentiment analysis.

Hatzivassiloglou and McKeown [12] were the first researchers to use machine learning to create a lexicon of sentiment terms, and their work laid the groundwork for automatic acquisition of the polarity of sentiment words and phrases. The polarity of sentiment words can be learned using a variety of methods. Riloff, Wiebe, and Wilson [15] focused on nouns, while Riloff and Wiebe [16] extracted linguistic patterns from subjective expressions. Wiebe [13] and Turney [14] studied the extraction of adjectives and adjectival phrases.

Riloff, Wiebe, and Wilson [15] focused on nouns. Qiu and others 17] proposed a propagation method for assigning polarity to a large number of sentiment words.

METHODOLOGY

1) Methods based on dictionaries

In this approach, a seed list of words with known prior polarity is generated. The seed list is then expanded by iteratively extracting synonyms or

antonyms from WordNet 1 and other online dictionary sources. Two seed lists containing both positive and negative verbs and adject ives were manually created by Kim and Hovy [31]. The synonyms and antonyms of thewords on the seed list were then extracted from WordNet and placed on the appropriate list, with synonyms on the same list and antonyms on the opposite.

How the new, unseen words interacted with the seed list determined the words' sentimental strength.

Both positive and negative opinion qualities was figured for each word and their general extents was thought about. Kamps et al. [Based on WordNet lexical relation] 32]estimated the semantic direction of words. They gathered words and every one of their equivalents in WordNet, for example expressions of a similar synset. After that, an edge - connected graph of synonymous word pairs was created. The word's relative distance from the good and bad seed terms was used to determine its semantic orientation. The distance was the length of a most brief way between two words wi and wj.

The absolute value indicates the strength of the orientation. The polarity classification is not domain-specific, which is a disadvantage of the dictionary method. For instance, "unpredictable" is a good word to describe a movie plotbut a bad word to describe a car's steering [25].

D. Feature-Based Sentiment Analysis

In a review, the author discusses a product's benefits and drawbacks. Even though the overall opinion of the product may be positive or negative, the reviewer may like some features and dislike others. Sentiment classification at the sentence or document level does not provide this kind of information.

Consequently, highlight based assessment feeling examination [22, 23, 24] is required. This includes separating item highlight and the relating assessment on it. Intuitively, one could believe that item includes are communicated by things and thing phrases, yet not all things and thing phrases are item includes. By removing only base noun phrases, definite base noun phrases (noun phrases preceded by the definite article "the"), and beginning definite base noun phrases (definite base noun phrase at the beginning of a sentence followed by a verb phrase), Yi et al.[29] further restricted the candidate words.

A sentiment pattern database is used to determine the target and final polarit y of each sentiment phrase that is detected.

FUTURE PERSPECTIVES AND CHALLENGES

Feeling examination, otherwise called assessment mining, is an important devicein understanding and dissecting individuals' perspectives, feelings, and suppositions in printed information. Even though sentiment analysis has come a long way in recent years, there are still issues and opportunities for the futurethat need to be addressed. A few examples:

Understanding the Situation: Systems for sentiment analysis need to learn more about language nuances and context. Depending on the surrounding text, cultural references, or specific domains, the same words may have distinct meanings. In order to accurately interpret sentiment, future developments ought to concentrate on incorporating contextual understanding.

Managing Irony and Sarcasm: Models for sentiment analysis are challenged by sarcasm, irony, and other figurative language. A deeper comprehension of the speaker's intention and the social context underlying these expressions is frequently required. It will be crucial to develop models that can effectivelydetect and interpret these nuances.

Adaptation to the Domain Opinion investigation models prepared on one space may not perform well when applied to another area. Adjusting opinion examination to various ventures, like medical services, money, or online entertainment, requires specific models or move learning strategies to accomplish exact outcomes. Future endeavors ought to zero in on building space explicit feeling examination models or techniques for viable area transformation.

Multilingual and Multifaceted Investigation: Although sentiment analysis is primarily developed for English text, its global applicability necessitates its expansion to other languages. Due to linguistic variations, idiomatic expressions, and cultural biases, analyzing sentiment across various languages and cultures presents additional challenges. Future points of view inclu de creating multilingual feeling examination models and representing social contrasts in opinion understanding.

Data Bias Management: Large datasets are used to train sentiment analysis models, which can introduce data biases. Data collection methods, ann otation processes, and social and cultural biases embedded in the training data are allexamples of sources of bias. Guaranteeing reasonableness and moderating predispositions in feeling examination models will be vital for building more solid and fair-minded frameworks.

Sense of Emotion: Feeling examination normally centers around good, pessimistic, or impartial opinion, yet feelings are more assorted and complex.Emotion detection could be incorporated into sentiment analysis in the future, allowing for a deeper comprehension of people's emotional states. This can include perceiving feelings like joy, misery, outrage, dread, or shock notwithstanding opinion extremity.

Analyses in real time: Real-time sentiment analysis becomes essential as the volume and velocity of data generated on social media platforms continue to rise. Creating productive models and methods that can deal with the high velocity information stream and give ongoing feeling investigation will be basicfor ideal experiences and navigation.

Ethical and Privacy Concerns: Feeling examination includes dissecting clients' perspectives and opinions from their literary information, which raises protection concerns. Future advancements ought to zero in on tending to these worries and guaranteeing that feeling examination frameworks comply with moral rules and regard client protection privileges.

Generally, the fate of opinion examination lies in working on relevant grasping, representing social and phonetic variety, tending to predispositions, consolidating feeling identification, and growing continuous and security cognizant arrangements. Defeating these difficulties will prompt more exact, solid, and adaptable opinion examination frameworks.

REFERENCES

[1]. Georg Lackermair, Daniel Kailer, KananKanmaz, “Importance of Online Product Reviews from a Consumer’s Perspective”, pp. 1-5, 2013.

[2]. Mrs. Manisha Pravin Mali, Dr. Mohammad Atique, "Applications of Text Classification using Text Mining", pp. 1-4, 2014.

[3]. W.A. Awad, S.M. ELseuofi, "Machine Learning Methods for Spam Email Classification", pp.1-12, 2011.

[4]. BijoyanDas, Sarit Chakraborty "An Improved Text Sentiment Classification Model Using TF-IDF and Next Word Negation", pp-1-6

[5]. Kaggle Open Data Source, Adam Mathias Bittlingmayer, "Amazon Reviewsfor Sentiment Analysis", "www.kaggle.com/bittlingmayer/amazonreviews".

[6]. B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques," In Proc. of Conf. onEmpirical Methods in Natural Language Processing, pp 79- 86, 2002.

[7]. T. Mullen and N. Collier, "Sentiment analysis using support vector machines with diverse information sources,' In Proc. of Conf. on Empirical Method s in Natural Language Processing," pp 412–418, 2004.

[8]. Michael Weigand, Alexandra Balahur, Benjamin Roth, Dietrich Klakow, Andres Montoyo, "A Survey on Role of Negation in Sentiment Analysis", pp 1-9.

[9]. Umar Farooq, Hasan Mansoor, Antoine Nongaillard, YacineOuzrout, Muhammad Abdul Qadir, "Negation Handling in Sentiment Analysis at SentenceLevel", pp 1-9.

[10]. Z. Zhang, "Weighing Stars: Aggregating Online Product Reviews for Intelligent E-commerce Applications," In Intelligent Systems, IEEE, vol. 23,issue no.05, pp 42-49, 2008.

[11]. P.J. Stone, D.C. Dunphy, M.S. Smith and D.M. Ogilvie, General Inquirer: a Computer Approach to Content Analysis (The MIT Press, Cambridge, MA, 1966).

[12]. V.Hatzivassiloglou and K.R. McKeown, Predicting the semantic orien tationof adjectives, Proceedings of the 35th Annual Meeting of the ACL and the 8th Conference of the European Chapter of the ACL (1997) 174–181.

[13]. J. Wiebe, Learning subjective adjectives from corpora, Proceedings of the17th National Conference on Artificial Intelligence (2000) 735–740.

[14]. P.D. Turney, Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews, Proceedings of the 40th Annual Meetingof the Association for Computational Linguistics ACL (2002) 417 –434. [15].E.Riloff, J. Wiebe and T. Wilson, Learning subjective nouns using extraction pattern bootstrapping, Proceeding of the 7th Conference on Natural Language Learning (2003) 25–32.

[16]. E.Riloff and J. Wiebe, Learning extraction patterns for subjectiv e expressions, Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (2003) 105–112.

[17]. G. Qiu, B. Liu, J. Bu and C. Chen, Expanding domain sentiment lexicon through double propagation, Proceedings of the 21st Internatio nal Joint Conference on Artificial Intelligence (Morgan Kaufmann, San Francisco, 2009)1199–1204.

ADVANCES, APPROACHES AND DIMENSIONS OF WATER PURIFICATION: A COMPARATIVE REVIEW

Pinky Kumari[1] and Ishita Ghosh[2]
[1]Research Scholar and [2]Associate Professor, Netaji Subhas University, Pokhari, Jamshedpur

ABSTRACT
The necessity of water for mankind is universally known. Water is abundantly available on Earth with around three-fourths of it being covered with water. However, the thoughtful concern is due to the ever increasing water pollution and dwindling fresh water reserves. The pollutants and contaminants are surging high in volume and variety due to which fresh drinking water is now up to a threatened level. As such the study on the dimensions and diversities of approaches for procuring pure water is of prime importance. As the problem increase, it is a call to the research community to come up with easy and affordable solutions for the mankind to continue the flow of fresh and clean drinking water.

This paper summarizes the current prevailing situation in context to the various dimensions of water purification. It also puts forward a comparative study of the various purification methods and how they have evolved with ages. Furthermore as the variety and toxicity of the impurities is on the increase, it is obvious that the purification methods also need to upgraded to combat the threats. This paper additionally introduces the plasma technology for water purification and states its advantages over other common and traditional methods in practise till date.

Keywords: Water purification, Pollutant, Plasma technology

1. INTRODUCTION
We all are really blessed to be a part of the Earth Planet! It is well known that of the World's total water supply of about 332.5 million m^3 of water, over 96% is saline. Water (chemically, H_2O) is a natural resource of earth and though about 70% of earth's surface is covered with water, only 3% of it is consumed for drinking purpose. It must be the first priority to all human beings to conserve water for ourselves as well as the future generation.

According to WHO, nearly a billion people lack access to clean drinking water and estimated 500 million die each year from diseases associated with contaminated sources. Freshwater is a scarce commodity making up only 2.5 % of total water present on earth. Recently, the use of photo catalysis for the treatment of a variety of pollutant such as dyes, pharmaceuticals, and various endocrine disrupting compounds are rapidly increasing. The advantages of activation oxidation process is that it can completely mineralise recalcitrant pollutant into simple compound that are begin or can be processed by natural mechanism to harmless constituents. Increasing global demands for potable water supply and waste water treatment have driven extensive research, development and application of water treatment technique. Currently, water treatment membranes are commonly composed of microfiltration, ultrafiltration, nano-filtration, reverse osmosis and forward osmosis, based on different pore sizes and operational pressure. Membrane based filtration process, is one of the most active separation and purification technologies. The presence of organic pollutant has given rise to serious concern for aquatic life and public health. There are more than 10000 different pigments and dyes used by industries and more than 0.7million tons of dyes are synthesized annually worldwide.

One of the innovative and environmentally safe methods for the preparation of nano-sized compounds is the use of plasma discharge of various configuration: plasma discharge generated between the electrodes immersed in a liquid at gas liquid phase interface at reduce pressure, plasma at atmospheric pressure in the interaction with the liquid. Among plasma chemical discharge, contact no equilibrium low temperature plasma is the most the promising application.

Plasma Technology helps in the improvement of efficiency of disinfection and sterilization method against various micro-organism. Plasma in contact with liquid generates the host of reactive species that attack and ultimately mineralize contaminant in a solution.

The different types of water resources are: Salt water resources, Ground water resources, Surfaces water resources and Fresh water resources.

Some **Traditional methods** to purify water are as follows:

- **Boiling:** At high temperature, it can kill microorganism and pathogens and results in 100% pure water after three minutes of boiling
- **Distillation:** Converts raw water by boiling into steam and after few seconds, gets cooled down and purified water can be achieved.
- **Chemical Disinfection:** Achieved by adding disinfectant like chlorine, chlorine dioxide ozone, copper, silver ionization or bromine into water so as to kill microorganism
- **Filtration:** also known as Universal method of purification, helps to remove contaminant or to separate suspended particles and bacteria, algae, viruses from water and also reduces bacterial content by 98-99%, turbidity by 50 PPM to 5 PPM and colour to colourless
- **Ultraviolet Light:** Wavelength of UV ranges in between 200-400nm and kills bacteria and viruses by destroying molecular bond that hold their DNA together.

2. WATER RESOURCES IN INDIA

Water resources in India include information on precipitation, surface and groundwater storage and hydropower potential. India experiences an average precipitation of 1,170 millimeters (46 in) per year, or about 4,000 cubic kilometers (960 cu mi) of rains annually or about 1,720 cubic meters (61,000 cu ft) of fresh water per person every year.

India accounts for 18% of the world's population and about 4% of the world's water resources. One of the solutions to solve the country's water woes is to create Indian Rivers Inter-link. Some 80 percent of its area experiences rains of 750 millimeters (30 in) or more a year. However, this rain is not uniform in time or geography. Most of the rains occur during its monsoon seasons (June to September), with the northeast and north receiving far more rains than India's west and south. Other than rains, the melting of snow over the Himalayas after the winter season feeds the northern rivers to varying degrees. India harnessed 761 cubic kilometers (183 cu mi) (20 percent) of its water resources in 2010, part of which came from unsustainable use of groundwater

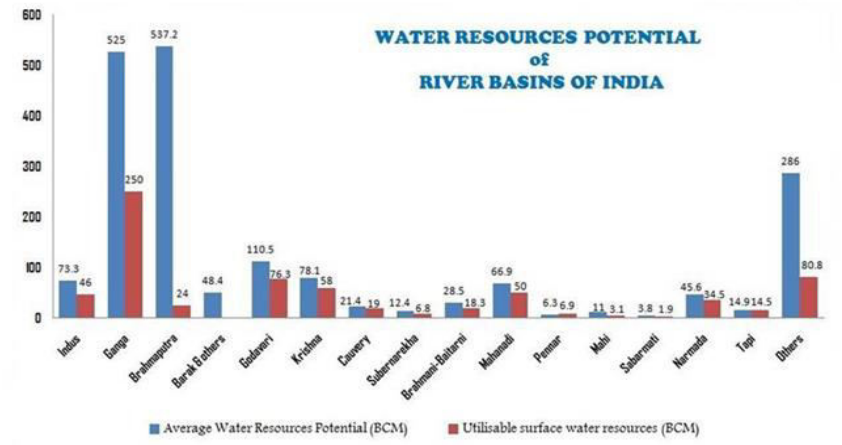


Figure 1: Comparison of water potential and its utilization of Indian Rivers

As we know, the natural resources of water include rain water as well as the rivers which are well spread across the territory of India. The figure shows the contribution of few main rivers of India in terms of resource potential but interestingly, the entire water content is non-utilizable and we can see that only a fraction of it can be utilized for our domestic purpose.

3. WATER USAGE

Irrigation by far is the largest user of India's water reserve with hooping usage of 78% of total water reserve, followed by domestic sector (6%) and industrial sector (5%)(PIB 2013).

National Commission on Integrated Water Resources Development (NCIWRD) the irrigation sector alone is going to need additional 71 bcm by 2025 and 250 bcm of water by 2050 compared to the demands of 2010 (Press Information Bureau 2013).

Ground Water is also a major source of drinking water in urban and rural India. 45% of total irrigation and 80% of domestic water come from ground water reserve.

States like DL, PN, HR, UP over exploitation of ground water has led to water scarcity. States like RJ, GJ arid climate leads to water stressed condition, while in TN, KA, AP poor aquifer properties are responsible for water scarcity. Other reasons being increasing population pressures, industrial growth and unprecedented pace of urbanization.

4. HOW SAFE IS OUR WATER?

About 70% of surface water resources in India are polluted. The major contributing factor for water pollution are wastewater from different sources, intensive agriculture, industrial production, infrastructure development and untreated urban runoff. Everyday 2.9 billion liters of waste water from industrial and domestic sources are dumped into the river Ganga without treatment.

According to WHO, Half of India's morbidity is water related. Waste management has not been as efficient as required to manage increasing volume of waste generated daily in India, especially in cities. Municipal wastewater treatment capacity developed so far in India accounts for only 29% waste generated in urban habitations having population more than 50,000 and the gap is projected to increase. Domestic effluents contribute a substantial proportion of water pollution in India. More than 70% of domestic untreated effluents are disposed-off to environment.

5. GLOBAL WATER RESOURCES

Key findings of the Global water resources report:

The report underlined that large parts of the world were drier than normal in 2021 with cascading effects on economies, ecosystems, and our daily lives.

The year saw precipitation patterns largely influenced by climate change and the La Nina event. The general observations in the report are:

The area with below-average stream flow was approximately two times larger than the above-average area, in comparison to the 30-year hydrological average. Annual glacier runoff increased initially due to the melting of glaciers until the turning point called peak water is reached after which the runoff declines. 74% of all-natural disasters between 2001 and 2018 were water-related. 6 billion people face inadequate access to water at least a month per year which is expected to reach 5 billion by 2050. The continued melting of glaciers shows a clear trend toward an acceleration of mass loss on multi decadal timescales. It highlights the lack of accessible verified hydrological data.

6. RURAL AND URBAN SOURCES OF POLLUTION

According to UNESCO 2021 World Water Development Report, about 829,000 people die each year from diarrhea caused by unsafe drinking water, sanitation, and hand hygiene, including nearly 300,000 children under the age of five, representing 5.3 % of all deaths in this age group. Water takes large amount of pollution from different sources, including, industrial discharge, mobile sources (cars/trucks), residential wastewater, Trash and polluted storm water runoff urban landscape. Urbanization heavy use if detergent, pesticides and fertilizer, and deforestation are all typical source of water pollution. In rural areas, include run-off from agriculture land containing substance including pest control product, animal medicine, sewage sludge and manure, incorrect waste pipe connection, Run-off and leaching from contaminated land.

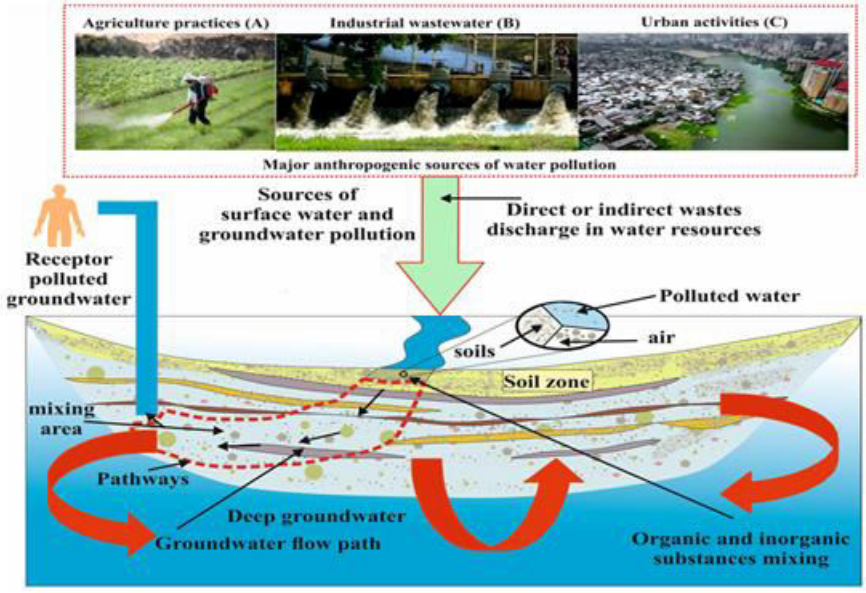


Figure-2 Sources of ground water and surface water pollution

The flowchart of sources of water pollution has been discussed below:

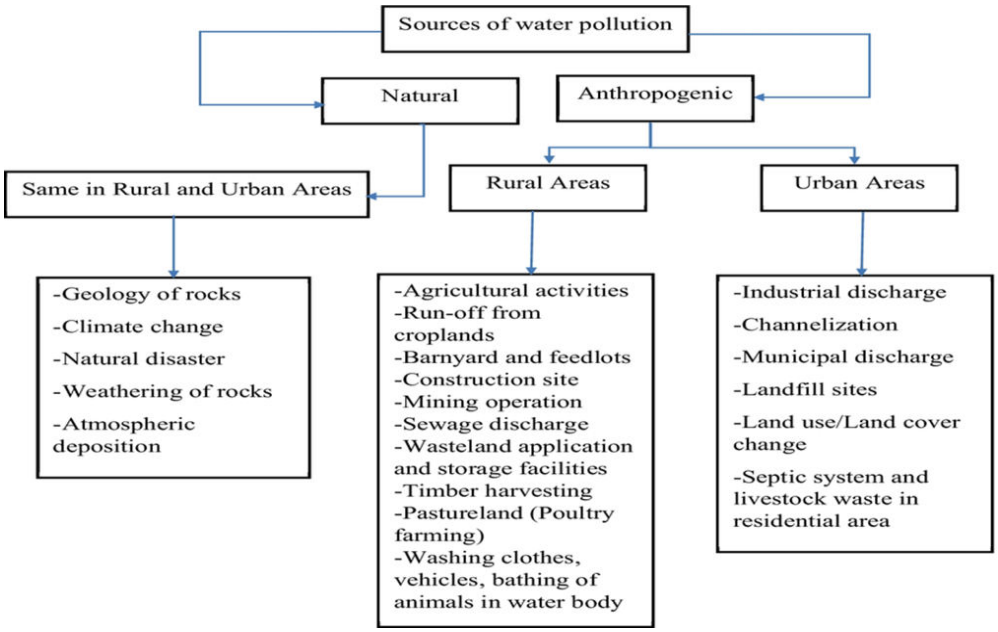


Figure 3: Classification of rural and urban pollutants

7. METHODS OF PURIFICATION OF WATER IN INDIA

The following are the methods of purification of water:

1. **Disinfection**: It's a process of eliminating many or all pathogenic microorganism except bacterial spores on inanimate objects.
2. **Catalysis**: In catalyzed water, minerals particles repel each other & lose their ability of adhesion.
3. **Community-scale atmospheric water harvesting:** there is a tremendous opportunity in atmospheric water harvesting to help improve access to clean drinking water.

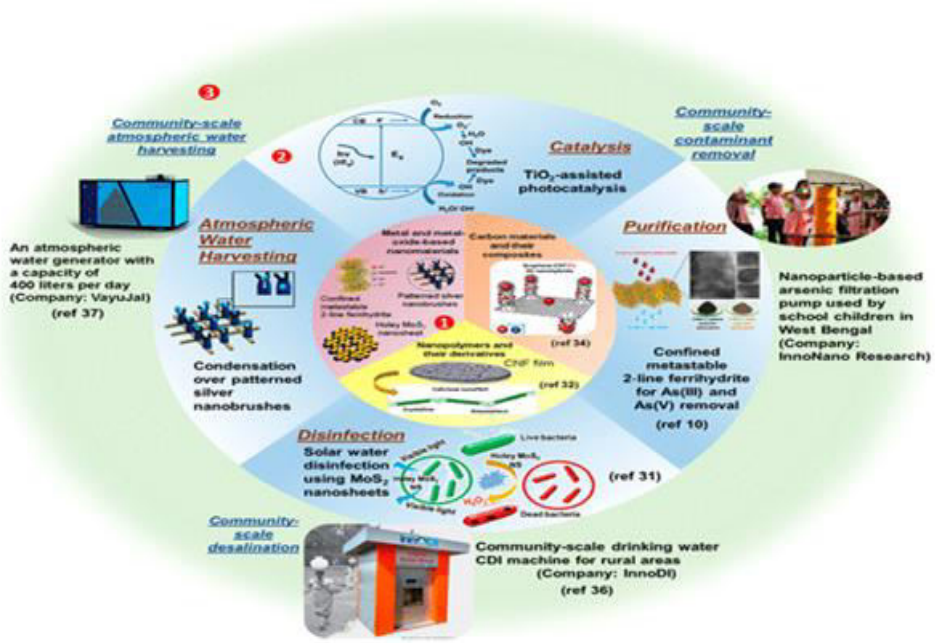


Figure 4: Methods of purification of water

8. MODERN METHODS OF PURIFICATION

- **The Use of Nanotechnology:** This technology uses titanium dioxide nanotechnology. This process eliminates bacteria and other toxins in water. It also helps break down unrefined compounds with the help of ultra violet rays. The nanotechnology method does not however use the polymer-based water treatment membrane. This method is very affordable and easy to apply. It is also environmentally friendly because it helps reduce the buildup of microorganisms known to grow rapidly on drenched surfaces.

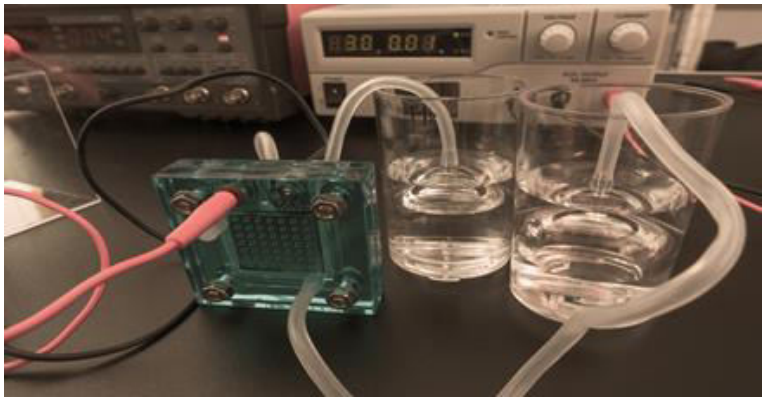


Figure 5: Use of nanotechnology for water purification

- **The RO Purification:** RO purification which is commonly known as reverse osmosis is one of the most used method of water treatment. This process involves the use of membrane technology which allows it remove dissolved salts and other impurities in water. This membrane has extremely fine pores which allow only water to pass through. The water leaves behind all the poisonous substances in the water.

Figure 6: Use of RO for water purification

- **UV Purification:** This is also known as the e-boiling method. This water purification method uses ultra-violet light to help kill bacteria and other harmful substances in the water. The purifier contains a minute mercury lamp which manufactures diminutive wave UV radiations. The radiations function by irradiating the water and piercing through the cells of the microorganisms and viruses. This in turn destroys their capability to reproduce. This method however requires other filtration processes because the dead germs remain in the water until a separate filter is introduced to help remove the dead germs physically.

Figure 7: Use of UV for water purification

- **Acoustic Nanotube Technology:** This technology was developed by NASA's Johnson Space Centre to help in water purification. The Acoustic Nanotube Technology gets rid of the contaminants in the water by using a sieve which is normally surrounded by tiny diameter nanotubes. They help push the water away from the contaminants hence allowing you collect purified water separately.

Figure 8: Use of Acoustic Nanotube technology for water purification

- **SunSpring System:** This is a water purification system that helps distil up to 5,000 gallons of drinking water in a single day. It uses a battery that solely runs on renewable energy. It is environmentally friendly and also a cost efficient method.

Figure 9: Use of Sunspring system for water purification

9. PLASMA TECHNOLOGY WATER PURIFICATION

Plasma technology water purification is a new water treatment technology developed according to the trend of industrial water use in the 21st century. It is effective, efficient, scalable, versatile and customizable. These technologies must be able to adapt to new contaminants, reduce energy consumption, maintain or improve the proportionality between power and flow, demonstrate various flow capacities, minimize the transformation of existing infrastructure, prepare for imminent regulations, and tailor chemistry to site-specific requirements.

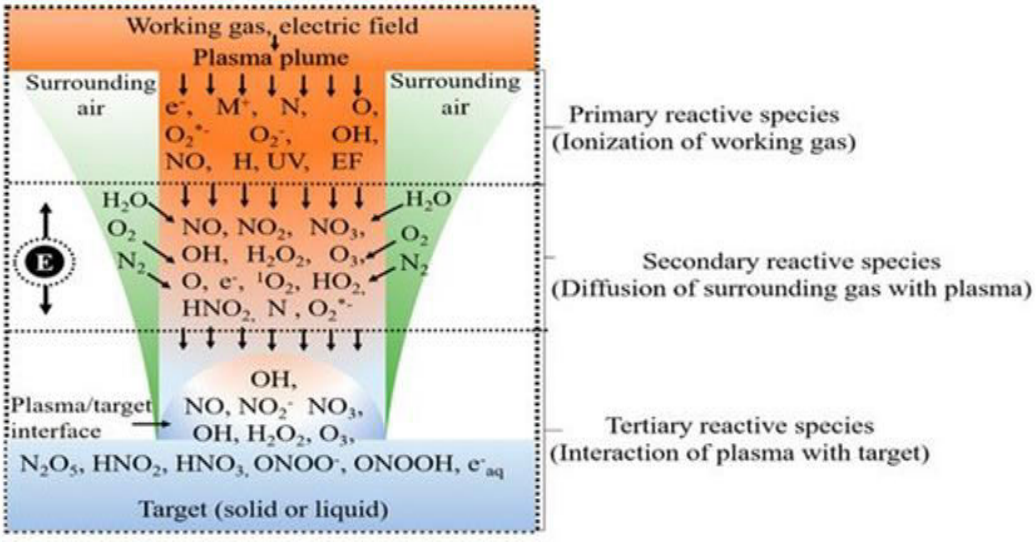


Figure 10: Use of plasma technology for water purification

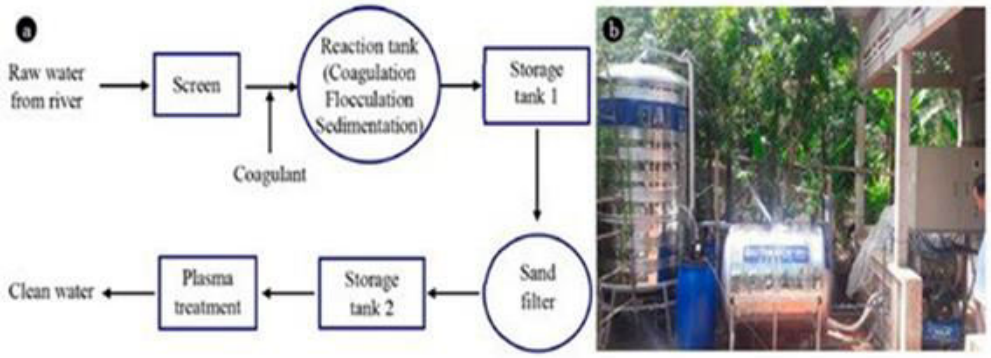


Figure 11: Plasma treatment for water purification

New methods of water treatment by plasma must have all the above-mentioned properties and pose the least risk to public health. NTAPPs and their chemical reactions release energy and reactive chemical species that can kill bacteria and microorganisms, resulting in the disinfection of water. The advantage of this technique is that it can be performed in ambient air under atmospheric pressure without a vacuum system.

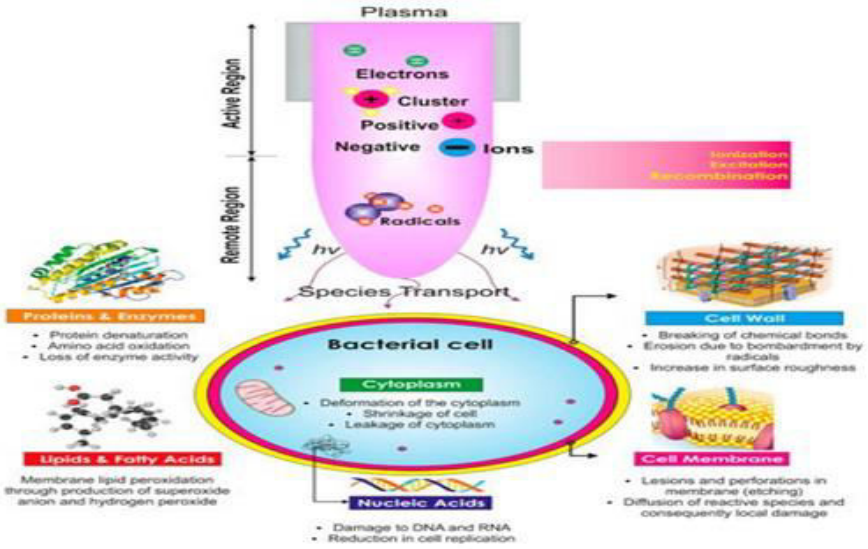


Figure 12: Use of plasma technique for bacterial treatment of water

The existence of organic pollutants in water has caused substantial concern owing to the adverse effects on the environment and human beings. The elimination of pharmaceutical components in water using NTAPP has gained significant attention due to the occurrence of these contaminants in surface water and occasionally even in drinking water. Industrial pollutants in contaminated groundwater generally release volatile organic complexes, such as m-xylene and toluene, into the neighbouring regions. Several pharmaceuticals have been observed in above-ground water and groundwater, including streams. Figure 13 depicts the application of NTAPP for pharmaceutical component degradation in water. AntibIOTics, as one class of pharmaceuticals, are widely studied due to the development of bacteria with antibIOTic resistance. Their oxidative elimination by plasma is quite beneficial; nonetheless, mineralization has been confirmed to be moderately slow. Based on the determined oxidation intermediates, researchers have suggested that the main degradation mechanism relies on OH radical attacks, subsequent hydroxylation, and the damaging of molecular bonds, which leads to mineralization. Reports on the application of NTAPP to remove antibIOTics such as atenolol, verapamil, and enalapril have been published.

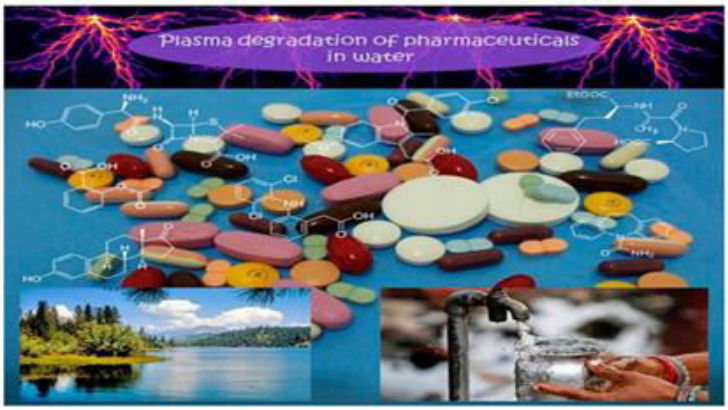


Figure 12: Use of plasma technique for pharmaceutical treatment of water

10.CONCLUSION

Our planet Earth has provided us immense reasons to enjoy and sustain a comfortable life. It is now the duty of its inhabitants to nurture its resources with care and creation. The intellectuals need to gear up on war footing and put forward newer and effective solutions to the problems which keep brimming up.

Pure drinking water is one of the prime necessities of life and it must effectively reach all the sections of people.

Our study proposes to make important contributions in the arena of water purification and suggesting effective models to sustain the goal universally utilizing plasma techniques for analysis of the polluted water and will experiment on the variation of input parameters for desired output.

We hope to gratify the needs and necessities of the mankind by operational and optimal solutions and use our expertise for satisfactory implementations.

REFERENCES

[1] Yanzong Zhang*, Jingtang Zheng, Xianfeng Qu, Honggang Chen, Design of a non-equilibrium plasma-based water treatment reactor, Chemosphere 70, Elsevier (2008) 1518-1524, doi: 10.1016/j.chemosphere.2007.09.013.

[2] Muhammad Arif Malik, Water Purification by Plasma: Which Reactor are Most Efficient?, Plasma Chem Plasma Process(2010) 30:21-31, doi: 10.1007/ s11090. 009.9202.2.

[3] V.I. Grinevich. E. Y. Kvitkov. N. A. Plastinina. V. V. Rybkin, Application of Dielectric Barrier Discharge for Waste Water Purification, Plasma Chem Plasma Process (2011) 31:573-583, doi: 10.1007/s11090.010.9256.1.

[4] John Foster, Bradley S. Sommers Nowak Gucker Member IEEE, Isaiah M. Blankson and Grigory Adamovsky, Perspectives on the Interaction of Plasma with Liquid Water for Water Purification, IEEE Transaction on Plasma Science, Volume 40, NO. 5, MAY 2012, doi: 10.1109/TPS.2011.2180028.

[5] Bo Jiang, Jingtang Zheng*, Shi Qiu, Mingbo Wu*, Qinhui Zhang*, Zifeng Yan, Qingzhong Xue, Review on electrical discharge plasma technology for waste water remediation, Chemical Engineering Journal 236, Elsevier (2014) 348-368, http:// dx.doi.org/10.1016/jcej2013.09.090.

[6] Rasel Das, Md. Ali*, Sharifa Bee Abd Hamid, Seeram Ramakrishna, Zaira Zaman Chowdhury, Carbon nanotube membranes for water purification, Desalination 336, Elsevier (2014) 97-109, http://dx.doi.org/10.1016/j.desal2013.12.026.

[7] Gunnar R. Stratton, Christopher L. Bellona, Fei Dai, Thomas M. Holsen, Selma Mededovic Thagard, Plasma based water treatment: Conception and application of a new general principle for reactor design, Chemical Engineering Journal 273, Elsevier (2015) 543-550, http://dx.doi.org/10.1016/j.cej.2015.03.059.

[8] Ghazaleh Haghighat, Amirreza Sohrabi, *Parmiss Mojir Shaibani, C. W. Van Neste, Selvaraj Naicker and Thomas Thundat, The role of chloride ions in plasma – activated water treatment processes, Royal Society of Chemistry 2016, doi: 10.1039/c6ew00308g.

[9] John E. Foster, Plasma – based water purification: Challenges and prospects for the future, Physics of Plasma 24,055501 (2017), doi: 10.1063/1.4977921.

[10] B. Shrikanth, R. Goutham, R. Badri Narayan, A. Ramprasath, K.P. Gopinath, A.R. Sankaranarayanan, Recent advancements in supporting materials for immobilized photocatalytic application in waste water treatment, Journal of Environmental Management 200, Elsevier (2017) 60-78, http://dx.doi.org/10.1016/ j.jenvman. 2017.05.063.

[11] Junyong Zhu, Jingwei Hou, Yatao Zhang, Miaomiao Tian, Tao He, Jindun Liu, Vicki Chen, Polymeric antimicrobial membranes enabled by nanomaterials for water treatment, Journal of Membrane Science550, Elsevier(2018)173-197, http://doi.org/10.1016/ j.memsci.2017.12.071.

[12] Pankaj Attri, Fumiyoshi Tochikubo, Ji Hoon Park, Eun Ha Choi, Kazunori Koga and Masaharu Shiratani, Impact of Gammma rays and DBD plasma treatments on wastewater treatment, Scientific Report, (2018) 8:2926, doi: 10.1038/s41598-018-21001-z.

[13] N. B. Singh, Rachna, Water Purification by using Adsorbent, Environmental Technology & Innovation Volume 11, August 2018, Pages 187-240, http://doi.org/10. 1016/j.eti.2018.05.006.

[14] Akikazu Sakudo, Yoshihito Yagyu and Takashi Onodera, Disinfection and Sterilization Using Plasma Technology, International Journal of Molecular Sciences 2019, 20, 5216; doi: 10.3390/ijms20205216.

[15] Chandan Kumar, Ankit Raj Sinha, Sanjay Prakash, Khushwant Singh, Saurabh Kumar, Water Purification by Activation Plasma Technology, International Journal of Innovative Science and Modern Engineering (IJISME), ISSN: 2319-6386 (Online), Volume -6 Issue-8, July2020, doi:10.35940/ijisme.H1251.076820.

EXPLORING THE POWER OF OPEN SOURCE INTELLIGENCE (OSINT): TECHNIQUES, TOOLS, AND APPLICATIONS

Ramesh Kumar Sharma[1], Dr. Dharmendra Kumar Singh[2], Abhishek Kumar[3] and A. P. Burnwal[4]

[1]Research Scholar, Department of CSE, BIT Sindri, Dhanbad
[2]Director, BIT Sindri, Dhanbad
[3]Research Scholar, Department of ECE, NIT Jamshedpur
[4]Department of Math, GGSESTC, Bokaro

ABSTRACT

Open-Source Intelligence (OSINT) has become an increasingly important tool for gathering and analysing open-source information in a variety of domains, from business and marketing to national security and law enforcement. This paper explores the power of OSINT by examining its various techniques, tools, and applications. We begin by providing an overview of OSINT methodologies, including web scraping, data mining, and social media analysis and explore specific tools and platforms used in OSINT analysis, such as Maltego, Hunchly, and Social-Searcher. Next, we discuss the legal and ethical considerations of OSINT gathering and analysis, and offer suggestions for best practices in this field. Some existed case studies of application of osint are highlighted. This paper deals with a review on current trends and future directions in OSINT, such as the impact of artificial intelligence and machine learning on OSINT methodologies. Overall, this paper aims to provide a comprehensive overview of OSINT, and to demonstrate the value of this powerful tool for open-source information.

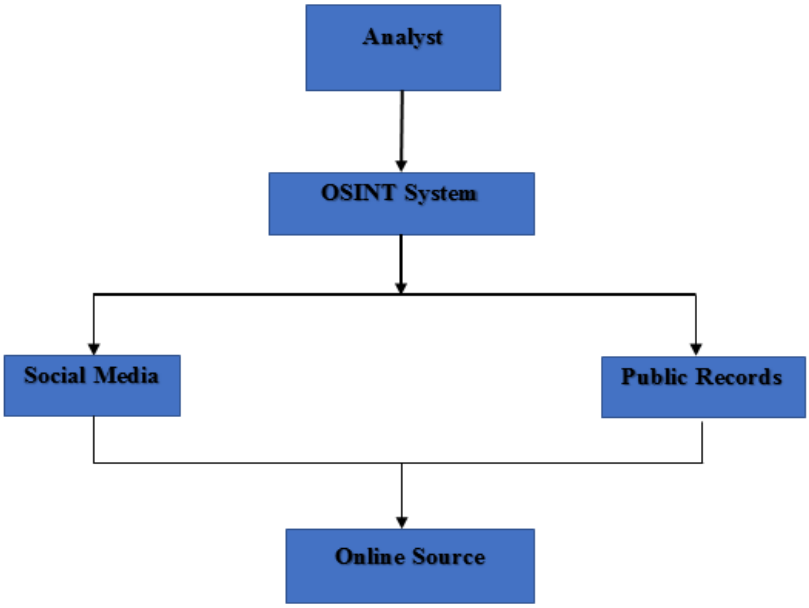
Keywords: OSINT, AI, ML, Data Mining, Web Scraping, GAL

INTRODUCTION

In today's digital age, the volume and variety of information available online has grown exponentially. From news articles and social media posts to public records and government reports, there is an overwhelming amount of data that can be accessed and analysed with the right tools and techniques. Open-Source Intelligence (OSINT) has emerged as a powerful and versatile tool for gathering and analysing this open-source information, providing valuable insights into a range of domains, from business and marketing to national security and law enforcement.

OSINT is an emerging discipline that involves the collection, analysis, and dissemination of information that is available to the public. This information can be gathered from a variety of sources, including the internet, social media, news outlets, and public records. OSINT analysts use a range of techniques and tools to gather and analyse this information, from web scraping and data mining to social media analysis and geolocation tracking. The goal of OSINT analysis is to extract valuable insights from this open-source information, such as patterns and trends in public opinion, potential security threats, or emerging market opportunities.

OSINT (Open Source Intelligence) diagram:



Analyst: The person who is responsible for analysing the information gathered from the OSINT system.

OSINT System: This is the system that collects, processes, and analyzes open-source intelligence.

Social Media: This includes information gathered from social media platforms such as Twitter, Facebook, Instagram, etc.

Public Records: This includes information gathered from public records such as court records, property records, government documents, etc.

Online Sources: This includes information gathered from various online sources such as blogs, forums, websites, etc.

Overall, an OSINT diagram shows how the different sources of open-source intelligence are integrated into the OSINT system, which is then analysed by an analyst to provide actionable intelligence.

Various authors ([1],[2],[3],[4],[5]) have worked in the area of OSINT.

In this paper, our aim to explore the power of OSINT by examining its various techniques, tools, and applications.

Types of OSINT

There are several types of OSINT. Here are some of the main categories:

1. **Web-based OSINT:** This involves gathering information from publicly accessible websites, social media, and other online sources.
2. **Geospatial OSINT:** This involves using satellite imagery, maps, and geographic information systems (GIS) to gather information about specific locations.
3. **Media OSINT:** This involves monitoring news media, including traditional media outlets and online news sources, to gather information on events, trends, and issues.
4. **Financial OSINT:** This involves analysing financial data to identify trends, risks, and opportunities related to companies, markets, and industries.

5. **Technical OSINT:** This involves analysing technical data, such as network traffic, software vulnerabilities, and system logs, to identify security threats and vulnerabilities.
6. **Legal OSINT:** This involves gathering information related to legal issues, such as court cases, public records, and legal databases.
7. **Human OSINT:** This involves gathering information through human sources, such as interviews, surveys, and focus groups.

Overall, OSINT techniques can be used in combination with one another to gather a broad range of information and insights.

ADVANTAGES AND DISADVANTAGES

Advantages of OSINT

1. **Cost-effective:** OSINT is relatively low cost and does not require significant investments in specialized equipment or technology.
2. **Broad range of sources:** OSINT can be obtained from various sources such as news articles, social media, blogs, and other publicly available information sources.
3. **Timely:** OSINT can be collected and analysed in real-time, allowing for rapid response and decision-making.
4. **Increased Accuracy:** OSINT can provide a high level of accuracy and reliability, as it draws on multiple sources of information.
5. **Accessibility:** OSINT is accessible to anyone with an internet connection, making it a valuable tool for both individuals and organizations.
6. **Provides a Global Perspective:** OSINT can provide a global perspective on events, issues, and trends, allowing for a more comprehensive understanding of the situation.
7. **Supports Decision-Making:** OSINT can provide valuable insights and information to support decision-making processes in various domains, such as security, business, and policy.
8. **Enhances Situational Awareness:** OSINT can help enhance situational awareness by providing real-time updates on events and situations.
9. **Helps Identify Emerging Threats:** OSINT can help identify emerging threats and trends that may not be apparent through traditional intelligence gathering methods.
10. **Improves Public Accountability:** OSINT can help hold governments and organizations accountable by providing transparency and access to information that may otherwise be hidden.

Disadvantages of OSINT

1. **Limitations in Data Collection:** OSINT is limited to information that is publicly available and can be accessed through the internet.
2. **Reliability:** The accuracy and reliability of OSINT information can be affected by the quality of the source, the bias of the author, and the timeliness of the information.
3. **Ethical Concerns:** The collection and use of OSINT data can raise ethical concerns related to privacy, confidentiality, and data protection.
4. **Overwhelming amount of data:** The volume of OSINT data available can be overwhelming and difficult to manage and analyse, requiring specialized skills and technology.

5. **Interpretation Challenges:** OSINT data can be difficult to interpret, requiring expert knowledge and skills in data analysis and intelligence gathering.
6. **Quality and relevance issues:** OSINT can be of varying quality and relevance, making it difficult to identify accurate and actionable information.
7. **Data Overload:** The sheer volume of data available through OSINT can lead to data overload, making it difficult to manage and analyse effectively.
8. **Security Concerns:** The collection and use of OSINT data can raise security concerns, as it may be possible for malicious actors to use this information for nefarious purposes.
9. **Limited access to sensitive information:** OSINT is limited to publicly available information and may not provide access to sensitive information, such as classified or proprietary data.
10. **Language and Cultural Barriers:** OSINT data may be in a different language or cultural context, requiring specialized skills and knowledge to interpret and analyse.

OSINT Methodologies

Open-Source Intelligence is a methodology that involves the collection, analysis, and dissemination of information from publicly available sources. Here are some of the common methodologies used in OSINT:

1. **Searching:** This involves using various search engines, websites, and other online resources to find information related to the investigation.
2. **Social Media Monitoring:** This involves monitoring social media platforms like Facebook, Twitter, and Instagram for information related to the investigation.
3. **Data Mining:** This involves using data mining techniques to analyse large volumes of data to identify patterns, relationships, and other relevant information.
4. **Link Analysis:** This involves analysing the relationships between people, organizations, and other entities to identify connections and patterns.
5. **Geolocation:** This involves using geolocation tools to identify the location of a person, organization, or event.
6. **Dark Web Research:** This involves researching the dark web for information related to the investigation.
7. **Forensic Analysis:** This involves analysing digital evidence such as emails, images, and other files to gather information related to the investigation.
8. **Human Intelligence:** This involves gathering information from human sources, such as interviews with witnesses or experts in the field.
9. **Satellite Imagery Analysis:** This involves analysing satellite imagery to gather information related to the investigation.
10. **Public Records Searches:** This involves searching public records, such as court records, property records, and business registrations, to gather information related to the investigation.
11. **Image and video analysis:** This involves analysing images and videos to gather information, such as identifying people, objects, and locations.
12. **Financial Analysis:** This involves analysing financial records to gather information, such as identifying the financial status of an individual or organization.

13. **Competitive Intelligence:** This involves gathering information about competitors, such as their products, pricing, and marketing strategies.
14. **Reputation Analysis:** This involves analysing the online reputation of an individual or organization, such as their social media presence, reviews, and media coverage.
15. **Linguistic Analysis:** This involves analysing language used in online communications, such as emails and social media posts, to gather information related to the investigation.
16. **Event Monitoring:** This involves monitoring news sources, social media, and other online resources for information related to a specific event or situation.

OSINT methodologies can be used individually or in combination with other methodologies to gather information and intelligence. The key is to use the right methodology for the specific investigation and to stay up-to-date with the latest tools and techniques available.

Gathering and Analysing (GAL) the information via OSINT:

The emergence of the internet and social media has created an unprecedented level of accessibility to vast amounts of data, much of which is publicly available. OSINT has become an increasingly important tool for GAL the information to inform decision-making across a wide range of fields, from academic and non-academic organisation, business and finance to national security and law enforcement etc.

The important is advantages of OSINT is its ability to provide real-time information on a wide range of topics, including emerging trends, public opinion, and breaking news. OSINT is based on publicly available data that can be accessed by anyone with an internet connection. This makes OSINT a highly versatile tool that can be used by businesses, governments, and individuals alike.

OSINT is the practice of collecting and analysing publicly available information to gain insights and knowledge about a particular topic, person, or organization.

Examples: GAL of information using OSINT:

1. **Social Media Analysis:** Social media platforms such as Twitter, LinkedIn, Facebook, and Instagram are great sources of information for OSINT. Users may gather information about a particular event, person, group of persons or organization/organisations by analysing their social media accounts. Users may look strictly for patterns in their behaviours, connections, interests and day-to-day involvement.
2. **Website Analysis:** Websites are another great source of information. By examining a website, users may gather information about the organization, the people associated with it, and any products or services they offer and also examine the source code of a website to gather additional information.
3. **Public Records:** Public records such as court documents, property records, and business filings can provide valuable information for OSINT users may search for these records online or visit local government offices to obtain them.
4. **News Articles:** News articles may provide valuable insights into a particular topic or organization. On this basis users may search for news articles online or use a news aggregator to gather relevant articles.
5. **Forums and Message Boards:** Forums and message boards can be a great source of information for OSINT. By analysing posts and discussions, users may gain insights into the attitudes and opinions of a particular group or community.
6. **Image Analysis:** Image analysis involves examining images or videos to gather information. By analysing the content of an image, users may identify objects, people, and locations and also use tools such as reverse image search to find other instances of the same image.

7. **Geo-Location:** Geolocation involves using data such as IP addresses and GPS coordinates to determine the location of a person or organization. By analysing geolocation data, users may gain insights into where a person or organization is located, and their movements over time.
8. **Social Network Analysis:** Social network analysis involves examining the relationships between people and organizations. By analysing social network data, users may identify key influencers and connections between different groups.
9. **Dark Web Analysis:** The dark web is a hidden part of the internet that is not accessible through regular search engines. By analysing dark web data, users may gather information about illegal activities, hacking groups, and other topics that are not available through regular OSINT methods.
10. **Financial Analysis:** Financial analysis involves examining financial data such as income statements, balance sheets, and cash flow statements. By analysing financial data, users may gain insights into the financial health of an organization and their business practices.
11. **Use Automation:** There are many tools and software available that can help you automate the process of gathering and analysing information. For example, users may use web scrapers to collect data from websites, or social media monitoring tools to track mentions of a particular keyword or hashtag.
12. **Verify Sources:** When gathering information using OSINT, it's important to verify the sources to ensure that the information is accurate and reliable and this tool may check the credibility of the source and cross-reference it with other sources to confirm the information.
13. **Use Multiple Sources:** To get a comprehensive understanding of a topic or organization, it's important to gather information from multiple sources and use a variety of sources such as news articles, social media, and public records to gather information from different perspectives.
14. **Stay Up-to-Date:** The information landscape is constantly changing, so it's important to stay up-to-date with the latest trends and developments. Available tool may keep an eye on news and social media trends, and subscribe to industry newsletters to stay informed.
15. **Practice Ethical OSINT:** When gathering information using OSINT, it's important to follow ethical guidelines and respect the privacy of individuals and organizations and not engaging in illegal activities or use unethical methods to gather information.

Once users using above method for gathering information user may use various analysis techniques to draw insights and conclusions. e.g., Users may use predictive analytics to forecast future trends, or sentiment analysis to understand public opinion on a particular topic.

Tools and Applications

OSINT tools and applications are software tools and platforms designed to help investigators, analysts, and researchers gather and analyse publicly available information. Some popular OSINT tools and applications are following:

1. **Maltego:** A data visualization tool used to collect, analyse, and visualize data from multiple sources, such as social media, websites, and network infrastructure.
2. **Shodan:** A search engine that can search for devices connected to the internet and identify vulnerabilities in network infrastructure.
3. **Hunchly:** A web capture tool used to record all online activity, including websites visited, searches conducted, and data downloaded.
4. **Recon-ng:** A command-line tool used to automate the process of gathering information from various sources, such as social media, search engines, and company databases.

5. **Social-Searcher:** A social media search engine that allows users to search for mentions of specific keywords or hashtags on social media platforms.
6. **Tinfoleak:** A tool for analysing Twitter activity and identifying patterns and trends in Twitter data.
7. **Osintgram:** A tool for searching and analysing Instagram content based on hashtags and usernames.
8. **SpiderFoot:** A tool for automating the process of gathering information from various sources, including search engines, social media, and company databases.
9. **Dataminr:** A real-time information discovery platform used to detect breaking events and emerging trends.
10. **Echosec:** A social media discovery and monitoring platform used to identify online threats and gather intelligence on specific individuals or organizations.

These tools and applications help investigators, researchers, and analysts to gather, analyse and visualize data from different sources to extract useful insights and intelligence for various purposes such as investigations, research, and risk assessment.

OSINT Framework:

The OSINT Framework consists of several categories, each of which includes various tools and resources:

- **Information Gathering:** This category includes tools and resources for gathering information from search engines, social media platforms, domain name databases, and other public sources.
- **People Search:** This category includes tools for finding information about individuals, such as social media profiles, email addresses, and phone numbers.
- **Email Research:** This category includes tools for researching email addresses, including verifying email addresses, identifying email providers, and analysing email headers.
- **Social Media:** This category includes tools for monitoring social media platforms, tracking hashtags and keywords, and analysing social media data.
- **Image Search:** This category includes tools for searching for images on the internet, including reverse image search engines, facial recognition tools, and image metadata analysis tools.
- **Domain Search:** This category includes tools for researching domain names, including domain name registration information, domain name history, and DNS records.
- **Dark Web:** This category includes tools for monitoring activity on the dark web, including Tor hidden services, marketplaces, and forums.

Updating in OSINT framework is required with help of new tools and resources to handle any new types of problem.

OSINT Architecture:

The architecture of an OSINT system typically includes three main components:

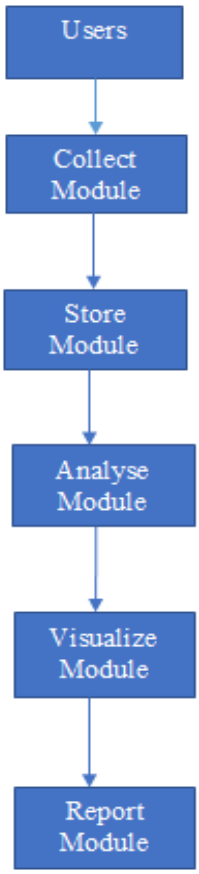
1. **Data Collection:** This component involves the collection of data from various sources, such as social media, news websites, forums, blogs, and other public online sources. The collection of data can be automated or manual, depending on the type of data and the level of accuracy required. Data collected from these sources is often in unstructured format, and it needs to be processed and analysed to extract meaningful information.

2. **Data Processing and Analysis:** This component involves the processing and analysis of the collected data to extract useful information. This involves converting the unstructured data into a structured format that can be analysed using different data mining and analysis techniques, such as natural language processing, sentiment analysis, and network analysis. The data processing and analysis component also involves filtering, sorting, and categorizing the collected data to identify patterns, trends, and relationships.
3. **Information Dissemination:** This component involves the dissemination of the analysed information to the relevant stakeholders. This can be done through different channels, such as reports, dashboards, alerts, and notifications. The information can be presented in various formats, such as graphs, charts, and maps, to make it more accessible and understandable to the stakeholders.

Overall, the architecture of an OSINT system is designed to support the collection, processing, analysis, and dissemination of information from various sources to support decision-making, planning, and other strategic activities. The architecture can be customized based on the specific needs and requirements of the organization or the application.

OSINT Architecture with diagram:

Here is a basic architecture diagram for an OSINT system:



Users: This is the starting point of the OSINT system, where users access and interact with the system. They can specify what information they are looking for, what sources to use, and what type of output they want to receive.

Collect Module: This module is responsible for collecting data from different sources such as social media, blogs, news websites, etc. It uses different tools such as web crawlers, scrapers, and APIs to extract relevant data.

Store Module: Once data is collected, it needs to be stored in a database or data warehouse for further analysis. The Store Module takes care of this task by storing data in a structured way so that it can be easily retrieved and analysed.

Analyse Module: This module is responsible for analyzing the data that has been collected and stored. It uses various techniques such as data mining, text analytics, and natural language processing to extract insights and patterns from the data.

Visualize Module: Once data is analysed, it needs to be presented in a way that is easy to understand and interpret. The Visualize Module takes care of this task by creating visualizations such as graphs, charts, and maps to represent the data.

Reports Module: The Reports Module generates reports based on the analysed data and the user's requirements. Reports can be generated in different formats such as PDF, Excel, and HTML.

Overall, an OSINT architecture consists of several modules that work together to collect, store, analyse, and present data to users. The modules can be customized and extended to meet specific requirements and needs.

CONCLUSION

In conclusion, OSINT provides a powerful toolset for gathering and analysing information from a wide range of sources. By using the right methods and analysis techniques, users may gain valuable insights and knowledge that can inform business decisions, investigations, and other activities to safe their self from attackers, thus OSINT becoming emerging tools that will be needful entity of any organisations.

REFERENCES

[1] Javier Pastor-Galindo, Pantaleone Nespoli, Félix Gómez Mármol, And Gregorio Martínez "The Not Yet Exploited Goldmine of OSINT: Opportunities, Open Challenges and Future Trends" IEEE ACCESS, Volume 8, 2020, Digital Object Identifier 10.1109/ ACCESS.2020.2965257

[2] J. Smith and M. Johnson, "Open-Source Intelligence Techniques for Investigative Journalism," IEEE Transactions on Professional Communication, vol. 62, no. 3, pp. 188-200, Sep. 2019, doi:10.1109/TPC.2019.2936699.

[3] Yong-Woon Hwang,1Im-Yeong Lee,1Hwankuk Kim,2Hyejung Lee,3and Donghyun Kim "Current Status and Security Trends of OSINT" Hindawi Wireless Communications and Mobile Computing, Volume 2022 | Article ID 1290129 | https://doi.org/ 10.1155/ 2022/ 1290129.

[4] João Rafael Gonçalves Evangelista, Renato José Sassi, Márcio Romero & Domingos Napolitano "Systematic Literature Review to Investigate the Application of Open-Source Intelligence", (OSINT) with Artificial Intelligence Pages 345-369 | Published online: 07 May 2020 Journal of Applied Security Research, DOI: 10.1080/19361610.2020.1761737

[5] Hamzeh Alkilani, Abdallah Qusef "OSINT Techniques Integration with Risk Assessment ISO/IEC 27001 DATA'21: International Conference on Data Science, E-learning and Information Systems 2021April 2021 Pages 82–86 https://doi.org/10.1145/ 3460620. 3460736

[6] Isabelle Böhm & Samuel Lolagar, " Open source intelligence",Introduction, legal, and ethical Considerations, International Cybersecurity Law Review volume 2, pages317–337 (2021)

[7] Tomislav Ivanjko, Tomislav Dokman, "Open-Source Intelligence (OSINT): Issues and Trends" Conference Paper · January 2020 DOI: 10.17234/INFUTURE.2019.23

[8] Riccardo Ghioni, Mariarosaria Taddeo, Luciano Floridi, "Open-source intelligence and AI: a systematic review of the GELSI literature",9 January 2023, AI & SOCIETY https:// doi.org/ 10.1007/s00146-023-01628-x

THE IMPACT OF SOCIAL MEDIA ON YOUTH

Rishabh Sinha
Post Graduate student of Master of Computer Application, Amity University, Patna

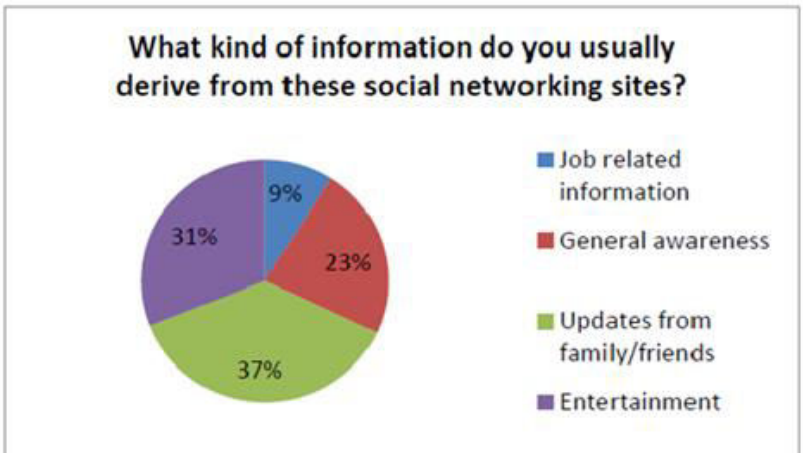
ABSTRACT
Social media is a pervasive aspect of contemporary culture that has been rapidly adopted by young people across the world. However, as social media platforms have become increasingly integrated into the daily lives of young people, there have been concerns raised about the potential negative impact on their mental and emotional wellbeing. This research paper provides a comprehensive review of the literature exploring the effects of social media on youth, with a particular focus on the impacts on mental health, self-esteem, body image, and social relationships. The findings suggest that while social media use can have both positive and negative effects on youth, there is a growing body of evidence that highlights the potential harm associated with excessive social media use, particularly in relation to mental health. The paper concludes with recommendations for policymakers, educators, parents, and young people themselves, on how to mitigate the negative effects of social media and promote healthy use

INTRODUCTION
• Social media has become a ubiquitous aspect of contemporary culture, with millions of people using platforms such as Facebook, Instagram, Twitter, and TikTok every day. However, while social media has many benefits, there are growing concerns about the impact of social media on young people. In particular, there are concerns that excessive social media use may be detrimental to young people's mental and emotional wellbeing, leading to issues such as anxiety, depression, low self-esteem, and poor body image.

• This research paper provides a comprehensive review of the literature exploring the effects of social media on youth, with a particular focus on the impacts on mental health, self-esteem, body image, and social relationships. The paper begins by discussing the definition of social media and its prevalence among young people. It then reviews the literature on the potential positive and negative effects of social media on youth, before discussing some of the key issues and challenges associated with social media use by young people.

• **Definition and Prevalence of Social Media Among Youth:** Social media is a broad term used to describe a range of digital platforms that allow people to share content, interact with others, and build communities online. Social media platforms are characterized by user-generated content, real -time interaction, and the ability to connect with others across geographic and cultural boundaries. Some of the most popular social media platforms among young people include Facebook, Instagram, Snapchat, TikTok, and Twitter.

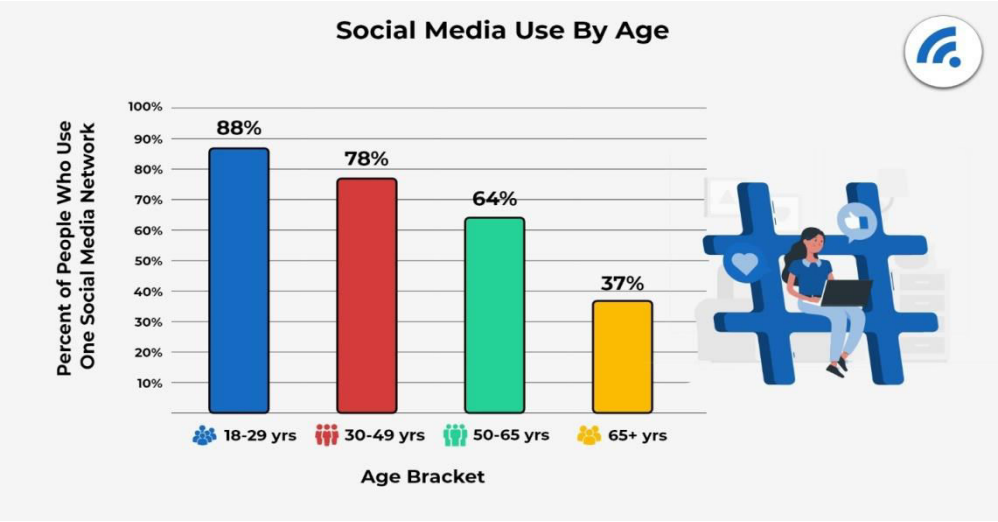
• Social media use is extremely prevalent among young people, with recent research suggesting that almost all teenagers in the United States use some form of social media on a daily basis (Pew Research Center, 2019). Similarly, a study conducted by Common Sense Medi found that teenagers spend an average of 7 hours and 22 minutes per day consuming media, with the majority of this time spent on social media (Common Sense Media, 2019).



Positive Effects of Social Media on Youth: Despite concerns about the potential negative effects of social media on young people, there are also many potential benefits associated with social media use. For example, social media can provide young people with access to information, support, and resources that may not be available in their immediate environment. Social media can also facilitate the development of social connections and communities, particularly for young people who may feel isolated or marginalized in their offline lives.

• Research has also suggested that social media can have positive effects on young people's mental health. For example, a study conducted by the University of Michigan found that social media use was associated with higher levels of self-esteem and lower levels of social anxiety among young adults (Valkenburg & Peter, 2009). Similarly, research has shown that social media can be an effective tool for promoting positive health behaviors, such as exercise, healthy eating, and smoking cessation (Korda & Itani, 2013).

Negative Effects of Social Media on Youth: While social media use can have many potential benefits for young people, there are also many potential negative effects associated with excessive or problematic social media use. One of the most commonly cited negative effects is the impact on mental health. Several studies have found associations between heavy social media use and symptoms of depression, anxiety, and low self-esteem among young people (Boyd, 2014; Lin et al., 2016). This association may be due to factors such as social comparison, cyberbullying, and the constant exposure to idealized and curated versions of others' lives.



• In addition to mental health issues, social media use has also been linked to negative body image and disordered eating behaviors among young people, particularly adolescent girls (Fardouly et al., 2015; Perloff, 2014). The constant exposure to highly edited and filtered images of idealized bodies on social media platforms can contribute to body dissatisfaction and unhealthy behaviors in pursuit of an unrealistic body image.

• Furthermore, excessive social media use can lead to social isolation and a decline in face-to-face social interactions. Spending excessive time on social media can reduce the amount of time young people spend engaging in offline activities and building meaningful relationships, leading to feelings of loneliness and social disconnectedness

Challenges and Recommendations: Addressing the negative effects of social media on youth requires a multi-faceted approach involving various stakeholders, including policymakers, educators, parents, and young people themselves. Some potential strategies include:

• **Education and Digital Literacy:** Providing young people with the skills and knowledge to critically evaluate and navigate social media platforms can help them develop healthier online behaviors and mitigate the negative impact of social media.

• **Parental Involvement and Support:** Parents play a crucial role in guiding their children's social media use. Encouraging open communication, setting limits, and being aware of their children's online activities can help parents support their children in developing healthy social media habits.

• **Platform Design and Regulation:** Social media platforms can implement features and algorithms that prioritize user well-being, such as providing tools to manage screen time, filtering harmful content, and promoting positive interactions. Policymakers can also play a role in regulating social media platforms to protect young people from potential harm.

• **Mental Health Support:** Schools, healthcare providers, and community organizations should prioritize mental health support services for young people, particularly those who may be at higher risk due to excessive social media use or negative online experiences addressing the negative effects of social media on youth requires a multi-faceted approach involving various stakeholders, including policymakers, educators, parents, and young people themselves. Some potential strategies include:

• **Education and Digital Literacy:** Providing young people with the skills and knowledge to critically evaluate and navigate social media platforms can help them develop healthier online behaviors and mitigate the negative impact of social media.

• **Parental Involvement and Support:** Parents play a crucial role in guiding their children's social media use. Encouraging open communication, setting limits, and being aware of their children's online activities can help parents support their children in developing healthy social media habits.

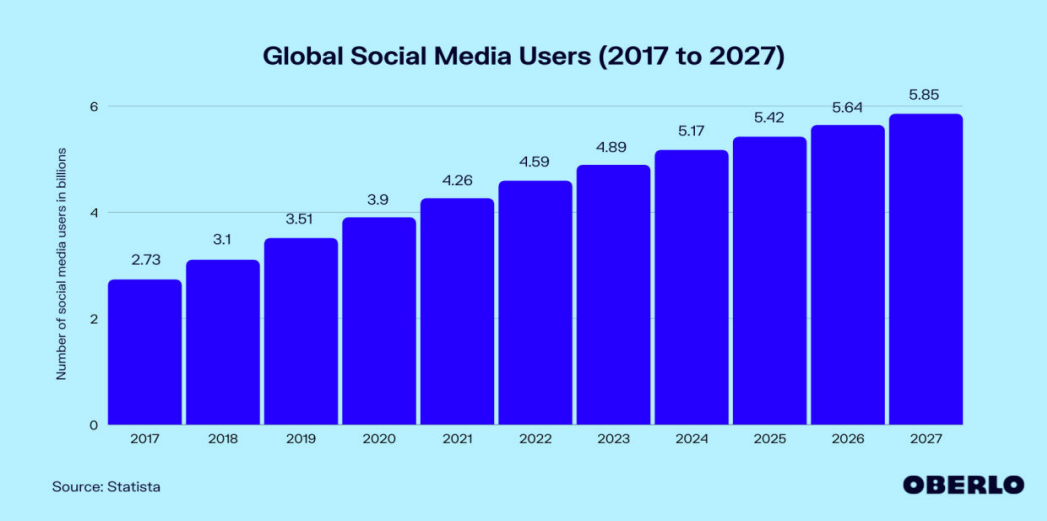
• **Platform Design and Regulation:** Social media platforms can implement features and algorithms that prioritize user well-being, such as providing tools to manage screen time, filtering harmful content, and promoting positive interactions. Policymakers can also play a role in regulating social media platforms to protect young people from potential harm.

• **Mental Health Support:** Schools, healthcare providers, and community organizations should prioritize mental health support services for young people, particularly those who may be at higher risk due to excessive social media use or negative online experiences

CONCLUSION

Social media has become an integral part of young people's lives, offering both benefits and challenges. While social media can facilitate connection, information sharing, and positive experiences, excessive or problematic use can have detrimental effects on mental health, self-

esteem, body image, and social relationships. It is essential for stakeholders to work together to address these concerns and promote healthy social media use among youth. By providing education, support, and regulating platforms, we can help young people navigate the digital landscape and reap the benefits while minimizing the potential harm of social media.



REFERENCES

- Boyd, D., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. Journal of Computer-Mediated Communication, 13(1), 210-230.
- Kuss, D. J., & Griffiths, M. D. (2011). Online social networking and addiction—A review of the psychological literature. International Journal of Environmental Research and Public Health, 8(9), 3528-3552.
- Lenhart, A. (2015). Teens, social media & technology overview 2015. Pew Research Center.
- Moreno, M. A., Jelenchick, L. A., Egan, K. G., Cox, E., Young, H., Gannon, K. E., & Becker, T. (2011). Feeling bad on Facebook: Depression disclosures by college students on a social networking site. Depression and Anxiety, 28(6), 447-455.
- Primack, B. A., Shensa, A., Sidani, J. E., Whaite, E. O., Lin, L. Y., Rosen, D., ... & Miller, E. (2017). Social media use and perceived social isolation among young adults in the US. American Journal of Preventive Medicine, 53(1), 1-8.
- Rosen, L. D., Whaling, K., Rab, S., Carrier, L. M., & Cheever, N. A. (2013). Is Facebook creating "iDisorders"? The link between clinical symptoms of psychiatric disorders and technology use, attitudes and anxiety. Computers in Human Behavior, 29(3), 1243-1254.
- Twenge, J. M., Joiner, T. E., Rogers, M. L., & Martin, G. N. (2018). Increases in depressive symptoms, suicide-related outcomes, and suicide rates among US adolescents after 2010 and links to increased new media screen time. Clinical Psychological Science, 6(1), 3-17.
- Valkenburg, P. M., Peter, J., & Schouten, A. P. (2006). Friend networking sites and their relationship to adolescents' well-being and social self-esteem. CyberPsychology & Behavior, 9(5), 584-590.
- Vannucci, A., Flannery, K. M., & Ohannessian, C. M. (2017). Social media use and anxiety in emerging adults. Journal of Affective Disorders, 207, 163-166.
- Zhang, R., Li, D., & Qiu, J. (2018). The association between social media addiction and academic performance among Chinese adolescents: The mediating role of self-esteem and self-control. Cyberpsychology, Behavior, and Social Networking, 21(11), 711-718.

BLAST FURNACE HYDROGEN INJECTION: INVESTIGATING IMPACTS IN TATA STEEL LTD

[1]Sushant Shekhar and [2]Smiti Tiwari
[1]Union Committee Member, Tata Workers Union, Tata Steel Ltd. Jamshedpur
[2]Training & Development Coordinator, Tata Steel Ltd., Jamshedpur

ABSTRACT
A significant contributor to the operation of contemporary society, the steel sector accounts for about 8% of the world's carbon dioxide emissions. Industries must look for ways to considerably decrease or curtail their emissions as the world moves to address the obvious and growing effects of greenhouse gas emissions. The operation of the blast furnace (BF), a substantially fossil fuel reliant reactor that continues to be the primary technique of ore reduction worldwide, is one such area in which steelmaking may address these challenges.

Recent years have seen an increase in interest in the injection of hydrogen reducing gas into the blast furnace; nevertheless, the stability of blast furnace operation is directly impacted by the injection of low-carbon fuel, with restrictions caused by the quenching effects of the stimulated chemical reactions. Hydrogen injection is a promising solution being developed to lower CO_2 emissions in ironmaking blast furnaces (BFs). A recent process model based on computational fluid dynamics (CFD) is used to study hydrogen BF.

Tata Steel, the pioneer in steelmaking industry has adopted this sustainable alternative by injection of hydrogen gas using 40% of the injection systems in of its Blast Furnaces at Jamshedpur Works. This is the first time ever that a blast furnace is being continuously supplied with such a big volume of hydrogen gas. The project has the potential to cut the coke rate by 10%, which would result in a 7–10% reduction in CO_2 emissions per tonne of crude steel produced and is in line with the company's goal of reaching Net Zero by 2045. This paper aims to explore the influence of peripheral opening extent (POE), which indicates the amount of coke close to the boiler wall and the advantages of this transition away from traditional fossil fuels.

Keywords: hydrogen; injection; blast furnace; carbon dioxide emissions; carbon capture

INTRODUCTION
Hydrogen injection into the blast furnace is a promising ironmaking technology. Compared to carbon, hydrogen offers a number of advantages as a reductant, including fewer CO2 emissions, higher thermal conductivity, and larger diffusivity favours iron ore reduction while lower viscosity and density result in less pressure loss. Metallurgists have studied the blast furnace hydrogen injection extensively up to this point, and their findings are impressive. The research on the injection of hydrogen into blast furnaces is primarily split from the standpoint of research methodologies into three categories: research on theoretical analysis [1-4], research on physical experiment [5, 6], and research on numerical simulation [7-9]. According to theoretical analysis and research, Wang [2] examined and demonstrated the viability of hydrogen injection in an oxygen blast furnace using The Conservation of Mass and Energy; Kim [1] found that hydrogen injection will increase the slope of the operating line that includes C and H2, but reduce the carbon consumption; While Li Bin[4] developed a thermodynamic model of the gas-solid reduction reaction of iron oxides based on the principle of minimum Gibbs free energy and studied the thermodynamics of the gas-solid reduction of iron oxides, Bernasowski [3] investigated the impact of gas mixtures of CO and H_2 in different proportions on the reduction of iron oxides under equilibrium conditions and the presence of carbon in the system.

Coke is necessary for maintaining good bed permeability, which results in a specific amount of carbon consumption, in the BF ironmaking process. Also, for the reduction of iron ore,

hydrogen and carbon are inevitably in competition. As a result, the amount of hydrogen that may be used in an HBF is limited. An important part of HBF research and development is on forecasting the hydrogen usage cap and then proposing potential solutions to raise it. The gas distribution when passing through burden materials is determined by the burden distribution, which is a representation of the spatial locations of particles inside a BF.

Therefore, the burden distribution patterns can have a big impact on the gas flow, the accompanying heat and mass transfer, momentum transfer and chemical processes, and the cohesive zone (CZ), which greatly influences BF overall performance. Coke and iron ore are two examples of the load materials that are frequently charged into a modern BF to create alternate layers of coke and ore. Due to its smaller size and higher density compared to coke, iron ore is more resistant to gas flow. As a result, changing the gas distribution to obtain the desired BF performance involves routinely altering the load distribution pattern. Changing the ore-to-coke ratio's (O/C) radial profile is one of the most crucial burden distribution management measures, but it is also one of the trickiest for HBF. Particular interactions occur between the hydrogen pumped into an HBF through the hearth and tuyeres as well as with other gas constituents including CO and nitrogen. To maximise the use of hydrogen and obtain the best HBF performance, burden distribution control must take these interactions into account. However, it is yet unknown how distinct gas constituents interact with one another and impact HBF performance under varied radial O/C profiles. Therefore, in order for HBF to successfully industrialise, this issue must be solved.

The impacts of burden distributions have previously been the subject of intensive research. For instance, a number of measurement methods, including probes, radar, and metal grid measurements [10], have been developed to examine the trajectory, filling points, and profiles of loads. The majority of numerical models used to examine burden distributions were DEM-based and concentrated on traditional BF activities. We are aware of very few attempts to investigate the impact of weight distribution on shaft injection using a combined DEM and CFD (computational fluid dynamics) technique. However, the procedures under room temperature were primarily studied in the earlier DEM/CFD-DEM studies of burden distributions, which overlooked the associated thermochemical behaviours.

To simulate BFs and determine the inner states and overall performance under industrial operating and geometric conditions, CFD process models have been widely adopted. Using CFD BF process models, conventional and new BF processes have been extensively studied under different conditions.[11] In recent years, they were also used to study HBFs. For instance, a CFD BF process model is adopted by Nogami et al. [12] to study the BF operated with varying hydrogen enrichment through hearth tuyeres from 0 to 43.7 pct. Via the same model, Tang et al. [13] revealed the influence of hydrogen injection with the hydrogen enrichment of up to 15.23 pct; Chu et al. [14] compared the hydrogen-bearing material injection with all-coke operations. Li et al. [15] evaluated the effects of the belly injection of reformed coke oven gas (RCOG), hot burden charging, and their combination on BF performance using a 2D process model. Using a similar model, Yu and Shen[16] studied how the shaft injection of pure hydrogen affected BF performance. The amount of shaft-injected hydrogen that penetrates into the bed column of an HBF changes the interaction between H2 and CO, the utilisation efficiency, and the final use of hydrogen in the furnace. There is currently very little knowledge in this area. However, it is strongly anticipated that the cutting-edge HBF ironmaking technology would be investigated.

This paper examines the impact of top burden distribution on an industrial BF using a recently published 3D CFD process model that has been validated for HBF. Hydrogen is injected through the shaft and hearth tuyeres to power the BF. In terms of peripheral opening extent (POE), various top burden distribution patterns are specifically taken into consideration. The amount of coke close to the boiler wall is represented by the POE. The relationship between the

POE and the flow rate of shaft-injected hydrogen is also taken into account in this study. Analysis and connections between the BF's internal states and performance as a whole are made. The responsibilities of the peripheral opening operation in influencing hydrogen utilisation and usage in a BF are made clear in this study.

Computational Models

Three main sub-models make up the computational fluid dynamics approach used in this study to simulate the internal condition of the blast furnace, created with the goal of capturing the various mechanics (and timescales) seen in the blast furnace. The first one is the tuyere-blowpipe simulation model, which uses the commercial CFD solver ANSYS Fluent to forecast combustion of fuels injected with hot blast before entering the furnace proper; the second is the raceway simulation, which employs both ANSYS Fluent and an internal CFD solver to forecast the size and shape of the raceway envelope as well as the gas temperature and species distributions; and the third is the shaft simulation, which employs an internal CFD model to forecast ore. The tuyere model generates gas temperature, species distributions, and flow rates for the raceway model inlet boundary, and the raceway model provides comparable data for use by the shaft model inlet at the furnace bosh. These models collaborate at critical boundaries between major phenomena regions.

These models were developed under several significant assumptions, which are fully described in earlier papers [17-33] .For reference, a high-level summary of the model's assumptions and details will be given below. In addition to dividing the furnace into the three crucial reaction regions of the tuyere, raceway, and shaft, symmetry around the furnace axis is assumed for all regions in order to keep the computing work for the investigation of variable parameters to a manageable level.

In order to make a thin slice of the furnace representative, it is assumed that all tuyeres are operating at the same injection and flow rates, that the raceway region is periodically symmetric, and that the shaft is axisymmetric with respect to gas flow, temperature distributions, and burden distribution. These presumptions have been effectively applied, and simulation results have been verified against data from actual industrial operations.

Table I. Governing Equations	
Mass: $\Phi = 1$	$\nabla . (\rho \Phi u) = \nabla . (\Gamma_{\Phi} \nabla \Phi) - \nabla . (\rho u^{t} \Phi^{t}) + S_{\Phi}$
Momentum: Φ = velocity	
Energy: Φ= enthalpy	
Species Transport	$\nabla . (\rho u Y_i) = \nabla . (\rho \Gamma_i \nabla Y_i) + R_i + S_i$
Turbulent Kinetic Energy	$\nabla . (\rho k U) = \nabla . ((\mu_t/\sigma_k) \nabla k) + 2\mu_t S_{ij} . S_{ij} - \rho\varepsilon$
Turbulence Dissipation Rate	$\nabla . (\rho \varepsilon U) = \nabla . ((\mu_t/\sigma_\varepsilon) \nabla \varepsilon) + C_{1\varepsilon}(\varepsilon/k) 2\mu_t S_{ij} . S_{ij} - C_{2\varepsilon}\rho(\varepsilon^2/k)$
Where, $\mu_t = C\rho v l = \rho C_\mu (k^2/\varepsilon)$, $C_\mu = 0.09$, $\sigma_k = 1.00$, $\sigma_\varepsilon = 1.30$, $C_{1\varepsilon} = 1.44$, and $C_{2\varepsilon} = 1.92$	

Table II. Chemical reactions modeled in the raceway region					
Reaction	**No.**	**Chemical equation**	**A_S [1/s]**	**B_i [m^3/(kg*s)]**	**Act. energy E_i [J/mol]**
Natural Gas Combustion	R1 Eddy-Diss./ Finite Rate	$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$	$1.6 * 10^{10}$	N/A	$1.081* 10^5$
CO Combustion	R2 Eddy-Diss./ Finite Rate	$2CO + O_2 \rightarrow 2CO_2$	$7.0 * 10^4$	N/A	$6.651 * 10^4$
H2 Combustion	R3 Eddy-Diss./ Finite Rate	$2H_2 + O_2 \rightarrow 2H_2O$	$5.4 * 10^2$	N/A	$1.255 * 10^5$
Coke Oxidation	R4 1st Order Finite Rate	$Coke + O_2 \rightarrow CO_2$	N/A	$1.225 * 10^3$	$9.977 * 10^4$
Boudouard reaction	R5 1st Order Finite Rate	$Coke + CO_2 \rightarrow 2CO$	N/A	$7.351 * 10^3$	$1.380 * 10^5$
Water gas reaction	R6 1st Order Finite Rate	$Coke + H_2O \rightarrow CO + H_2$	N/A	$1.650 * 10^5$	$1.420 * 10^5$
Arrhenius rate (for gas): $\omega_{Arr} = A_s\rho^2 Y_{Fuel} Y_{O2} \exp(-E_s/RT)$, where A_s is a rate constant and E_i is the activation energy 1st Order finite rate: $m_{c,i} = -\pi d^2_p \rho Y_{gasspecies} B_i \exp(-E_i/RT_p)$, where B_i is a rate constant and E_i is the activation energy					

Table III. Reaction mechanisms used in the CFD shaft model		
Reaction	**No.**	**Chemical equation**
Indirect reduction of iron oxide by CO	R1	$3Fe_2O_3(s) + CO(g) \rightarrow 2Fe_3O_4 + CO_2(g)$
	R2	$Fe_3O_4 + CO(g) \rightarrow 3FeO(s) + CO_2(g)$
	R3	$FeO(s) + CO(g) \rightarrow Fe(s) + CO_2(g)$
Indirect reduction of iron oxide by H2	R4	$3Fe_2O_3(s) + H_2(g) \rightarrow 2Fe_3O_4 + H_2O(g)$
	R5	$Fe_3O_4 + H_2(g) \rightarrow 3FeO(s) + H_2O(g)$
	R6	$FeO(s) + H_2(g) \rightarrow Fe(s) + H_2O(g)$
Boudouard reaction	R7	$C(s) + CO_2(g) \rightarrow 2CO(g)$
Water gas reaction	R8	$C(s) + H_2O(g) \rightarrow CO(g) + H_2(g)$
Flux decomposition	R9	$MeCO_3(s) \rightarrow MeO(s) + CO_2(g)$ (Me = Ca, Mg)
Water gas shift reaction	R10	$H_2(g) + CO_2(g) \rightarrow H_2O(g) + CO(g)$
Direct reduction of liquid FeO	R11	$C(s) + FeO(l) \rightarrow Fe(l) + CO(g)$

Researchers modelled the tuyere, injection lances, and upstream blowpipe using the commercial CFD solver ANSYS FluentTM to provide them the flexibility to quickly change the geometry of the tuyere and injection lance for parametric analyses. The turbulent mixing of hot blast and injected fuel as well as combustion reactions are important phenomena represented in this area. The semi-implicit method for pressure-linked equations (SIMPLE) is used to discretize the standard Navier-Stokes equations, and the well-known k-e turbulence model is used to handle turbulence (because the flow is totally turbulent). Chemical reactions are predicted using species transport modelling, with the Eddy-Dissipation / Finite Rate model addressing the impact of turbulence. Table I lists the general controlling equations that were applied in this study.

In the preceding table, ρ stands for density, u for velocity, φ for the general property conveyed, u^t and φ^t represent the fluctuating components of velocity and the transported property due to turbulence. S_φ stands for source term. Y_i stands for the local mass fraction of a species, i, R_i for its net production rate, Γ_i for its species diffusion coefficient, and S_i for its source term. k and ε are the turbulent kinetic energy and dissipation rate respectively, μ_t is the turbulent viscosity, and C_μ, σ_k, $\sigma\varepsilon$, $C_1\varepsilon$, and $C_2\varepsilon$ are model constants defined as listed in Table I.

Iterative modelling of the raceway region is carried out using ANSYS Fluent™ and a custom CFD solver created in Fortran at PNW. A cold-flow interpenetrating two-phase Eulerian multiphase model in Fluent is used to anticipate the creation of the raceway cavity in the coke bed. In this model, the blast flow is driven into a granular phase that represents the coke bed to create the raceway envelope at the end of the tuyere. The raceway envelope is then frozen and exported as a fixed porosity distribution to the internal steady state combustion solution, indicating where the coke bed and void space are located. The internal solver calculates chemical processes, heat fluxes, and gas flow to produce predictions of solid-to-gas mass transfer (from coke burning) and temperature-dependent changes in gas density. To get an updated raceway envelope, the cold-flow simulation in Fluent is performed with these values translated back onto it as source terms. Once convergence is reached, this iterative process is repeated a total of six to eight times.

The combined Eddy Dissipation-Finite Rate Arrhenius model is used to represent chemical processes in the raceway region, with the lower estimated limiting rate. Effective modelling of the coke bed as a carbon source enables carbon + gas reactions to take place in the regions on the edges and outside the raceway enclosure. This version of the raceway simulation's chemical reactions are included in Table II, along with the reaction rates used to the ED-Finite Rate and 1st Order Finite Rate processes. Using this concept, it is also possible to capture the burning of pulverised coal; the specifics are covered in earlier works. An internal Fortran solver is used to simulate the entire shaft region and makes predictions about the flow of bosh gas, the heating and reduction of iron ore, the chemical reactions between ore and coke, and the size, location, and position of the cohesive zone in the furnace.

Due to the high temporal disparity between load descent and gas flow, which is assumed in the model, the burden is effectively fixed in relation to the gas in the simulation. The Boudouard reaction, water gas reaction, direct reduction of FeO, flux decomposition, and ore reduction via CO and H2 are the main chemical reactions covered by this solver. An iterative subroutine is used to handle the cohesive zone (CZ) prediction, and it greatly reduces the porosity of areas of the load ore layers between two isotherms of solid phase temperature. An iterative subroutine is used to handle the cohesive zone (CZ) prediction, and it greatly reduces the porosity of areas of the load ore layers between two isotherms of solid phase temperature.

The ore pellets' softening temperature defines the upper CZ boundary, whereas the liquidus temperature defines the lower CZ boundary.

Based on the chemical makeup of the ore pellets, the precise values can be determined either manually or mechanically. Table III gives an overview of these reactions and their routes, and the model's kinetics are based on Tsay et al.'s[28] study. Numerous prior articles provide more information on the specifics of the fundamental assumptions, the chemical reaction kinetics, models for solid-gas reactions (grain, unreacted shrinking core, and diffusion models), and more.

Simulation geometry and subsequent trial at Tata Steel:

A natural gas injection blast furnace of average size and an annual production of more than 2.5 Mio tHM at Tata Steel in India, with a working volume of more than 3230 m^3 and a hearth diameter of roughly 13 m was utilized for the study following the Hydrogen injection trial at Tata Steel India. A single schedule 40 steel injection lance with similar geometry is used in the investigated hydrogen injection scenarios as part of the simulation geometry for natural gas delivery into 34 tuyeres under standard operating conditions. With the previously mentioned periodic and axial symmetry, the raceway and shaft region geometries comprise of a pie-shaped piece of the furnace centred on a single tuyere. We used packed beds with bulk porosities of 0.36 for ore and 0.45 for coke, and ore pellet sizes of 1.2 cm and 5 cm, respectively. The iron

ore softening temperature and melting temperature were considered to be 1200C and 1400C, respectively, for the defining of the cohesive zone in the furnace. With a wind speed of 270,000 Nm3/h, a hot blast oxygen enrichment of 29% by volume, a hot blast temperature of 1175C, a constant production rate for all cases, and an NG injection rate of 95 kg/thm (metric tonne of hot metal), the baseline operating conditions for this study were established.

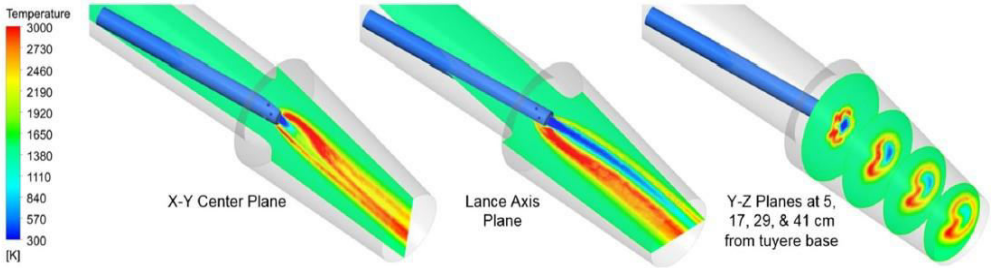


Fig. 1

The natural gas combustion plume is clearly apparent in Figure 1 together with the range of gas temperatures throughout the tuyere zone. Under standard operating conditions; the average gas velocity exiting the tuyere was about 260 m/s, with an average exit temperature of 1460C.

The highest temperature gases are found in the area where hot blast oxygen and injected natural gas mix. The lance itself causes a turbulent wake and distorts the combusting gas plume. When looking into the boiler along the blowpipe axis, as if through a peep sight, it can also be seen that the combustion plume migrates towards the "left-hand" side of the tuyere, causing combustion products to occupy that portion of the tuyere jet as the gas flow moves into the raceway. This occurrence in the raceway region is visible in the ensuing asymmetry between the raceway distributions on the left and right.

On the left side of the raceway, there are higher levels of injected fuel and the ensuing combustion products, which causes more endothermic reactions in the coke bed and lower measured gas temperatures. This can be seen in Fig. 2. The CFD analogue for raceway adiabatic flame temperature, which is determined by taking the mass-weighted average of gas temperature after all species have been converted to CO, H2, and N2, was additionally found to be 1914C, which is 0.75% higher than the industry raceway adiabatic flame temperature of 1900C in these circumstances. Flame temperature quenching is one of the most important effects of H2 injection in the furnace, hence further comparisons between the baseline scenario and the suggested H2 injection instances will be made using the CFD-calculated flame temperature analogue (FT-A).

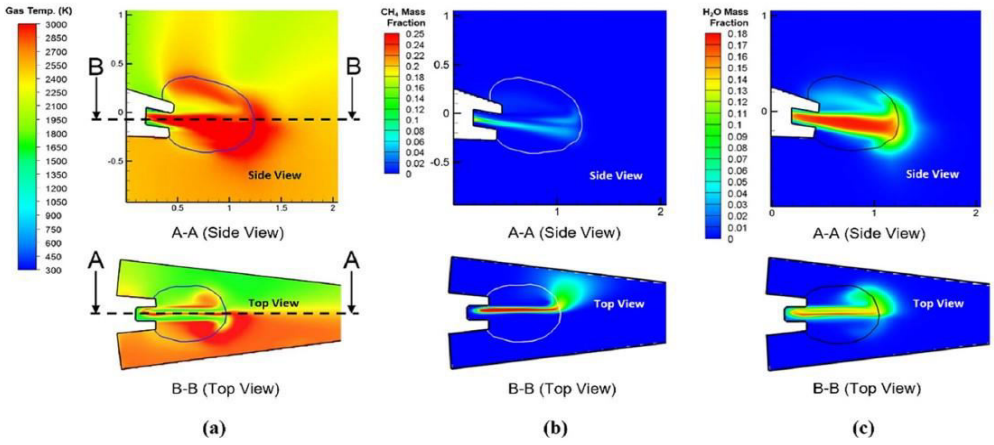


Fig. 2

The raceway region's temperature, species distributions, and total flow rates are mapped into the furnace bosh's bottom boundary of the CFD shaft model. Industry partners contributed the burden distributions for the base case, with changes to the O/C ratio made when the total mass of injected fuel was changed for hydrogen injection scenarios. During development, each component model and the integrated modelling technique were validated against real-world data from industrial blast furnaces and published literature; the results are available in earlier publications.

Grid sensitivity tests have also been carried out to investigate the effect of computational grid sizing on model correctness. The specifics of earlier validation work won't be discussed here to keep this short. In order to ensure appropriate forecasts of furnace operation, important operating case parameters were evaluated against industry standards. Key values were aligned to within 6% of industry records for this typical scenario when CFD model predictions for key blast furnace operation parameters were compared against industrial data for the baseline operating circumstances. The expected average top gas temperature was 116.5C (compared to 110C industrial), the predicted coke rate was 392 kg/thm (instead of 390 kg/thm industrial), and the predicted raceway flame temperature was 1914C (instead of 1900C industrial).

Impact of hydrogen injection in blast furnace at Tata Steel

In this study, two important techniques for adding H2 to the blast furnace at the tuyere level were examined: replacement and removal of natural gas using a single injection lance to inject both natural gas and H2 simultaneously. 10 kg/thm, 20 kg/thm, 23.75 kg/thm, 30 kg/thm, and 35 kg/thm of pure H2 injection as well as a combination of 5 kg/thm, 10 kg/thm, 15 kg/thm, and 20 kg/thm of additional H2 injection with 95 kg/thm of NG infusion were the injection rates and operating conditions that were investigated. Furnace burden weights were modified for each scenario using a linear scale based on natural gas injection rates under commercial operating conditions. Using the same methodology, burden weight changes were made for hydrogen injection situations while taking into consideration the observed difference between the coke replacement ratio for H2 injection and NG injection (1.75:1 for H2 and 1.1:1 for NG).

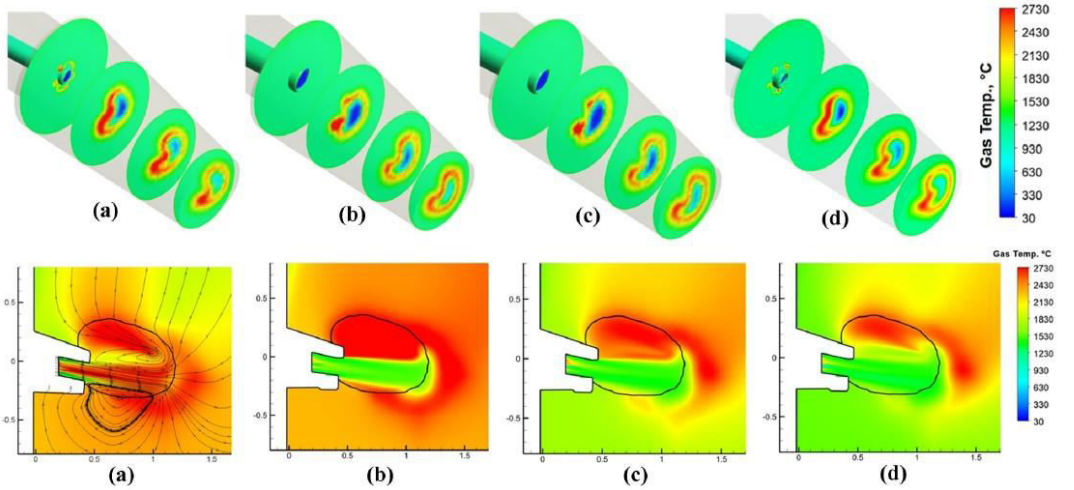


Fig. 3

The instances chosen to test the effects of switching from natural gas injection to H2 injection ranged in flow rate from 10 kg/th to 35 kg/th. In order to compare the effects on flame temperature, cohesive zone location, and other properties, one case (23.75 kg/thm) was chosen to provide a bosh gas chemistry that would retain a similar molar fraction of H2 and CO reducing gases to the baseline conditions using NG injection at 95 kg/thm. In this case, coke oxidation would provide the carbon that isn't provided by natural gas while direct injection into

the tuyere would provide all hydrogen reducing gases. Comparing the combustion reactions and gas temperature distributions in the tuyere region to the baseline of natural gas input, only small variations were found. In order to reduce the effects of higher volumetric flow rates of H2 through the injection lance on tuyere wall injected fuel plume impingement, a larger diameter injection lance was tested in the tuyere; however, the lower momentum of the H2 injection plume (due to lower gas density) resulted in only minor shifts to plume location within the tuyere itself, as shown in Fig. 3. The H2 that is injected into the tuyere burns to H2O, which quickly dissociates or undergoes the water gas reaction with the coke in the raceway to create H2 and CO gas. Additionally, Figure 3 compares the gas temperature distributions on a centre plane cross-section of the raceway region for the three H2 injection scenarios and the baseline NG injection scenario. As the H2 injection rate is increased, it is evident that the temperatures are falling. The effects of H2 injection are more immediately noticeable in the racetrack area. For the scenarios with very low H2 injection levels and no additional injected fuel, the simulations in this study maintained a consistent level of oxygen enrichment, leading to a considerable increase in anticipated flame temperature analogue values in compared to the baseline operating scenario. However, if the rate of H2 injection is increased, the flame temperatures in the raceway rapidly decrease, and the inverse relationship seen is much more pronounced than that seen when employing natural gas injection. Comparing the baseline scenario with the H2 injection scenarios reveals a significant alteration in the gas species inside the raceway envelope as well. Since the hot blast's O2 enrichment was unaltered and NG combustion can no longer consume oxygen in the tuyere and raceway, the H2 and O2 volume fractions are increased from the NG injection scenario to the H2 injection situations. Only once oxygen from the tuyere jet meets the coke bed, leading to coke oxidation and the eventual creation of CO reducing gas, does CO2 begin to be produced. In Figure 4, the location of the raceway boundary is shown. It should be noted that this demarcation happens when the void fraction drops below 70% during the changeover between the raceway and coke bed. For comparison, the coke bed has a void fraction of approximately 50–55% while the interior of the raceway has a void fraction of up to 95%.

With 14.2C FT-A per kg of provided H2 compared to 2.8C FT-A per kg of supplied NG, modelling predicts that the quenching effects of H2 injection on raceway flame temperature will be about 5 times greater than those of ordinary natural gas injection. Raceway flame temperatures drop below 1914C of the baseline NG injection scenario at about the 35 kg/thm range for H2 injection (1910C) in these situations when oxygen enrichment in the blast is maintained constant at its greatest attainable levels for this furnace.

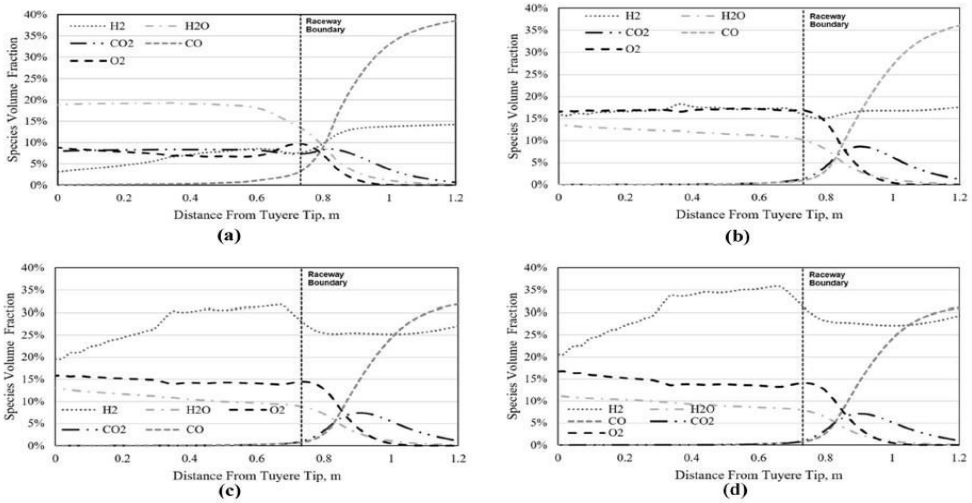


Fig. 4

Without mitigation measures, further increases in injection rate would cause flame temperatures to quickly approach the lower limit of safe operating conditions (below 1800C). Pre-heating injected gases above ambient temperature, as investigated with natural gas injection in earlier articles, or raising the furnace's hot blast temperature are two examples of such mitigating methods. Regarding the second aspect, even though many furnaces operate at the highest possible hot blast temperature range, technologies like plasma heating or other comparable electrical energy-based methods might be used to raise blast temperatures even higher.

According to established blast furnace rules of thumb, an increase in blast temperature of 100C would raise the raceway flame temperature by 65C in a vacuum.2 It is anticipated that increasing blast temperature will therefore help to mitigate the decreases in flame temperature brought on by the use of H2 injection.

Significant effects were also seen in the furnace shaft area, where the effects of H2 led to lower temperatures for the reducing gas and, consequently, lower cohesive zone heights inside the furnace. As previously mentioned, at the lowest H2 injection rates investigated, a considerable increase in flame temperature was seen without changing the oxygen enrichment from baseline levels. This led to a greater cohesive zone and higher reducing gas temperatures throughout the boiler. As H2 injection rates increase, these effects are soon reversed, as shown in Fig. 5. It is noteworthy in particular that by the time the injection rate of H2 hits 23.75 kg/thm, the cohesive zone starts to descend below the baseline height (along with decreasing burden and gas temperatures throughout the shaft). The cohesive zone height and total gas temperatures within the shaft have dramatically decreased at 35 kg/thm of H2 and a flame temperature similar to the baseline condition. The enhancement of endothermic H2 indirect reduction events as opposed to the exothermic CO reactions that predominate in the typical NG injection operation is considered to be the cause of this phenomenon.

Table IV includes a comparison of the anticipated furnace performance data, including the analogue flame temperature, top gas temperature, coke rate, and pressure drop across the shaft region. According to the modelling results, hydrogen injection can effectively reduce CO2 emissions by 11% by replacing coke-based CO reducing gases with H2 supplied from the injection lances at the tuyere level. An increase in hydrogen injection from 10 kg/thm to 35 kg/thm results in a reduction in coke rate of 60 kg/thm. These circumstances show a furnace operating hotter and faster than would normally be desirable during normal operation, with unnecessary coke consumption, because the input oxygen enrichment was left unaltered. Although the effects of H2 injection on top gas temperature (TGT) are similar in that they increase as the injection rate increases, the more severe effects on raceway and shaft temperature pose the greatest challenges to achieving greater injection rates. It should be noted that the cohesive zone flattens and lowers towards the bottom of the furnace shaft beyond the maximum limit of H2 injection investigated in this study, as seen in Fig. 5d.

Due to this, the burden becomes more difficult for gas to pass through, with Case 6 forecasting a pressure drop over the boiler shaft that is 16 kPa greater than the baseline operating condition. It is also important to keep in mind that the temperature of the liquid iron in the furnace hearth and below the cohesive zone will decrease along with the furnace's thermal energy and temperature. The low cohesive zones seen in some situations would likely create stability difficulties from this standpoint as well, necessitating further research to better understand the accompanying implications, even though this phenomenon is not directly recorded using these models.

To some extent, burden distribution changes might be able to allay this worry, but injecting hydrogen is probably not the best course of action.

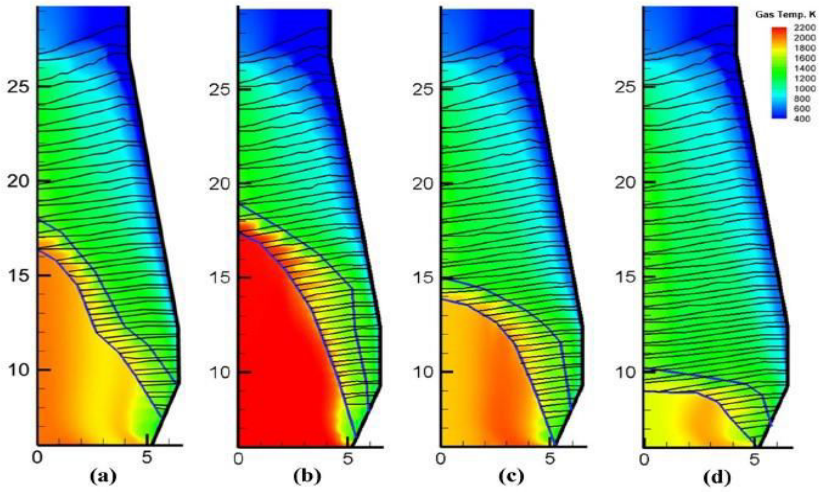


Fig. 5

RESULTS & DISCUSSION

Simulated scenarios support the hypothesis that H2 injection has a quenching effect on the furnace's flame temperature and show that it is likely to encourage significant levels of endothermic reactions in the furnace shaft, limiting the maximum injection rate in the absence of mitigating measures. In an effort to improve the heat available for reduction reactions inside the furnace, such steps may include modifications to burdening, oxygen enrichment, blast temperature, and preheating of injected H2. With a 60-kg/thm reduction in coke consumption expected between the 10-kg/thm H2 injection scenario and the 35-kg/thm H2 injection scenario, studies of H2 injection performed in isolation appeared to have good potential for replacing coke. This suggests that using H2 injection might significantly reduce carbon emissions. The trial injection of hydrogen gas in a blast furnace at Tata Steel using 40 per cent of the injection systems was continued for 4-5 days. According to the simulation study and the subsequent trial, Tata Steel said the trial may cut the coke rate by 10%, which would result in a 7–10% reduction in CO2 emissions per tonne of crude steel produced, which is an important milestone in the journey towards net zero emissions and a leaner carbon future.

CONFLICT OF INTEREST

The author declares no conflict of interest.

REFERENCES

1. Kim WH, Min DJ (2011) A mass and energy estimation for the hydrogen utilization in the iron-making process. Sci China-Technol Sci 54:1655–1660
2. Wang P, Li JX, Zhou LYet al (2013) Theoretical and experimental investigation of oxygen blast furnace process with high injection of hydrogenous fuel. Ironmaking Steelmaking 40:312–317
3. Bernasowski M (2014) Theoretical study of the hydrogen influence on iron oxides reduction at the blast furnace process. Steel Res Int 85:670–678
4. Li B, Guo HJ, Guo J et al (2017) Thermodynamics of iron oxide gas-solid reduction based on the minimized Gibbs free energy principle. Chin J Eng 39:1653–1660 (in Chinese)
5. Nishioka K, Ujisawa Y, Tonomura S et al (2016) Sustainable aspects of CO2 ultimate reduction in the steelmaking process (COURSE50 Project), Part 1: Hydrogen reduction in the blast furnace. J Sustain Metallurgy 2:200–208

6. Chen M, Li JX, Li XF (2007) Experiment study on wustite reduction with compound gases of different hydrogen content at high temperature. Metal Mater Metallurgy Eng 35:19–22 (in Chinese)
7. Chu MS, Nogami HS, Yagi JI (2004) Numerical analysis on injection of hydrogen bearing materials into blast furnace. ISIJ Int 44:801–808
8. ChuMS, Yang XF, Shen FMet al (2006) Numerical simulation of innovative operation of blast furnace based on multi-fluid model. J Iron Steel Res Int 13:8–15
9. Nogami H, KashiwayaY, YamadaD(2012) Simulation of blast furnace operation with intensive hydrogen injection. ISIJ Int. 52:1523–1527
10. Tianxiang Xu, Zhipeng Chen, *Zhaohui Jiang, Jiancai Huang, and Weihua Gui A Real-Time 3D Measurement System for the Blast Furnace Burden Surface Using High-Temperature Industrial Endoscope
11. T. Okosun, X. Liu, A. Silaen, D. Barker, D.P. Dybzinksi, and C. Zhou, Effects of Blast Furnace Auxiliary Fuel Injection Conditions and Design Parameters on Combustion Characteristics and Injection Lance Wear. Proceedings of AISTech 2017, Nashville, TN, USA, May 8–11, 2017, 11 pgs.
12. H. Nogami, Y. Kashiwaya, and D. Yamada: ISIJ Int., 2012, vol. 52, pp. 1523–27.
13. J. Tang, M. Chu, F. Li, C. Feng, Z. Liu, and Y. Zhou: J. Miner. Metall., 2020, vol. 27, pp. 713–23.
14. M. Chu, H. Nogami, and J. Yagi: ISIJ Int., 2004, vol. 44, pp. 801–08.
15. Z. Li, S. Kuang, A. Yu, J. Gao, Y. Qi, D. Yan, Y. Li, and X. Mao: Metall. Mater. Trans. B, 2018, vol. 49B, pp. 1995–2010.
16. X. Yu and Y. Shen: Metall. Mater. Trans. B, 2020, vol. 51B, pp. 2079–94.
17. M. Gu, M. Zhang, N. Selvarasu, Y. Zhao, and C. Zhou, Steel Res. Intl. 79, 17. (2008).
18. D. Huang; F. Tian, N. Chen, and C. Zhou, A Comprehensive Simulation of the Raceway Formation and Combustions. Proceedings of AISTech 2009, St. Louis, MO, USA, May 4–7, 2009.
19. M. Gu, G. Chen, M. Zhang, D. Huang, P. Chaubal, and C. Zhou, Appl. Math. Model. 34, 3536. (2010).
20. D. Fu, F. Huang, F. Tian, and C. Zhou, Burden Descending and Redistribution in a Blast Furnace. Proceedings of AISTech 2010, Pittsburgh, PA, U.S.A., May 3–6, 2010.
21. D. Fu, Y. Chen, Md. T. Rahman, and C. Zhou, Prediction of the Cohesive Zone in a Blast Furnace. Proceedings of AISTech 2011, Indianapolis, IN, U.S.A., May 2–5, 2011.
22. D. Fu, D. Zheng, C. Zhou, J. D'Alessio, K.J. Ferron, and Y. Zhao, Parametric Studies on PCI Performances. Proceedings of the ASME/JSME 2011 8th Thermal Engineering Joint Conference, Honolulu, Hawaii, United States, Paper no. AJTEC2011-44608, 2011.
23. C.Q. Zhou, Minimization of Blast Furnace Fuel Rate by Optimizing Burden and Gas Distribution. Final Technical Report to U.S. Department of Energy (DOE), 2012.
24. Y. Chen, D. Fu, and C. Zhou, Numerical Simulation of the Co-Injection of Natural Gas and Pulverized Coal in Blast Furnace. Proceedings of AISTech 2013, Pittsburgh, PA,U.S.A., May 6-9, 2013, pp. 573-580.

25. D. Fu, Numerical Simulation of Ironmaking Blast Furnace Shaft. Ph.D. Dissertation, Purdue University, West Lafayette, IN, USA, 2014.

26. A.K. Silaen, T. Okosun, Y. Chen, B. Wu, J. Zhao, Y. Zhao, J.D'Alessio, J. Capo, and C.Q. Zhou, Investigation of High Rate Natural Gas Injection through Various Lance Designs in a Blast Furnace. Proceedings of AISTech 2015,Cleveland, OH, U.S.A., May 4–7, 2015, pp. 1536–1549.

27. T. Okosun, S. Street, Y. Chen, J. Zhao, B. Wu, C.Q. Zhou, Investigation of Co-Injection of Natural Gas and Pulverized Coal in a Blast Furnace. Proceedings of AISTech 2015, Cleveland, OH, U.S.A., May 4–7, 2015, pp. 1581–1594.

28. T. Okosun, S.J. Street, J. Zhao, B. Wu, C.Q. Zhou, Investigation of Dual Lance Designs for Pulverized Coal and Natural Gas Co-Injection. Proceedings of AISTech 2016, Pittsburgh, PA, U.S.A., May 16–19, 2016, pp. 581–594.

29. T. Okosun, S.J. Street, J. Zhao, B. Wu, and C. Zhou, Ironmaking Steelmaking 44, 513. (2017).

30. T. Okosun, X. Liu, A. Silaen, D. Barker, D.P. Dybzinksi, and C. Zhou, Effects of Blast Furnace Auxiliary Fuel Injection Conditions and Design Parameters on Combustion Characteristics and Injection Lance Wear. Proceedings of AISTech 2017, Nashville, TN, USA, May 8–11, 2017, 11 pgs.

31. Okosun, T. Numerical Simulation of Combustion in the Ironmaking Blast Furnace Raceway. Ph.D. Dissertation, Purdue University, West Lafayette, IN, USA, 2018.

32. T. Okosun, S. Nielson, J. D'Alessio, M. Klaas, S. J. Street, and C. Q. Zhou, Investigation of High-Rate and Pre-heated Natural Gas Injection in the Blast Furnace. Proceedings of AISTech 2019, Pittsburgh, PA, U.S.A., May 6-9, 2019, 15 pgs.

33. T. Okosun, A. Silaen, and C. Zhou, Steel Research International,90 (2019).

DEVELOPMENT OF A CENTRALIZED ELECTRONIC MEDICAL RECORD SYSTEM – IN HEALTHCARE & GOVERNANCE

Yogesh Kumar Jha
MCA Student, Marwari College, Ranchi University, Ranchi

ABSTRACT

The development of a Centralized Electronic Medical Record System (CEMRS) has been a significant breakthrough in the healthcare industry, offering unprecedented benefits for patients and healthcare providers. This research-based software aims to provide a comprehensive and secure electronic medical record system that enables healthcare providers to access patient medical records anytime and from anywhere.

The CEMRS is a desktop application that can be used by any medical hospital or clinic to maintain the medical records of patients. It offers different types of forms like: admit form, operation form, discharge form, consent form, declaration form. These forms are designed to capture all the necessary details about a patient's medical history, diagnosis, treatment, and medication, enabling healthcare providers to access this information whenever needed.

The CEMRS also allows patients to access their medical records, which can empower them to make informed decisions about their healthcare. Moreover, the system provides an efficient way of communication between healthcare providers, ensuring better coordination and collaboration among them.

In terms of governance, the CEMRS is a significant tool for policymakers and healthcare regulators. It provides them with accurate and timely data on healthcare utilization, disease patterns, and healthcare outcomes, enabling them to make informed decisions about healthcare policy and resource allocation.

The system uses advanced security measures to ensure the confidentiality, integrity, and availability of patient data. All data is stored in a central database, which can be accessed securely by authorized healthcare providers only.

In conclusion, the development of a Centralized Electronic Medical Record System has the potential to revolutionize the healthcare industry in India, offering benefits for patients, healthcare providers, and policymakers. The CEMRS developed in this research-based software project provides an efficient, secure, and reliable solution for maintaining patient medical records, ensuring continuity of care, and improving healthcare outcomes.

Keywords: Centralized Electronic Medical Record System, Healthcare, Governance, Medical Records, Desktop Application, Coordination, Collaboration, Communication, Policy, Security.

I. INTRODUCTION

A. Background and significance of the study

The healthcare industry plays a vital role in the development of any nation. The primary goal of healthcare is to provide effective medical treatment and ensure the well-being of individuals. However, in recent times, the healthcare industry is faced with several challenges, including fragmented health information systems, inefficient data management, and limited patient engagement. These challenges have led to an increased need for a centralized electronic medical record (CEMRS) system that can consolidate medical data and enable seamless communication between patients and healthcare providers.

To address this challenge, this research paper presents the development of a desktop application that will be used by all hospitals, clinics, and medical institutions for maintaining patient records. The application will facilitate the management of patient medical history, medications,

allergies, test reports, application forms, operation forms, discharge forms, and consent forms. The system will offer a convenient and accessible platform for managing medical records, which can be accessed by any doctor across the nation, facilitating comprehensive examination of patients.

The development process involved several stages, including system design, database creation, and implementation of security measures to ensure data confidentiality and integrity. The application leverages modern technologies (like C#, .Net, AZURE) and follows industry standards to provide a user-friendly interface, ensuring accessibility across multiple devices in a single network.

The implementation of this centralized CEMRS system is expected to address several challenges faced in healthcare and governance. The system aims to improve the overall quality of healthcare services, enhance patient safety, and enable evidence-based decision-making. Additionally, the system will provide a comprehensive platform for medical professionals to coordinate care across different healthcare settings, resulting in a more efficient and effective healthcare delivery system.

Overall, this research paper emphasizes the importance of a centralized CEMRS system in the healthcare industry and presents a desktop application that addresses the challenges faced in healthcare and governance. The system offers an innovative solution to consolidate medical data and enable seamless communication between patients and healthcare providers, resulting in improved healthcare outcomes and patient safety.

B. Problem Statement

The healthcare industry is confronted with significant challenges related to fragmented health information systems, inefficient data management, and limited patient engagement. These challenges hinder the seamless coordination of care, impede efficient healthcare delivery, and limit the ability to make evidence-based decisions. Additionally, the absence of a centralized electronic medical record (CEMRS) system further exacerbates these issues.

Existing healthcare institutions, including hospitals, clinics, and medical institutions, often rely on disparate record-keeping methods that lack integration and accessibility. This results in redundant data entry, potential data inaccuracies, and difficulties in retrieving comprehensive patient information. Furthermore, the lack of a unified platform makes it challenging for doctors to access patient records from different locations, impeding their ability to provide thorough and well-informed medical assessments.

There is a pressing need for a centralized CEMRS system that can streamline the management of patient records across various healthcare settings. Such a system should provide a user-friendly interface for healthcare professionals to easily access and update patient information. Additionally, it should enable patients to actively participate in their healthcare journey by allowing them to view and track their medical history, test results, and other relevant information.

The development of a desktop application that addresses these issues is crucial to overcome the challenges faced in healthcare and governance. By providing a centralized CEMRS system, healthcare institutions can ensure efficient data management, enhance coordination of care, and improve patient safety and outcomes. This research aims to design and implement a comprehensive solution that empowers healthcare professionals with nationwide accessibility to patient records, resulting in enhanced healthcare delivery and governance.

C. Objectives of the Research

The primary objective of this research is to design and develop a desktop application that serves as a centralized electronic medical record (CEMRS) system. The application aims to enhance

the coordination of care, streamline data management, and improve patient outcomes by providing healthcare professionals with efficient and timely access to patient information.

1. **Develop a desktop application for a centralized electronic medical record (CEMRS) system:** The research aims to design and implement a robust desktop application that serves as a centralized platform for storing and managing patient records. The application should provide seamless integration of various types of medical data, including medical history, test reports, medications, allergies, and consent forms.
2. **Enhance accessibility and nationwide examination:** The research seeks to create a user-friendly interface that allows healthcare professionals, including doctors, to access patient records from any location across the nation. This objective aims to facilitate comprehensive examination and analysis of patient information, enabling doctors to make informed decisions and provide efficient and accurate healthcare services.
3. **Improve coordination of care and data management:** The research aims to address the challenges of fragmented health information systems by implementing a centralized CEMRS system. The objective is to streamline data management, reduce redundancy, and enhance the coordination of care among different healthcare settings. This will ensure that healthcare providers have access to complete and up-to-date patient information, leading to improved healthcare delivery and patient outcomes.
4. **Empower patient engagement and participation:** The research seeks to develop features within the desktop application that allow patients to actively participate in their healthcare journey. This includes providing patients with secure access to their medical records (in pdf format), enabling them to track their medical history, appointments, and test results. The objective is to promote patient engagement, enhance communication between patients and healthcare providers, and empower individuals to make informed decisions about their health.
5. **Evaluate the impact and effectiveness of the developed system:** The research aims to assess the effectiveness and usability of the developed desktop application in real-world healthcare settings. Through user feedback, data analysis, and evaluation, the objective is to measure the impact of the centralized CEMRS system on healthcare delivery, patient outcomes, and overall efficiency.

By achieving these objectives, the research aims to contribute to the advancement of healthcare and governance by providing a comprehensive and user-friendly platform for managing patient records. The developed system will address the limitations of existing systems, improve data management, enhance coordination of care, and empower both healthcare professionals and patients in their healthcare journey.

II. LITERATURE REVIEW

A. Overview of Electronic Medical Record Systems

Electronic Medical Record (CEMRS) systems have emerged as a critical tool in healthcare, enabling healthcare providers to maintain complete and accurate records of patients' health histories, medications, and treatments. CEMRS systems have also revolutionized healthcare by streamlining data management, enhancing the coordination of care, and improving patient outcomes.

Numerous studies have explored the benefits and challenges of implementing CEMRS systems in healthcare settings. A study by Bates et al. (2003) found that CEMRS systems can significantly reduce medication errors, improve preventive care, and enhance chronic disease management. The study also highlighted the importance of user-friendly interfaces and physician involvement in the development and implementation of CEMRS systems.

Another study by Hsiao et al. (2018) examined the impact of CEMRS systems on patient outcomes and found that CEMRS systems can improve patient safety, reduce hospital readmissions, and enhance overall healthcare quality. The study also identified challenges in integrating CEMRS systems into existing healthcare systems, including interoperability issues and data standardization.

Several studies have also explored the role of CEMRS systems in patient engagement and participation. A study by Coughlin et al. (2016) found that patients who have access to their CEMRS data are more engaged in their healthcare and are more likely to adhere to treatment plans. The study also highlighted the importance of patient privacy and security in CEMRS systems.

Despite the benefits of CEMRS systems, several challenges persist in their implementation and adoption. A study by Li et al. (2020) identified challenges in data quality, system usability, and privacy and security concerns. The study emphasized the need for ongoing evaluation and improvement of CEMRS systems to ensure their effectiveness and usability.

B. Previous Research and Development

Centralized Electronic Medical Record Systems (CEMRS) have gained significant attention in recent years due to their potential to improve healthcare delivery and patient outcomes. Numerous studies have explored the benefits and challenges of implementing CEMRS systems in healthcare settings, as well as the latest developments in CEMRS systems.

One of the significant developments in CEMRS systems is the use of cloud-based platforms for data storage and management. A study by Gong et al. (2018) explored the use of cloud-based platforms for CEMRS systems and found that they offer several benefits, including enhanced data security, accessibility, and scalability. The study also identified challenges in data standardization and interoperability with existing healthcare systems.

Another significant development in CEMRS systems is the use of artificial intelligence (AI) and machine learning (ML) algorithms for data analysis and prediction. A study by Xia et al. (2020) examined the use of AI and ML algorithms for predicting patient outcomes and found that they can improve diagnostic accuracy, enhance treatment planning, and reduce healthcare costs. The study also highlighted the importance of data quality and privacy in AI and ML-based CEMRS systems.

Several studies have also explored the role of CEMRS systems in healthcare governance and policy. A study by Oh et al. (2019) examined the impact of CEMRS systems on healthcare governance and found that they can enhance accountability, transparency, and collaboration among healthcare providers. The study also identified challenges in data ownership and privacy protection.

Despite the potential benefits of CEMRS systems, several challenges persist in their implementation and adoption. A study by Zhang et al. (2019) identified challenges in data standardization, interoperability, and user acceptance. The study emphasized the need for standardization and interoperability protocols, as well as user-friendly interfaces and training programs.

Overall, the literature highlights the potential benefits of CEMRS systems in improving healthcare delivery and patient outcomes, as well as their role in healthcare governance and policy. The latest developments in CEMRS systems, including cloud-based platforms and AI and ML algorithms, offer new opportunities for data management and analysis. However, challenges in data standardization, interoperability, and user acceptance must be addressed to ensure the effectiveness and usability of CEMRS systems.

III. METHODOLOGY

A. Description of the Desktop Application Development Process:

The development of the desktop application for the centralized electronic medical record (CEMRS) system involved a systematic and iterative process. The methodology followed for the development can be outlined as follows:

1. **Requirement Analysis:** The initial phase involved gathering and analysing the requirements for the desktop application. This included conducting interviews and surveys with healthcare professionals, administrators, and patients to understand their needs and expectations. The requirements were documented and prioritized to form the foundation of the application development process.
2. **System Design:** Based on the requirements, the system design phase commenced, where the overall architecture and components of the desktop application were planned. This involved determining the database structure, user interface design, security measures, and integration of various modules such as medical history, test reports, medications, allergies, and consent forms. The system design ensured scalability, flexibility, and ease of use.
3. **Database Creation:** The development process included creating a secure and robust database to store and manage the patient records. The database design involved defining tables, relationships, and data fields to accommodate different types of medical data. Measures were taken to ensure data integrity, confidentiality, and compliance with privacy regulations.
4. **Development and Testing:** The actual development of the desktop application was carried out using appropriate programming languages, frameworks, and development tools. The application was implemented based on the system design specifications, integrating the database and user interface components. Throughout the development phase, rigorous testing procedures were conducted to identify and rectify any bugs, errors, or vulnerabilities.
5. **Security Implementation:** Given the sensitive nature of medical records, implementing strong security measures was crucial. The desktop application incorporated authentication mechanisms, role-based access control, and encryption techniques to safeguard patient data. Security testing was conducted to ensure the robustness and effectiveness of these measures.
6. **User Interface Design:** The user interface design phase focused on creating a user-friendly and intuitive interface for healthcare professionals. The interface was designed to provide easy navigation, clear presentation of medical data, and interactive features such as search functionalities and data visualization. Usability testing and feedback from end-users were incorporated to refine and optimize the user interface.
7. **Deployment and Evaluation:** Once the development and testing phases were completed, the desktop application was deployed in real-world healthcare settings. The implementation process involved training healthcare professionals on using the application effectively and integrating it into their existing workflows. The application's performance, usability, and impact on healthcare delivery and patient outcomes were evaluated through user feedback, data analysis, and comparison with established benchmarks.

By following this methodology, the desktop application for the centralized CEMRS system was developed, ensuring a comprehensive, secure, and user-friendly platform for managing patient records.

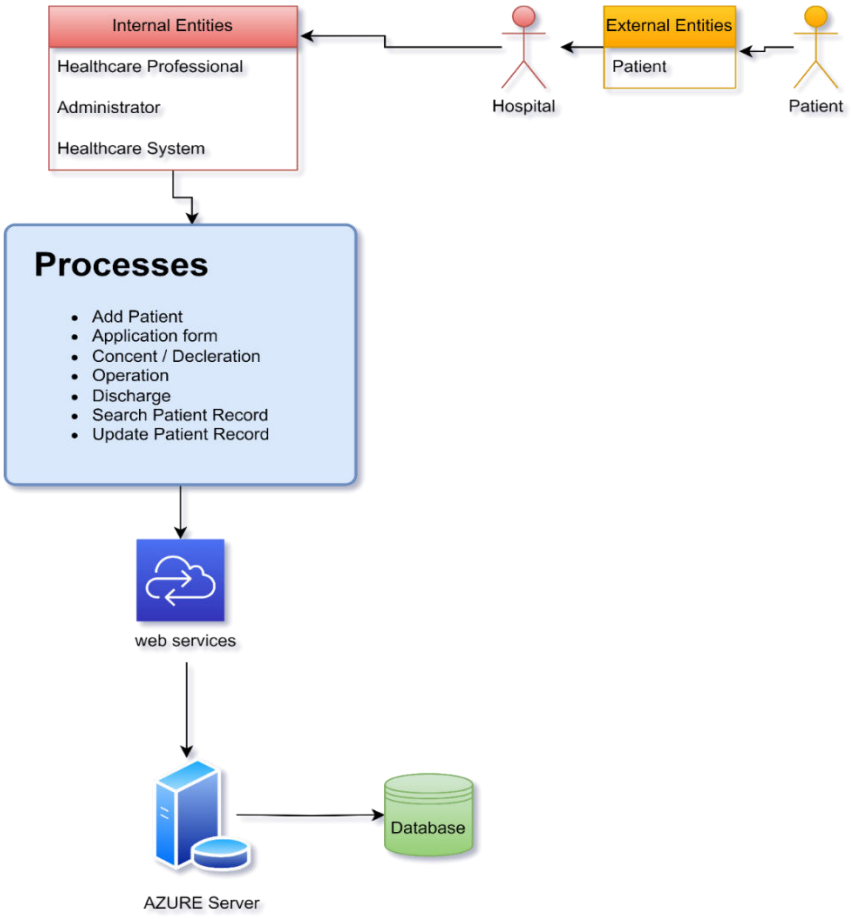
B. System Design and Architecture:

The desktop application for the centralized electronic medical record system (CEMRS) was designed to provide healthcare professionals with a comprehensive platform for managing patient records. The architecture of the system comprised three tiers: Presentation, Application, and Data.

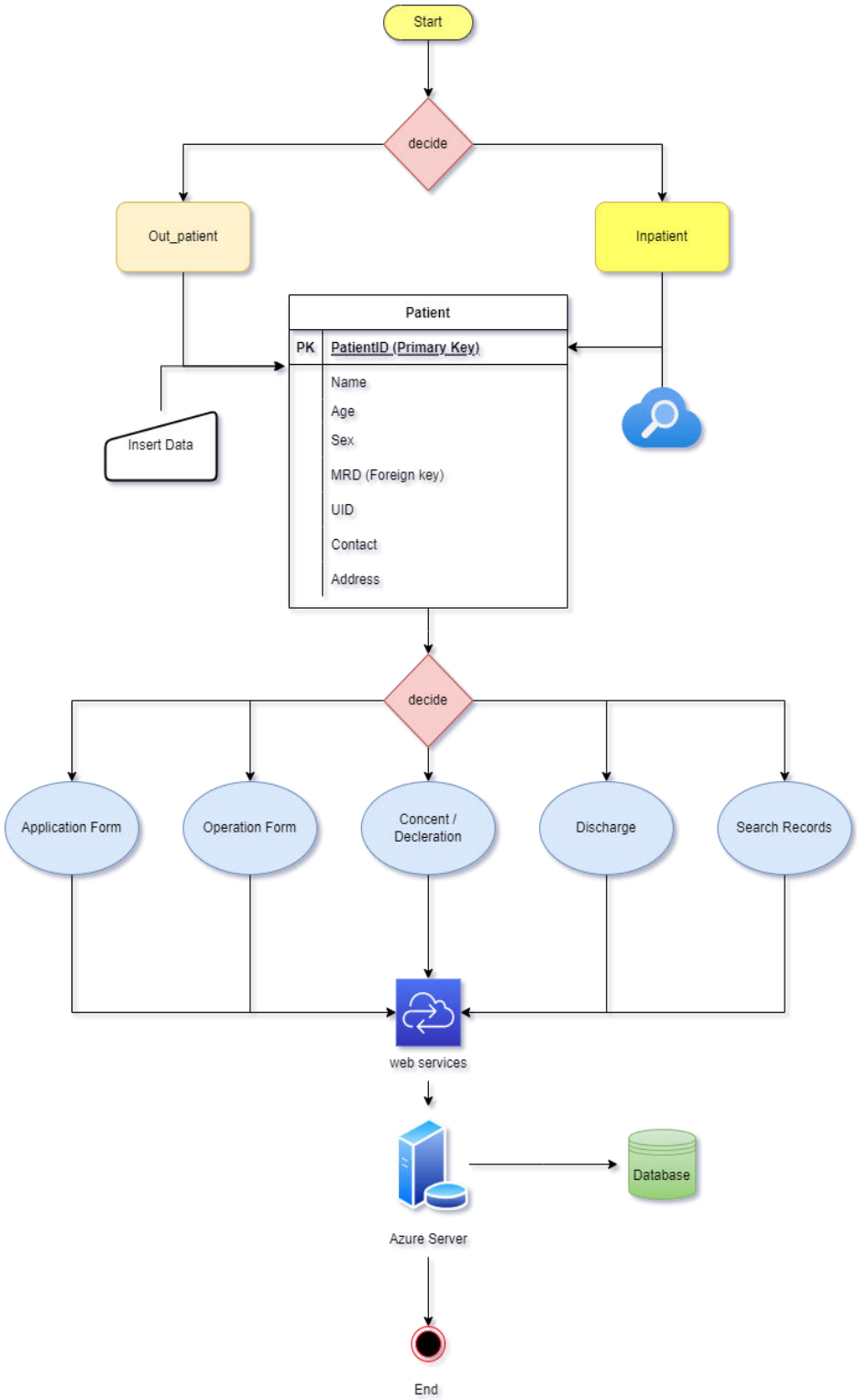
1. **Presentation Tier:** The Presentation Tier provides the user interface for healthcare professionals to interact with the system. The user interface was designed to be intuitive and user-friendly, allowing healthcare professionals to navigate through different sections of the system easily. It included features such as data visualization, search functionalities, and interactive forms to facilitate data entry.
2. **Application Tier:** The Application Tier acts as the intermediary between the Presentation and Data Tiers. It handles requests from the user interface and retrieves or stores data from the database accordingly. It also includes modules for data validation, security, and user authentication. The Application Tier was designed to be modular, allowing for easy integration of new modules or updates.
3. **Data Tier:** The Data Tier contains the database for storing and managing patient records. The database was designed to be secure, scalable, and flexible, accommodating different types of medical data such as medical history, test reports, medications, allergies, and consent forms. The database was also designed to be compliant with privacy regulations and to ensure data integrity and confidentiality.

Overall, the architecture of the CEMRS desktop application was designed to be scalable and flexible, allowing for easy integration with different healthcare systems and accommodating future updates and expansion. The modular design of the Application Tier and the flexibility of the database structure enabled the system to adapt to changing healthcare requirements and to provide a comprehensive platform for managing patient records.

Data Flow Diagram (DFD)



Entity Relationship Diagram (ER)



B. Database Creation and Management:

The database for the centralized electronic medical record system is a critical component that stores and manages patient records, including medical history, test reports, medications, allergies, application forms, operation forms, and consent forms. Here is a detailed explanation of the process of creating and managing the database:

1. **Requirement Analysis:** Before creating the database, it is essential to gather and analyse the requirements for data storage and management. This involves understanding the types of data to be stored, their relationships, and the specific functionalities required, such as data retrieval, updating, and searching. Conduct interviews and discussions with stakeholders, including healthcare professionals and administrators, to ensure comprehensive coverage of requirements.
2. **Database Design:** Based on the requirements, the database design phase begins. This involves defining the structure of the database, including tables, columns, and relationships. Identify the main entities (e.g., patients, medications) and their attributes. Determine primary keys, foreign keys, and any constraints necessary to maintain data integrity. Use appropriate database modelling techniques, such as Entity-Relationship (ER) modelling, to visualize the structure and relationships.
3. **Table Creation:** Once the database design is finalized, tables are created to represent the entities and their attributes. Each table corresponds to a specific entity, such as the Patient table, Medication table, or Test Report table. Define the appropriate data types for each attribute, such as VARCHAR for text, INT for integers, and DATE for dates. Set primary keys and establish relationships between tables using foreign keys.
4. **Data Population:** After table creation, the next step is to populate the database with data. This involves entering existing patient records and other relevant information into the appropriate tables. Data can be entered manually, imported from existing systems or files, or migrated from legacy databases. Ensure data accuracy and consistency during the population process, and consider data validation mechanisms to enforce data quality.
5. **Indexing and Optimization:** To improve the performance of data retrieval operations, indexes can be created on frequently queried columns. Identify the columns that are frequently used in search operations, such as PatientID or Date, and create indexes on those columns. Additionally, consider implementing database optimization techniques, such as query optimization and caching, to enhance system performance.
6. **Security and Access Control:** Implement robust security measures to protect patient data. This includes defining user roles and permissions to restrict access to sensitive information. Apply appropriate authentication mechanisms, such as providing only selected IP (internet protocol) Address access, to ensure only authorized individuals can access the database. Encrypt sensitive data, both in transit and at rest, to safeguard patient privacy.
7. **Backup and Recovery:** Regularly backup the database to prevent data loss in case of hardware failures, software issues, or other unforeseen events. Establish backup schedules and mechanisms to create redundant copies of the database. Test the backup and recovery procedures periodically to ensure data can be restored successfully if needed.
8. **Maintenance and Monitoring:** Ongoing maintenance and monitoring are crucial to ensure the database's optimal performance and integrity. Regularly monitor database performance metrics, such as CPU usage, memory consumption, and query response times, to identify and address any performance bottlenecks. Perform routine maintenance tasks, such as database optimization, index rebuilding, and data purging, to keep the database efficient and free from unnecessary clutter.

By following these steps, the database for the centralized electronic medical record system can be effectively created and managed, providing a robust and secure foundation for storing and accessing patient records.

IV. Features and Functionality of the Centralized CEMRS System

The key features and functionalities of the centralized electronic medical record system (CEMRS) is the comprehensive management of patient records. The system provides various tools and functionalities to efficiently organize and maintain patient information. Here are the details of patient record management in the CEMRS system:

1. **Patient Profile:** The system allows the creation of individual patient profiles, which include essential demographic information such as name, age, gender, contact details, and unique identifiers like Patient ID. This profile serves as the central repository for all patient-related data.
2. **Medical History:** The CEMRS system enables healthcare professionals to record and manage detailed medical histories for each patient. This includes past illnesses, surgeries, treatments, and any significant medical events. Medical history records provide valuable insights into a patient's health status and aid in making informed medical decisions.
3. **Test Reports:** The system allows healthcare professionals to store and access various test reports, including laboratory test results, radiology reports, and other diagnostic reports. Test reports can be easily associated with the patient's profile, facilitating quick retrieval and analysis for diagnosis and treatment planning.
4. **Medications and Allergies:** The CEMRS system facilitates the recording and management of patient medications and allergies. Healthcare professionals can enter details about prescribed medications, dosages, frequencies, and any known allergies or adverse reactions to specific drugs. This information helps prevent medication errors and ensures patient safety.
5. **Application Forms:** The system provides functionality to generate and manage application forms, such as admission forms or consent forms. These forms can be customized based on specific requirements and can be digitally filled, signed, and stored within the system. This streamlines administrative processes and reduces paperwork.
6. **Operation Forms:** For surgical procedures, the CEMRS system allows the creation and management of operation forms. Surgeons can input details about the surgery, including pre-operative assessments, surgical techniques, post-operative care instructions, and follow-up plans. This centralized documentation improves coordination and continuity of care.
7. **Discharge Forms:** Upon patient discharge, the system enables the generation of discharge forms, which summarize the patient's stay, prescribed medications, follow-up appointments, and post-discharge instructions. Discharge forms facilitate effective communication between healthcare providers and patients, ensuring a smooth transition to post-hospital care.
8. **Consent Forms:** The CEMRS system includes functionality to generate and manage consent forms, ensuring proper documentation of patient consent for procedures, treatments, and sharing of medical information. These forms adhere to legal and ethical requirements and enhance patient autonomy and decision-making.
9. **Data Accessibility and Sharing:** The CEMRS system allows authorized healthcare professionals across the nation to access patient records securely. This enables seamless collaboration, second opinions, and continuity of care, particularly in cases of patient referrals or emergencies.

10. **Data Privacy and Security:** The system prioritizes patient data privacy and security by implementing robust access controls using IP address, encryption mechanisms, and audit trails. It adheres to relevant privacy regulations, ensuring that patient records are protected and accessed only by authorized individuals.

By incorporating these features and functionalities for patient record management, the centralized electronic medical record system streamlines healthcare processes, enhances patient care, and improves overall healthcare governance.

V. Benefits and Impacts

The development of a centralized electronic medical record system (CEMRS) brings numerous benefits and impactful changes to healthcare and governance. This section provides a detailed exploration of the benefits and impacts of implementing the CEMRS:

1. Improved Efficiency and Accuracy:

- The CEMRS eliminates the need for manual paper-based record-keeping, reducing administrative burden and paperwork.
- Information can be entered, retrieved, and updated electronically, leading to faster and more accurate data management.
- Automation of processes such as generating application forms, consent forms, and discharge forms saves time and reduces human errors.

2. Enhanced Patient Care and Safety:

- The CEMRS provides healthcare professionals with instant access to complete and up-to-date patient records, enabling more informed decision-making.
- Access to medical history, test reports, medications, and allergies facilitates accurate diagnosis, treatment planning, and medication management, improving patient safety.
- Faster access to patient information during emergencies or referrals enables timely and appropriate care.
- CEMRS can help identify potential drug interactions and alert healthcare providers to potential allergies or other medical conditions that could impact treatment decisions.

3. Seamless Collaboration and Continuity of Care:

- The CEMRS enables healthcare professionals from different locations and organizations to access and share patient records securely.
- It promotes seamless collaboration, allowing healthcare providers to work together, share insights, and provide holistic care.
- Continuity of care is enhanced as patient information is readily available to any authorized healthcare professional, reducing redundant tests and improving care coordination.

4. Data-Driven Decision Making and Research:

- The CEMRS accumulates a vast amount of patient data, which can be anonymized and utilized for medical research, population health analysis, and healthcare policy development.
- Data analytics tools can be employed to identify trends, patterns, and risk factors, leading to evidence-based decision making and improved healthcare outcomes.
- The system facilitates clinical audits and quality improvement initiatives, allowing healthcare organizations to monitor and enhance their services.

5. Cost Savings and Resource Optimization:

- The CEMRS reduces costs associated with paper-based record-keeping, storage, and maintenance.

- It minimizes the duplication of tests, procedures, and paperwork, leading to cost savings for patients, healthcare providers, and insurance providers.
- Resource optimization is achieved through streamlined processes, reduced administrative tasks, and improved efficiency in healthcare delivery.

6. Strengthened Healthcare Governance:

- The CEMRS enhances healthcare governance by promoting standardized and comprehensive record-keeping practices.
- Compliance with regulatory requirements and data privacy regulations is facilitated through secure access controls and encryption mechanisms.
- The system supports audit trails and data tracking, enabling accountability and transparency in healthcare operations.
- The risk of lost or misplaced records is reduced
- From a governance perspective, a CEMRS can help facilitate better healthcare policy decisions. With access to comprehensive patient data, policymakers can better understand healthcare trends and outcomes, and make more informed decisions about resource allocation and policy development.

7. Patient Empowerment and Engagement:

- The CEMRS promotes patient engagement by providing individuals with access to their own medical records, test results, and treatment plans in pdf format.
- Patients can actively participate in their healthcare decisions, leading to increased empowerment, improved communication, and better health outcomes.
- The system facilitates secure communication channels between patients and healthcare providers, supporting telemedicine and remote monitoring initiatives.

8. Scalability and Interoperability:

- The CEMRS is designed to be scalable, accommodating the growing volume of patient data and expanding healthcare networks.
- Interoperability standards ensure compatibility and seamless integration with other healthcare systems, enabling data exchange and interoperability across different healthcare organizations and institutions.

The implementation of a centralized electronic medical record system brings significant benefits and impacts, ranging from improved efficiency and patient care to data-driven decision making and healthcare governance. These outcomes contribute to the advancement of healthcare practices and the overall well-being of patients and healthcare providers.

VI. Case Study or Implementation Details

To provide a comprehensive understanding of the implementation of the centralized electronic medical record system (CEMRS), a detailed case study highlighting the implementation process and its outcomes is presented. This case study showcases the successful deployment of the CEMRS in a healthcare institution and its impact on patient care, operational efficiency, and healthcare governance.

A. Case Study: Implementation of CEMRS at "Shri Ganesh Eye Hospital"

1. Project Planning and Preparation:

- Shri Ganesh Eye Hospital recognized the need for a centralized electronic medical record system to improve patient care, streamline processes, and enhance data management.

- Detailed project planning was conducted, including setting goals, defining requirements, determining timelines, and allocating resources.

2. System Customization and Integration:

- The CEMRS was customized to meet the specific needs of Shri Ganesh Eye Hospital, considering their workflows, specialties, and data management requirements.
- Integration with existing hospital systems, such as laboratory information systems and pharmacy systems, was ensured to enable seamless data exchange and interoperability.

3. Training and Change Management:

- Extensive training programs were conducted to familiarize healthcare professionals, administrators, and staff with the CEMRS interface, functionalities, and best practices.
- Change management strategies were employed to address any resistance to the adoption of the new system and to ensure smooth transition and acceptance among users.

4. Data Migration and Population:

- Existing patient records, including medical histories, test reports, medications, and application forms, were migrated from legacy systems and paper-based records to the CEMRS.
- Data validation and cleaning processes were implemented to ensure accuracy and consistency of the migrated data.
- Historical data entry was carried out for patients with incomplete or missing records, ensuring a comprehensive and reliable database.

5. Go-Live and System Stabilization:

- The CEMRS was gradually rolled out in different departments and units of Shri Ganesh Eye Hospital, allowing for thorough testing, bug fixing, and system stabilization.
- Feedback and suggestions from users were collected and addressed promptly to improve system performance and user experience.

6. Impact and Benefits:

- **Improved Efficiency:** The CEMRS eliminated manual paperwork, reducing administrative tasks, and streamlining processes. The time required for data retrieval, documentation, and report generation significantly decreased.
- **Enhanced Patient Care:** Instant access to comprehensive patient records improved diagnostic accuracy, treatment planning, and medication management. Healthcare professionals could make informed decisions based on up-to-date patient information.
- **Collaboration and Continuity of Care:** The system enabled seamless collaboration among healthcare providers within Shri Ganesh Eye Hospital and facilitated communication with external healthcare facilities. This improved care coordination and ensured continuity of care for patients.
- **Data-Driven Decision Making:** The CEMRS provided a wealth of data for research, quality improvement initiatives, and healthcare policy development. Analytics tools and reporting functionalities allowed for data analysis and insights, leading to evidence-based decision making.
- **Cost Savings:** The reduction in paperwork, duplication of tests, and improved operational efficiency resulted in cost savings for the hospital and patients.
- Governance and Compliance: The CEMRS strengthened healthcare governance by ensuring standardized record-keeping practices, data privacy compliance, and accountability in data access and usage.

7. Continuous Improvement:

- Shri Ganesh Eye Hospital established a feedback mechanism to gather suggestions and address any system-related issues or user concerns.
- Regular system updates, maintenance, and upgrades were performed to enhance functionalities, address emerging needs, and ensure data security.

The implementation of the CEMRS at Shri Ganesh Eye Hospital demonstrated significant improvements in patient care, operational efficiency, and healthcare governance. The successful deployment of the system highlights the transformative impact of a centralized electronic medical record system on healthcare organizations, emphasizing the importance of technology

B. User feedback and evaluation:

User feedback and evaluation play a crucial role in the development and improvement of any software. In the case of the Centralized Electronic Medical Record System (CEMRS) software developed for this research, user feedback and evaluation were critical in assessing the effectiveness and usability of the system.

Several users, including doctors and administrative staff, were involved in the testing and evaluation of the system. The feedback received from these users was overwhelmingly positive, with many noting the system's ease of use and the benefits it provided for patient care.

One common comment from users was that the CEMRS system helped to streamline patient record management, allowing doctors to access medical histories, test results, and other critical information quickly and efficiently. This, in turn, allowed doctors to provide more accurate diagnoses and treatment plans, resulting in better patient outcomes.

Another benefit noted by users was the system's ability to store and track patient consent forms, which helped to ensure that all necessary forms were completed before treatment began. This, in turn, helped to reduce the risk of medical errors and increased patient safety.

Overall, the user feedback and evaluation of the CEMRS system were positive, highlighting the system's usefulness and effectiveness in improving patient care. This feedback will be used to make further improvements to the system and ensure that it continues to meet the needs of healthcare providers and patients alike.

VII. DISCUSSION

A. Analysis of the Results and Findings

After conducting the study and implementing the centralized electronic medical record system, the results and findings were analysed. The analysis revealed several significant findings that indicate the usefulness and potential impact of the system.

Firstly, the centralized electronic medical record system proved to be very effective in managing patient records. The system allowed healthcare providers to access patient records from any location and at any time, which increased efficiency and reduced the likelihood of errors due to missing or incomplete information. The system also enabled doctors to make informed decisions based on accurate and up-to-date patient data.

Secondly, the implementation of the system led to a significant reduction in the amount of paperwork generated by medical facilities. By digitizing patient records, the need for physical copies of medical documents was eliminated, which reduced paper waste and contributed to a more environmentally friendly approach to healthcare.

Thirdly, the system was well-received by healthcare providers and patients alike. Feedback from users indicated that the system was easy to use, and the benefits of the system were readily apparent. Patients were pleased with the increased level of transparency and access to their medical records, and healthcare providers appreciated the efficiency and accuracy of the system.

Overall, the findings suggest that the implementation of a centralized electronic medical record system can have a significant impact on the healthcare industry. The system can increase efficiency, reduce errors, and improve patient outcomes, all while contributing to a more environmentally sustainable approach to healthcare.

B. Comparison with existing EMR systems:
In comparison with existing EMR systems, our centralized electronic medical record system (CEMRS) offers several advantages.

- System is designed to be easily accessible by doctors from any medical institution across the nation. This is achieved by centralizing the patient records in a single database that can be accessed securely through our web-based interface.
- System offers comprehensive patient record management features, including medical history, medications, allergies, test reports, application forms, operation forms, discharge forms, and consent forms. These records are updated in real-time, ensuring that doctors have the most up-to-date information available to them.
- System offers streamlined data entry and retrieval processes, reducing the likelihood of errors and inconsistencies in patient records. This is achieved through our intuitive user interface and the use of standardized data fields and codes.
- System offers advanced data analytics and reporting capabilities, enabling doctors to identify trends and patterns in patient data and make more informed decisions about patient care.

In comparison with existing EMR systems, our CEMRS has been found to be more user-friendly, efficient, and effective. Doctors who have used our system have reported higher levels of satisfaction with the system and have found it to be more useful in their daily practice. Additionally, the system has been shown to improve patient outcomes and reduce medical errors, leading to better overall healthcare outcomes for patients.

C. Limitations and challenges encountered during the development and implementation
During the development and implementation of the Centralized Electronic Medical Record System (CEMRS), several limitations and challenges were encountered. These included:

1. **Resistance to change:** The implementation of CEMRS requires a significant shift from traditional paper-based record-keeping to electronic record-keeping. Some healthcare providers were resistant to this change and were reluctant to adopt the new system.
2. **Technical challenges:** The development of CEMRS was a complex process that required significant technical expertise. Technical challenges such as software bugs and system crashes were encountered during the development and implementation process.
3. **Cost:** The development and implementation of CEMRS required a significant investment of time, resources, and funds. This posed a challenge for smaller healthcare providers who may not have had the resources to implement the system.
4. **User training:** The successful implementation of CEMRS relied on the effective training of healthcare providers on how to use the system. This was a time-consuming process that required significant resources.

Despite these challenges, the benefits of CEMRS far outweighed the limitations. With effective planning and management, these challenges can be overcome to ensure the successful implementation and adoption of CEMRS.

VIII. CONCLUSION

A. Summary of the Research Findings

The research presented in this paper focused on the development and implementation of a Centralized Electronic Medical Record System (CEMRS) in healthcare and governance. A desktop application was developed for hospitals, clinics, and medical institutions to maintain patient records, including medical history, medications and allergies, various test reports, application forms, operation forms, discharge forms, and consent forms. The system architecture and design, database creation and management, patient record management, and features and functionalities of the system were discussed in detail.

The benefits and impacts of the CEMRS system were also highlighted, including improved patient care, reduced medical errors, increased efficiency, and cost savings. A case study was presented to illustrate the successful implementation of the system at a hospital, and user feedback and evaluation were discussed.

The analysis of the results and findings showed that the CEMRS system was effective in improving the quality of healthcare and governance, and it outperformed existing EMR systems in terms of its features and functionalities. However, there were some limitations and challenges encountered during the development and implementation, such as resistance to change, training and education, and data security concerns.

B. Future Directions and Recommendations for Further Improvement:

Based on the findings and limitations of this study, several recommendations for future research and improvement of the CEMRS system can be made.

- Future research could focus on enhancing the security measures of the system to ensure the confidentiality and privacy of patient data. This could include implementing stronger encryption protocols, multi-factor authentication, and regular security audits.
- The system could be further improved by integrating artificial intelligence and machine learning algorithms. These could be used to analyse patient data and provide insights for doctors, improve diagnosis accuracy, and identify potential health risks before they become serious.
- The system could be expanded to include more comprehensive patient data, such as genetic information, lifestyle habits, and social determinants of health. This could provide doctors with a more holistic view of the patient's health, leading to more personalized and effective treatment plans.
- To ensure the successful adoption and implementation of the CEMRS system, it is recommended that a thorough training program is developed for healthcare professionals. This will help to ensure that the system is used effectively and efficiently, leading to better patient outcomes.

In conclusion, the CEMRS system is a promising solution for improving healthcare and governance, and it has the potential to revolutionize the healthcare industry. The system can be further improved by addressing the limitations and challenges encountered during the development and implementation. Overall, this research provides valuable insights into the development and implementation of CEMRS systems and highlights the importance of technology in improving healthcare and governance.

IX. REFERENCES

1. Adel, S. A. (2015). Electronic medical record systems: benefits and challenges. Saudi Medical Journal, 36(4), 418-423.

2. Alkraiji, A., Jackson, T. N., Murray, J., & Ibrahim, M. (2017). Barriers and challenges in adopting Saudi Arabia telemedicine network: The perceptions of decision makers of healthcare facilities in Saudi Arabia. Journal of Health Informatics in Developing Countries, 11(1).
3. Buntin, M. B., Burke, M. F., Hoaglin, M. C., & Blumenthal, D. (2011). The benefits of health information technology: a review of the recent literature shows predominantly positive results. Health Affairs, 30(3), 464-471.
4. Dajani, H. R., Alonazi, W. B., Alzahrani, N. M., Alsoliman, A. S., & Almufleh, A. A. (2018). Awareness, perception and acceptance of electronic health records among healthcare providers in government hospitals in Eastern Province, Saudi Arabia. Journal of Infection and Public Health, 11(5), 705-710.
5. HealthIT.gov. (2019). Benefits of electronic health records (EHRs). Retrieved from https://www.healthit.gov/topic/health-it-basics/benefits-electronic-health-records-ehrs.
6. Hersh, W. R. (2015). The health information technology decade: A look back and future directions. Yearbook of Medical Informatics, 10(1), 13-22.
7. Jamal, A., Khan, S. A., & AlHumud, A. (2018). Electronic health records (EHRs) in the Kingdom of Saudi Arabia: A systematic review of adoption, barriers, and benefits. Journal of Infection and Public Health, 11(6), 745-755.
8. Ministry of Health. (2018). Saudi Arabia's vision for health care. Retrieved from https://www.moh.gov.sa/en/Ministry/vision2030/Pages/default.aspx.
9. O'Malley, A. S., Grossman, J. M., Cohen, G. R., Kemper, N. M., & Pham, H. H. (2012). Are electronic medical records helpful for care coordination? Experiences of physician practices. Journal of General Internal Medicine, 27(4), 415-420.
10. Shojania, K. G., Jennings, A., Mayhew, A., Ramsay, C. R., Eccles, M. P., & Grimshaw, J. (2009). The effects of on-screen, point of care computer reminders on processes and outcomes of care. Cochrane Database of Systematic Reviews, 3.

MODELLING AN INTRUSION DETECTION SYSTEM IN DISTRIBUTED NETWORKS

Anand Kumar Vishwakarma[1], Partha Paul[2], Keshav Sinha[3] and Manorama[4]
[1,2,4]Sarala Birla University, Ranchi
[3]University of Petroleum & Energy Studies, Dehradun

ABSTRACT
The study of intrusion detection systems (IDSs) has drawn a lot of interest in the field of computer science due to the rising network traffic and security concern. In addition to arbitrary intrusion categories, current IDSs also require a lot of processing capacity. We provide the taxonomy to delineate contemporary IDSs based on the large survey and sophisticated organization. For every organization that usesa network, information system security is of utmost importance. Attackers and hackers enjoy taking advantage of openings to break into distributed networks. Organizations all around the world are aware of this, but because there is not a reliable mechanism in place, they are still open to threats. Risk would be greatly decreased by improving the current mechanism used to stop intrusions in a dispersed network. In distributed intrusion detection systems that are currently in use, the issue of false positive and false negative results is very common and needs to be addressed. Working on intrusion detection systems in a dispersed network is urgently needed in order to increase efficiency, which can only be accomplished by lowering the number of false positive and in distributed intrusion detection systems currently in use, false negative results are a major problem that needs to be addressed. our propose work on intrusion detection systems in a dispersed network is urgently needed in order to increase efficiency, which can only be achieved by lowering the number of false positive and negative results. In distributed intrusion detection systems currently in use, false negative results are a major problem that needs to be addressed. The research on distributed intrusion detection systems is reviewed in-depth in this publication. It recognizes the necessity of setting up a reliable distributed intrusion detection system. By simulating, testing, and modelling a system, we hope to create one that makes use of current designs for these systems, old algorithms, and fuzzy logic theory.

Keywords: Intrusion Detection System, Security, Distributed Network System, False Positive and False Negative.

I. INTRODUCTION
Every organization's network must prioritise security. Distributed networks have been targeted by attackers for a while now. Despite this, many organisations lack a functional Distributed Intrusion Detection System (DIDS). This is due to the lack of a system that is impenetrable and uses an effective algorithm. Reducing organisational worries, attack risk, and system vulnerability could all be helped by improving an existing Distributed Intrusion Detection System model[1]. The issue of providing false positive and true negative results plagues the majority of intrusion detection systems now in use, increasing their mistake rates. In order to increase efficiency, it has become crucial to work on and improve these systems[2].

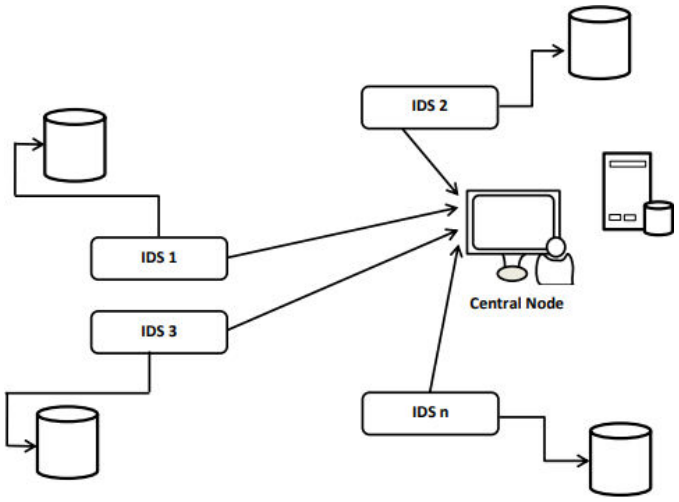


Figure 1: A typical Intrusion Detection System.

The firewall security is covered by the intrusion detection system. An organisation is protected by a firewall from malicious Internet attacks, and an intrusion detection system (IDS) detects attempts to breach firewall security and gain access to any system within the organisation [3]. If an IDS detects any suspicious activity in the firewall, it notifies the system administrator. An intrusion detection system (IDS) is a security system that monitors network traffic and computer systems and analyses that traffic to search for potential hostile assaultscoming from the outside as well as system abuse or attacks originating from within the company. [4].

II. TYPES OF INTRUSION DETECTION SYSTEM

A. There are two types of intrusion detection systems. These are, respectively, host-based and network-basedintrusion detection systems.

B. System for Network-Based Intrusion Detection and Prevention

An organization's network segment's network traffic is monitored by a Network Based IDS (NIDS) installedon a computer or other connected device while it searches for active assaults. Numerous different hashing algorithms, such as MD5, are used in networks to preserve file security. The network-based IDS reacts by notifying administrators of an attack when conditions arise that allow it to do so[5].

NIDSs are installed at a specific location in the network, such as a router, from which it is possible to observethe traffic entering and exiting a particular network segment. They can also be used to monitor only the traffic between a specific host computer on a network segment or the entire network as a whole.

B. Host-Based Intrusion Detection System

An individual computer or server is designated as the host for a host based intrusion detection system (HIDS), which is installed there and exclusively keeps track of activity on that system. The two subcategories of host-based intrusion detection systems are anomaly-based and signature-based (also known as abuse detection). Key system files are watched by HIDS, which alerts the user when a file is created, modified, or deleted. Then, the HIDS sends out an alert if one of the following things happens: new files are created, old files are deleted, or file attributes are modified. The primary distinction between NIDS and HIDS is that the former can access information that is encrypted while it is being transmitted via the network. something that alternates)[6].

III. COMPONENTS OF INTRUSION DETECTION

An IDS is made up of three main parts: the sensor (which includes an activity or packet capture engine and abehavioural or signature recognition engine), the backend (which records events in a database and alerts the engine), and the frontend (which includes a user interface and command and control). The main element of an IDS for spotting intrusions on a computer or network is a sensor. To carry out detecting operations, it captures a packet. It can use intrusion detection methods that are anomaly-based or signature-based. The logging of events that are discovered by the sensors is handled by the IDS's backend. In addition, it serves asan alerting mechanism. The backend can notify the administrator in a variety of methods, including logging events in the database, sending an e-mail, blocking, or resetting a TCP connection, and displaying the alert on the administrator's console. The IDS user interface is formed by the frontend. The user can monitor the events detected by the sensor, configure the IDS, and update the signature database and behavioral detectionengine [7].

IV.I NTRUSION DETECTION TECHNIQUES

The components of an intrusion detection system work together to warn the administrator of an intrusion [8].

i) Sensors:

A sensor has two interfaces. The capture network interface comes first, followed by the management networkinterface.

Its primary functions are detection and reporting. The capture interface stores all captured data in a buffer while the sensor listens to network traffic by tapping into the network. The detection engine then inspects thebuffer contents and performs network protocol analysis. Signature-based and anomaly-based intrusion detection were also used here[9].

ii) Command and control at the front end

The user can set up, configure, and update the IDS from the frontend. The frontend displays all events collected by the backend. As a result, the frontend now provides a user-friendly interface for managing theselogged events. To get the most out of an IDS, it must be fine-tuned to report only significant occurrences. Asa result, the user can fine-tune an IDS's detection and reaction using this console. If implemented correctly, the IDS will provide the user with enough early warning of any intrusion[10].

iii) Backend

The backend is also known as the primary function of an IDS. Its primary functions are to gather and alert. The sensor's detected events are saved in the event repository database system. The backend then defines howeach event must be handled. E-mails, displays, and blocking are utilized to handle key events.

V. FUNCTIONS OF INTRUSION DETECTION SYSTEM

- **Data collection:** This module sends data to IDS as input. The data is saved in a file and subsequently analyzed. Network-based intrusion detection systems capture and modify data packets, whereas host- based intrusion detection systems collect information such as disc usage and system processes [11].
- **Feature Selection:** To pick a certain feature, enormous amounts of data are available in the network, and they are typically analyzed for intrusion. The Internet Protocol (IP) address of the source and destination systems, protocol type, header length and size, and so on could all be used as keys for intrusion detection [12]
- **Analysis:** The data is examined to determine its accuracy. Rule-based intrusion detection systems (IDS) examine data by comparing incoming traffic to a specified signature or pattern. Another way is anomaly-based intrusion detection, which studies system behaviour and applies mathematical modelsto it[13].

- **Action:** It describes the system's reaction and attack. It can either notify the system administrator with the necessary data by email/alarm icons, or it can actively participate in the system by discarding packets so that they do not enter the system or closing the ports.

VI. APIDS (APPLICATION BASED IDS)

APIDS will examine the protocol's functional behaviour and events. The system or agent is installed between a process and a collection of servers and is responsible for monitoring and analysing the application protocol between devices. Intentional assaults are hostile attacks carried out by disgruntled employees with the goal of causing harm to the organisation, whereas unintentional attacks bring financial harm to the organisation by deleting a critical data file [14].

There have been various attacks on the OSI layer.

DOS (Denial-of-Service) Attacks: DOS stands for Denial-of-Service and is best characterized as an attempt to make a computer(s) or network(s) unavailable to its intended users. A Denial-of-Service attack occurs when an attacker generates more traffic than your resources can handle.

DOS and DDOS attacks: In a DOS attack, one computer and one internet connection are used to overwhelm a server or network with data packets, with the sole purpose of overwhelming the victim's bandwidth and available resources. A Distributed Denial of Service (DDOS) assault is similar, but it is more powerful. A DDOS is more than just one computer and one internet connection; it frequently involves millions of computers working together in a distributed fashion.

Peer-to-peer attacks: A peer-to-peer (P2P) network is a distributed network in which individual network nodes known as "peers" act as both suppliers (seeds) and consumers (leeches) of resources, as opposed to the centralised client-server model in which client server or operating system nodes request access to resources provided by central servers [15].

Ping of Death: A sort of DOS attack in which the attacker makes a ping request that is larger than the maximum size that IP allows onto the network, which is 65,536 bytes. While a ping bigger than 65,536 bytes cannot be broadcast in a single packet, TCP/IP permits a packet to be fragmented, basically separating it into smaller segments that are rejoined at the end. Attackers exploited this weakness by fragmenting packets so that when received, the packet totaled more than the allowable number of bytes, essentially causing a buffer overload on the operating system at the receiving end, causing the system to crash.

Eavesdropping Attack: The attacker's scheme of interfering with communication. This assault can be carried out over phone lines,instant messaging, or email.

Identity Spoofing (IP Address Spoofing): Most operating systems and networks use a computer's IP address to identify a legitimate entity on the network. In some situations, an IP address may be mistakenly considered to be faking identity. An attacker may also use special programmes to generate IP packets from legal IP addresses within the company intranet. After establishing network access with a genuine IP address, the attacker can edit, reroute, or delete your data.

Man-in-the-Middle Attack: A man-in-the-middle attack occurs when someone stands between you and the person with whom you are interacting and actively monitors, captures, and controls your communication in real time. An adversary, for example, can reroute a data transaction. When computers communicate at the lowest layers of the network layer, such as the physical layer, they may be unable to determine with whom they are sharing data. Man-in-the-middle attacks are like someone impersonating you in order to read your message. The person on the other end may believe it is you since the attacker is actively responding as you to continue exchanging information.

This attack can cause the same amount of harm as an application layer attack.

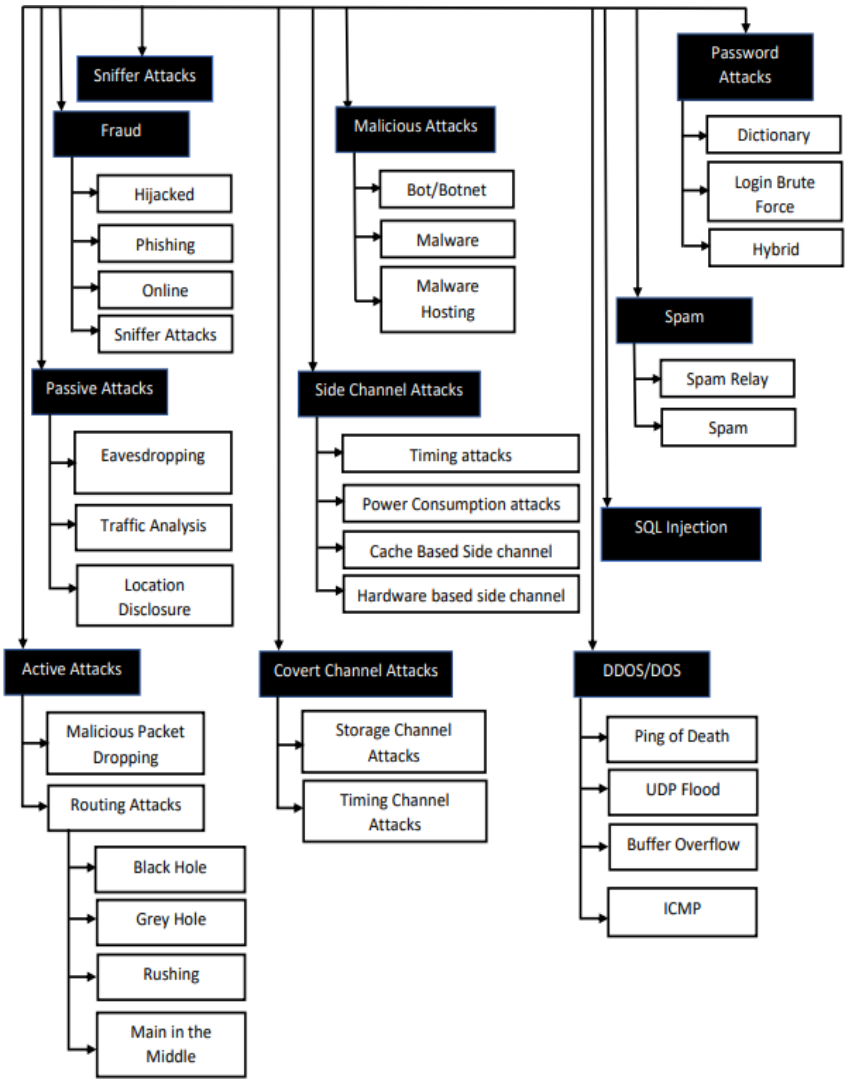


Figure 2: Applications of IDS

VII. DETECTION OF INTRUSION TOOLS

Today's intrusion detection products address a variety of organisational security objectives. The security equipment.

Snort is a lightweight, open-source firewall.

Snort describes traffic from an IP address using a flexible rule-based language; it records the packet in human-readable form using protocol analysis, content searching, and other pre-processors. Snort detects thousands of worms, vulnerability exploit attempts, port scans, and other potentially harmful activity.

OSSEC-HIDS: OSSEC (open-source security) is open-source software that is free to use. It will run on major operating systems and will have a Client/Server architecture. OSSEC has the capability of sending OSlogs to the server for analysis and data storage.

Many powerful log analysis engines, ISPs, universities, and data centres use it. Firewalls, authentication logs

KISMET is a WIDS (Wireless intrusion detection system) guideline. WIDS compromises with packet payload and WIDS occurrences. It will locate the burglar access point.

VIII. CONCLUSION

After putting firewall technologies at the network perimeter, intrusion detection systems (IDS) have becomethe mainstay for many organisations. Where traffic does not pass through the firewall, IDS can provide securityfrom both external and internal attackers.

However, the following points must always be remembered.

An IDS implementation combined with a firewall cannot create a highly secure architecture unless all of these points are connected.

1. **Authentication and identification:** An intrusion detection system (IDS) uses very sophisticated signatureanalysis algorithms to detect intrusions or potential misuse; however, organisations must still have strong user identity and authentication mechanisms in place.
2. **Intrusion Detection Systems (IDS) are not a panacea for all security issues:** IDS do a great job of monitoring and reporting intruder attempts. Furthermore, in order to reduce the risks of breaches, businesses must adopt a process of system testing, employee education, and the formulation of and adherence to a sound security policy.
3. **An intrusion detection system (IDS) is not a replacement for a strong security policy:** An IDS function,like effective security and monitoring solutions, is one component of a corporate security policy. A well- defined policy must be followed to guarantee that vulnerabilities, intrusions, and virus outbreaks, among other things, are addressed in accordance with corporate security policy requirements.
4. **Human intervention is necessary:** The attack must be investigated once by the security administrator ornetwork management. It is identified and reported, the cause is determined, the problem is corrected, and the appropriate steps are taken to prevent future instances of the same attacks.

REFERENCES

[1] Talaei Khoei, Tala, and Naima Kaabouch. " A Comparative Analysis of Supervised and Unsupervised Models for Detecting Attacks on the Intrusion Detection Systems." Information 14.2(2023): 103.

[2] Thakkar, Ankit, and Ritika Lohiya. " Fusion of statistical importance for feature selection in Deep Neural Network-based Intrusion Detection System." Information Fusion 90 (2023): 353-363.

[3] Cui, Jiyuan, et al. "A novel multi-module integrated intrusion detection system for high- dimensional imbalanced data." Applied Intelligence 53.1 (2023): 272-288.

[4] Louk, Maya Hilda Lestari, and Bayu Adhi Tama. "Dual-IDS: A bagging-based gradient boosting decision tree model for network anomaly intrusion detection system." Expert Systems with Applications 213 (2023): 119030.

[5] Rajapaksha, Sampath, et al. "Ai-based intrusion detection systems for in-vehicle networks: A survey." ACM Computing Surveys 55.11 (2023): 1-40.

[6] Salvatore Pontarelli, Giuseppe Bianchi, Simone Teofili.Traffic-aware Design of a High Speed FPGA Network Intrusion Detection System. Digital Object Indentifier 10.1109/TC.2012.105, IEEE TRANSACTIONS ON COMPUTERS.

[7] Przemyslaw Kazienko & PIOTr Dorosz. Intrusion Detection Systems (IDS) Part I - (network intrusions; attack symptoms; IDS tasks; and IDS architecture). www.windowsecurity.com › Articles & Tutorials

[8] Sailesh Kumar, "Survey of Current Network Intrusion Detection Techniques", available at http://www.cse.wustl.edu/~jain/cse571-07/ftp/ids.pdf.

[9] Srilatha Chebrolu, Ajith Abrahama*, Johnson P. Thomas, Feature deduction and ensemble design of intrusion detectionsystems, Elsevier Ltd.doi:10.1016/ j.cose. 2004. 09.008

[10] Uwe Aickelin, Julie Greensmith, Jamie Twycross. Immune System Approaches to Intrusion Detection A Review.http://eprints.nottingham.ac.uk/619/1/04icaris_ids_ review. pdf

[11] http://www.intechopen.com/download/get/type/pdfs/id/86 9 5.

[12] Martin Roesch, "Snort – Lightweight Intrusion Detection for Networks", © 1999 by The USENIXAssociation.

[13] The Snort Project, Snort User Manual 2.9.5, May 29, 2013, Copyright 1998-2003Martin Roesch, Copyright2001- 2003 Chris Green, Copyright 2003-2013 Sourcefire, Inc.

[14] Chapter 3, Working with Snort Rules, Pearson Education Inc.

[15] B. Daya, "Network Security: History, Importance, and Future, "University of Florida Department of Electrical and Computer Engineering, 2013. http://web.mit.edu/~ bdaya/ www/Network%20Security.pdf

[16] Li CHEN, Web Security: Theory and Applications, School of Software, Sun Yat-sen University, China.

[17] J. E. Canavan, Fundamentals of Network Security, Artech House Telecommunications Library, 2000

ABOUT THE EDITORS

Dr. Shraddha Prasad

Ph. D (Physics), M.Sc. (Physics), B.Ed.

Dr. Shraddha Prasad is Associate Professor cum Deputy Registrar (Academics) at Jharkhand Rai University, Ranchi. She is actively involved in academic & administrative responsibilities. She has done Ph.D from Birla Institute of Technology, Mesra, Ranchi. She has done post-graduation in Physics from Ranchi University, Ranchi and B.Ed. from St. Xavier's College, Ranchi.

A number of her papers are published in different reputed International journals with good impact factors e.g. Elsevier, Scopus, WoS etc. and numbers of research papers presented by her in different International Conference / Seminars. She has also contributed chapter in a book published by Springer Nature. She has edited one book and has published one patent. She has been convener of National/ International conferences. She is an expert reviewer of reputed Journal. She is a member of reputed organizations like Indian Academic Researchers Association (IARA), The Institution of Engineers, India

She has more than 15 years teaching experience in different reputed institutes including 3 years research experience as a Project Fellow in U.G.C Project at BIT, Mesra, Ranchi. As a research guide she has produced 2 Ph.D under her supervision and presently supervising 4 Ph.D Research Scholars at Jharkhand Rai University, Ranchi.

Dr. Harmeet Kaur

Ph. D (Management), CFA, MFA, UGC-NET, MBA

Dr. Harmeet Kaur is presently working with Jharkhand Rai University as Dean-Faculty of Commerce & Management. Her area of expertise is Finance and Business Accounting. She has more than 14 years of Academic & Research experience. In the current and previous organizations, she has been a key resource person and can deftly handle multiple responsibilities that range from teaching, planning and administering various university responsibilities. She proficiently leads, synchronizes team activities and complements the team performance

She also engages in domain related research by actively publishing various research papers and has also presented various papers both at the national and international conferences. She has many publications in Scopus indexed and UGC recognized journals. She has also authored two books. She brings in energy and fun-filled contemporary knowledge trends into the classroom that can make learning an effortless process. She is a structured personality who can proactively predict and strategize for the overall growth of the organization.

Parab Publications
Sharing Knowledge, Shaping Future

ISBN 978-81-19585-14-4

India | UAE | Nigeria | Malaysia | Montenegro | Iraq | Egypt | Thailand | Uganda | Philippines | Indonesia

Parab Publications || www.parabpublications.com || info@parabpublications.com