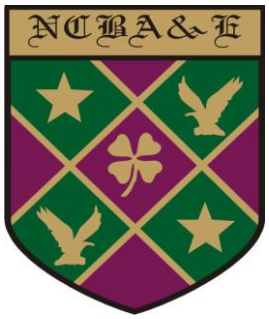
National College of Business Administration & Economics Lahore



ENHANCING CYBERSECURITY IN HEALTHCARE: AN EXPLAINABLE AI APPROACH TO IMPROVE INTRUSION DETECTION SYSTEM

BY

AMNA BATOOL

MASTER OF PHILOSOPHY
IN
COMPUTER SCIENCE

MARCH, 2024

NATIONAL COLLEGE OF BUSINESS ADMINISTRATION & ECONOMICS

ENHANCING CYBERSECURITY IN HEALTHCARE: AN EXPLAINABLE AI APPROACH TO IMPROVE INTRUSION DETECTION SYSTEM

BY

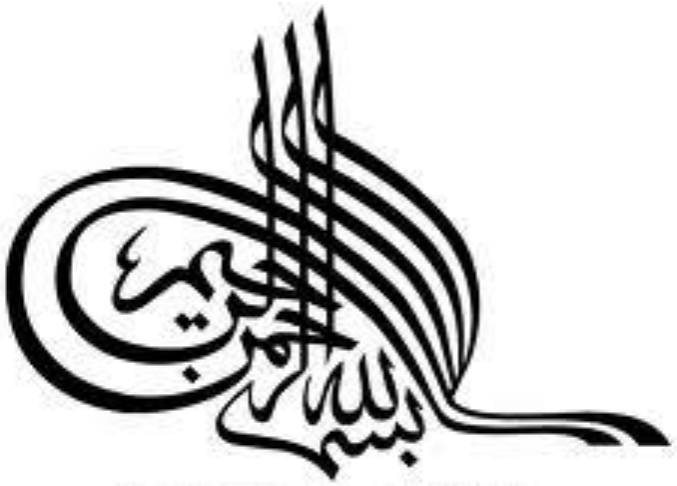
AMNA BATOOL

A dissertation submitted to
Faculty of Computer Sciences

In Partial Fulfillment oft he
Requirements for the Degree of

MASTER OF PHILOSOPHY
IN
COMPUTER SCIENCE

MARCH, 2024



In the name of ALLAH,
The Most Beneficial,
The Most Merciful,

NATIONAL COLLEGE OF BUSINESS ADMINISTRATION & ECONOMICS LAHORE

ENHANCING CYBERSECURITY IN HEALTHCARE: AN EXPLAINABLE AI APPROACH TO IMPROVE INTRUSION DETECTION SYSTEM

BY

AMNA BATOOL

A dissertation submitted to Faculty of Computer Sciences, in partial fulfillment of the requirements for the degree of

MASTER OF PHILOSOPHY IN COMPUTER SCIENCE

Dissertation Committee:

_____
Chairman

_____
Member

_____
Member

DECLARATION

It is to declare that this research work has not been submitted for obtaining similar degree from any other university/college.

AMNA BATOOL
March, 2024

Dedicated

To

My Parents

&

My Family

ACKNOWLEDGEMENT

I have the deepest gratitude to Allah Almighty, the “Omnipotent” the “Beneficent” the “Merciful” and the “All-Powerful” "who guided me to the path of knowledge and learning. My special appreciation and heartiest gratitude to our beloved Prophet Mohammad (S.A.W) who was a great seeker of knowledge and advised us to learn from cradle to grave.

I would like to extend my gratitude to the National College of Business Administration and Economics for providing me with such environment and resources, which enabled me to complete this research successfully.

I feel great honor and pleasure in expressing my profound and cordial gratitude to my Supervisor, Dr. Muhammad Saleem for academic support in the whole process from the beginning to the end of this thesis. Thanks to all my teachers in the college for giving me the different concepts throughout my studies.

I have furthermore to thank my friends in the Department of Computer Science for making my studies at an institution full of learning and friendship.

Especially, I have no words to pay gratitude to my parents, continuous encouragement enabled me to complete this research work. In the end, I pray to Almighty Allah to give me wisdom and strength to use this knowledge the way he wants.

RESEARCH COMPLETION CERTIFICATE

Certified that the research work contained in this thesis entitled **"Enhancing Cybersecurity in Healthcare: An Explainable AI Approach to Improve Intrusion Detection System"** has been carried out and completed by **Ms. Amna Batool** under my supervision during her **M.Phil. Computer Science** Programme.

(Dr. Muhammad Saleem)
Supervisor

SUMMARY

The ever-increasing digitalization of healthcare systems has ushered in a new era fraught with escalating cybersecurity threats. This thesis paper presents a pioneering solution aimed at fortifying cybersecurity within the healthcare sector, addressing the growing menace of cyber-attacks. The research at the nexus of cybersecurity and Artificial Intelligence (AI) introduces an innovative framework for Explainable Artificial Intelligence (XAI). The integration of machine learning and XAI technologies has had a profound impact on the capabilities and services provided by the healthcare industry.

XAI, with its promise of full privacy and anonymity, is instrumental in transactions and in identifying individuals involved therein. Recent years have witnessed the exceptional reliability of XAI technology across various sectors, spanning smart homes, medical services, banking, data storage administration, and cybersecurity. The burgeoning volume of data collected in medical and healthcare domains has underscored the urgency of timely medical information analysis, which is paramount for delivering optimal patient care. Data mining techniques have emerged as a potent tool for revealing latent patterns within vast medical datasets.

By amalgamating XAI with conventional intrusion detection systems, this research endeavors to offer lucid insights into AI-driven decision-making processes, thereby ensuring precise threat detection and the effective mitigation of potential risks. The proposed Explainable AI for Healthcare Intrusion Detection System (XAI-HCIDS) framework not only justifies the predictions and recommendations made by AI models but also serves as a robust guardian of data security.

The efficacy of this framework is empirically validated through real-world case studies, exemplifying its adaptability to the ever-evolving landscape of cyber threats. Additionally, ethical considerations underscore the imperative of safeguarding patient data and upholding the integrity of healthcare systems in the digital domain. In conclusion, this innovative approach fosters trust, resilience, and security in the realm of healthcare cybersecurity, offering a promising avenue for enhancing data protection and fortifying critical healthcare infrastructure.

LIST OF ABBREVIATION

Abbreviation	Description
XAI	Explainable Artificial Intelligence
ML	Machine Learning
IDS	Intrusion detection system
IoT	Internet of Things
NIDS	Network-based Intrusion detection system
HIDS	Host-based Intrusion detection system
DL	Deep Learning

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

In an era characterized by the increasing digitalization of healthcare systems and the growing threat of cyberattacks, the imperative to safeguard sensitive patient data and critical medical infrastructure has never been more pressing. Healthcare institutions and systems are increasingly reliant on interconnected digital platforms, opening new avenues for sophisticated cyber threats. As the healthcare sector experiences a paradigm shift towards data-driven decision-making and telemedicine, the protection of patient records, critical infrastructure, and medical devices becomes a paramount concern.

While healthcare systems invest in security platforms to protect their servers, the extensive data exchange between local user terminals and cloud servers heightens the vulnerability to network intrusions. Malicious hackers employ various forms of attacks, including DDoS attacks, injection attacks, infiltration attacks, and port scans, to hijack valuable information and incapacitate servers, preventing user access. In response, Intrusion Detection Systems (IDS) have been implemented to monitor network traffic for threats and detect anomalies before intruders can complete their objectives. When a threat is identified, the system promptly notifies the IT team to address anomalies before hackers can achieve their goals.

However, the lack of transparency and interpretability in current AI techniques diminishes user confidence in cybersecurity models, especially in the face of complex cyber-attacks. To address this critical issue, the application of eXplainable Artificial Intelligence (XAI) is crucial for creating models that are not only highly accurate but also understandable. XAI empowers users to comprehend and have confidence in advanced cyber defense systems of the future, combining the sophisticated defense capabilities of AI with transparency and user understanding.

This proposed research, "Explainable AI for Healthcare Intrusion Detection System (XAI-HCIDS)," aims to underscore the significance of XAI as an enabling technology that enhances the confidence and trust of healthcare professionals in the decision-making processes of IDS. By addressing these challenges, future research can unlock the full potential of XAI in developing the cybersecurity landscape of the healthcare industry, ultimately enhancing cybersecurity in healthcare and preserving the integrity of patient data and critical medical infrastructure.

1.1 BACKGROUND STUDY

Securing data is crucial in today's context where extensive data exchange occurs among hospitals, corporations, and consumers. Given the trust placed in them, protecting this data is paramount. Despite extensive investments in secure servers by the healthcare sector, a single hacker can undermine the trust established between parties. To counter such threats, numerous automated security solutions have been created, yet none have seen as much adoption as Intrusion Detection Systems (IDS), which play a vital role in safeguarding against malicious attacks.

1.2 HEALTH MANAGEMENT SYSTEM AND SECURITY CHALLENGES

Healthcare management faces numerous challenges to safeguarding the network and patient sensitive data from data breaches and the privacy of personal health data in an increasingly digital and interconnected healthcare landscape. By adopting the new technologies in the healthcare field with best security measures will maintain a reliable and trustworthy healthcare system.

1.3 IMPORTANCE OF HEALTH MANAGEMENT

Effective healthcare management is essential for the smooth and safe functioning of healthcare facilities, particularly hospitals. Healthcare managers are responsible for overseeing various aspects of the medical industry to ensure efficiency. Their role includes timely healthcare delivery, clinical data management, healthcare policy, economy of healthcare, quality assurance and secure the networks. Through strategic planning and effective leadership, healthcare managers contribute to improved and more efficient healthcare services, ultimately benefiting patients and the entire healthcare system. (Castonguay, 2021).

1.4 SECURITY RISKS IN HEALTH INFRASTRUCTURES

Security risks in health infrastructures pose significant challenges to patient safety, data privacy, and operational continuity. Some key security risks in health infrastructures include:

1.4.1 Cyberattacks

Health infrastructures are prime targets for cyberattacks, including denial-of-service (DoS) attacks, ransomware, virtual attacks, and has the potential to disturb medical operations, expose patient information, and result in substantial financial losses.(Bhosale, Nenova, & Iliev, 2021).

1.4.2 Weak Access Controls and Authentication

Insufficient authentication protocols and lenient access controls can lead to illegal entry to sensitive patient information or critical systems, leading to data breaches and privacy violations.

1.4.3 Insider Threats

Employees or staff with access to healthcare systems may intentionally or unintentionally cause security breaches, either through malicious intent or negligence.

1.4.4 IoT Vulnerabilities

The increasing use of IoT devices in healthcare, for example medical devices and wearables, introduces additional security risks, as these devices may lack robust security features and become potential entry points for cyberattacks. (Tsantikidou & Sklavos, 2021).

1.4.5 Outdated Software and Systems

Running outdated or unpatched software and operating systems can expose health infrastructures to known vulnerabilities that attackers can exploit.

1.4.6 Third-Party Vendor Risks

Healthcare institutions frequently depend on third parties for the range of facilities, and if these external vendor have weak security measures, it can lead to supply chain vulnerabilities and data breaches. (Bandari, 2023).

1.4.7 Social Engineering

Hackers may use social engineering tactics, such as scams and phishing emails, to trick healthcare employees by uncovering their sensitive information and granting an unauthorized access.

1.4.8 Physical Security Breaches

It's an unauthorized access to restricted areas of medical equipment, that can compromise patient safety and disrupt healthcare operations. (Gopalan, Raza, & Almobaideen, 2021).

1.4.9 Data Integrity and Availability

Ensuring the integrity and availability of healthcare data is crucial. Data manipulation or loss can lead to incorrect medical decisions and patient harm.

1.4.10 Regulatory Compliance Challenges

Healthcare facilities need to adhere to different rules and criteria, like the Health Insurance Portability and Accountability Act (HIPAA), in order to meet health infrastructure requirements. (Alder, 2020) in United States, which adds complexity to security management and risk mitigation.

Addressing these security risks requires a comprehensive approach, including robust cybersecurity measures, regular audits and security assessments, employee training practice on security process, and a proactive incident response plan. Healthcare organizations must prioritize security to safeguard patient information and ensure the continuity of critical medical services.

1.5 INTRUSION DETECTION SYSTEM

Intrusion involves unauthorized entry by intruders exploiting the network. An Intrusion Detection System (IDS) is a tool that consistently watches both incoming and outgoing network data to identify shifts or patterns in activity. When it identifies abnormal conduct, it notifies an administrator, who subsequently evaluates alerts and responds to eliminate the risk. The primary

aim of an IDS is to recognize anomalies before hackers achieve their goals. Once a threat is detected, the IT team is notified and informed action is taken.

When the hacker tries to breach the External network, originating beyond the organization's boundaries, this can be achieved by getting direct access to the network’s credentials or it can be done using multiple attacks like DDOS (Distributed Denial of Service) attacks, injection attacks etc. Two types of intruders can exploit user privacy and confidential information known as.

1.5.1 Masqueraders and Misfeasors

A Masquerader is an external intruder with no direct system access, who maliciously targets the system to unlawfully extract data or information. Misfeasor are people that are insiders which is dangerous. They unethically aim to attain direct system access through attacks for the purpose of wrongfully obtaining data or information.

To protect your network from intruders and their exploits, two methods are commonly used for intrusion detection. The initial approach is Signature-based intrusion detection, while the second method involves Anomaly-based intrusion detection.

1.5.2 Signature-Based Intrusion Detection

Signature-based intrusion detection identifies cyber threats by comparing known patterns, called signatures, of malicious activity with the network's current traffic. When a match is found, it signals a potential intrusion attempt.

1.5.3 Anomaly-Based Intrusion Detection

Anomaly-based intrusion detection is tailored to identify unknown attacks like new malware by utilizing machine learning to adapt dynamically. It establishes trust models based on normal activity and detects anomalies by comparing new behaviors against these models. However, false alarms can arise due to legitimate but unfamiliar network traffic being misidentified as malicious. Figure 1 is representing the architecture of an anomaly based intrusion detetion system (Otoum & Nayak, 2021).

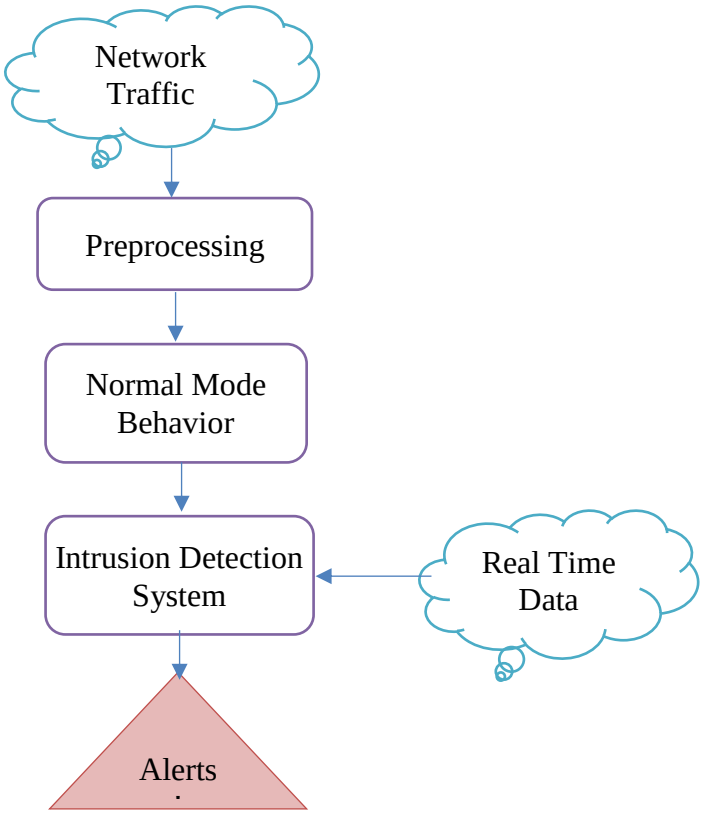


Figure 1: Generic Architecture of an Anomaly-based Intrusion Detection System

1.5.4 Hybrid Intrusion Detection

The hybrid intrusion detection system combines signature-based and anomaly-based approaches to improve threat detection. It can identify a broader range of threats and understands evasion techniques used by cybercriminals, like fragmentation and low bandwidth attacks. (Maseno, Wang, & Xing, 2022).

Intrusion detection systems can be broadly classified into three main types.

1.5.5 Network-based IDS (NIDS)

Network-based intrusion detection deploys sensors strategically within or at the network's perimeter to monitor, analyze and capture inbound and outbound network traffic for malicious activity. Multiple instances of NIDS may be required based on network architecture and traffic volume. (Ashoor & Gore, 2011).

1.5.6 Host-based IDS (HIDS)

In Host-based intrusion detection system agent runs all the devices, endpoints and servers with internet and internal network access. HIDS analyzes specific activities like file system modifications, registry changes, and access control lists, as well as monitors system application logs. It complements NIDS by detecting anomalous traffic originating from within the organization or the monitored host. (Ahmad, Emad Ul Haq, Imran, O. Alassafi, & A. AlGhamdi, 2022).

1.5.7 Cloud-based IDS

The third variant is Cloud-based IDS, designed to monitor cloud environments effectively. Traditional on-premises IDS solutions may not be optimized for the cloud, so cloud-specific sensors use cloud service provider APIs to gain visibility into the cloud environment.

1.6 ROLE OF INTRUSION DETECTION IN HEALTHCARE MANAGEMENT

In healthcare management systems, IDS are essential due to the growing reliance on digital technologies and the associated cybersecurity risks. Implementing IDS in healthcare offers several advantages. This involves safeguarding confidential patient information against unauthorized access, ensures data integrity by detecting and preventing data manipulation, and identifying and responding to cyberattacks such as ransomware, phishing, and malware. (Kumaar, et al., 2022) Additionally, IDS helps maintain uninterrupted healthcare services, secures medical devices connected to the internet, and guaranteeing compliance with regulatory requirements like HIPAA. (Alder, 2020). Early detection of insider threats and real-time incident response capabilities further enhance cybersecurity posture, demonstrating the organization's commitment to safeguarding patient information. Overall, deploying IDS in healthcare management systems is crucial to proactively manage cybersecurity risks, protect patient data privacy, and maintain the reputation of healthcare organizations.

1.7 MACHINE LEARNING

A branch of artificial intelligence called "machine learning" has been devoted to "learning." It involves developing statistical models and algorithms that, given data, can learn to carry out tasks on their own without needing to be specifically programmed for each activity.

There are different types of Machine Learning approaches that are used for different types of problems. The three most well-known types are:

1.7.1 Supervised Learning

This is the most widely used type of Machine Learning in data science. In supervised learning, a model is trained on a “labelled dataset”. Labelled datasets have both input and output parameters. The algorithms learn to map points between inputs and correct outputs. For example, while inspecting damages on concrete material, photos of previous examples are being used of which it is known whether they show damage or not. Each of these photos has been given a label: ‘damaged’ or ‘not damaged’, which helps in further classification.

1.7.2 Unsupervised Learning

In unsupervised learning, labels are not used, but the model itself tries to discover relations in the data. This is mainly used for grouping (clustering) the examples from the data. For example, creating different customer groups where customers with similar characteristics are in the same group. Beforehand it is unknown, which groups of customers there are and which characteristics they meet, but the algorithm can distinguish the different groups with enough data available.

1.7.3 Reinforcement Learning

Reinforcement learning models learn on the basis of trial and error. By rewarding good choices and punishing bad ones, the model learns to recognize patterns. This technique is mainly used in understanding (computer) games (such as Go) and in robotics (a robot which learns to walk through falling and standing up again). This type of Machine Learning usually falls outside of data science, because the purpose of ‘learning’ a task is the goal and not understanding and using the underlying data.

Similarly, there are numerous machine learning models and algorithms used for data classification, providing accuracy in the results. Chapter 3 of this research discusses some of these models for testing and evaluating the outcomes.

1.8 EXPLAINABLE ARTIFICIAL INTELLIGENCE

In the present era, Explainable Artificial Intelligence (XAI) has gained widespread usage, mainly to justify the predictions and recommendations of AI models. XAI provides insight into the decision-making mechanism of AI models, playing a vital role in establishing trust and ensuring accountability in AI systems. In contrast to traditional AI, XAI prioritizes transparency, fairness, and ethics, making it a crucial tool for addressing the challenges posed by AI, such as bias and ethical concerns. In this research, XAI techniques used to detect the data breaches of healthcare organizations using an intrusion detection system.

1.9 INTEGRATION OF XAI IN HEALTHCARE MANAGEMENT FOR INTRUSION DETECTION SYSTEM

XAI has risen as an important asset within healthcare administration, particularly in the context of intrusion detection. With the increasing reliance on digital technologies and the rising threat of cybersecurity breaches in healthcare systems, the need for robust intrusion detection mechanisms is more critical than ever. Traditional Intrusion Detection Systems (IDS) often operate as black-box models, causing difficulty to understand the issue behind their choices, hindering effective interpretation and trust among healthcare professionals and administrators.

In this context, XAI offers a promising solution by providing transparency and interpretability to the intricate workings of AI-based intrusion detection systems. By integrating XAI techniques into healthcare management systems, such as those used in hospitals and medical facilities, it becomes possible to gain a deeper understanding of the IDS predictions, thus empowering cybersecurity experts to detect and respond to potential threats effectively. (Rehman & Farrakh, Improving Clinical Decision Support Systems: Explainable AI for Enhanced Disease Prediction in Healthcare, 2023)

This research explores the integration of XAI in healthcare management systems for intrusion detection. The aim of this research is to demonstrate how

XAI techniques can enhance the accuracy, reliability, and explainability of intrusion detection, thereby assisting healthcare professionals in making well-informed decisions to ensure patient data privacy and system security. Through this research, it intends to shed light on the transformative potential of XAI in safeguarding healthcare infrastructures from cyber threats and fostering trust in the overall cybersecurity measures deployed in the healthcare industry.

1.10 PROBLEM STATEMENT

In the rapidly evolving landscape of healthcare technology, the critical need for robust cybersecurity measures has become paramount. Healthcare systems, reliant on interconnected networks and digital data exchange, face an escalating risk of cyber intrusions, data breaches, and malicious attacks. Although intrusion detection systems (IDS) are conventionally employed to safeguard against such threats, their efficacy in the intricate healthcare environment remains limited. Moreover, the lack of transparency and interpretability in these systems hinders swift and informed decision-making, undermining proactive response strategies.

This research aims to address the shortcomings of existing cybersecurity protocols in healthcare by pioneering an innovative approach. By harnessing the power of Explainable Artificial Intelligence (XAI), we seek to revolutionize intrusion detection within healthcare systems. Our focus lies in developing an advanced IDS that not only boasts heightened accuracy and detection capabilities, but also provides clear and intelligible insights into its decision processes. Through a comprehensive exploration of AI-driven techniques, real-time monitoring, and the incorporation of domain expertise, we aspire to create a paradigm shift in healthcare cybersecurity.

1.11 RESEARCH QUESTIONS

- What are the key factors contributing to the limitations of current intrusion detection systems in the healthcare sector?
- How can XAI techniques be integrated into an intrusion detection system to provide interpretable insights into detected threats and anomalies?
- How can real-time monitoring be achieved to enhance the system's responsiveness and accuracy in detecting emerging cyber threats?

- What are the potential challenges and ethical considerations associated with implementing XAI-enhanced intrusion detection systems in healthcare, and how can they be addressed?

1.12 RESEARCH OBJECTIVE

- To identify and analyze the key factors contributing to the limitations of current intrusion detection systems in the healthcare sector.

- To design and develop a XAI-powered intrusion detection system tailored to the healthcare domain, using methods of machine learning and conducting real-time analysis of data.

- To implement a real-time monitoring mechanism that continuously adapts and updates the IDS based on evolving cyber threats and attack patterns.

- Explore challenges and ethical considerations of integrating XAI-enhanced intrusion detection systems in healthcare while proposing mitigation strategies.

This research aims to address the above mwentioned questions and objectives by strengthening cybersecurity in the healthcare sector. It involves creating an XAI-powered intrusion detection system developed to efficiently recognize and counter cyber threats. This system aims not only to safeguard against threats but also to offer transparent decision-making support for healthcare professionals.

CHAPTER 2

LITERATURE REVIEW

This section analyses the present literature work and discusses the challenges in different existing approaches. In the healthcare field, numerous techniques are focused on. The objective of the literature survey is to practice XAI by using intrusion detection system in the healthcare area. In this literature review, numerous researches shared the perspective by using different cases of XAI techniques in the healthcare field. Researchers share various articles by concentrating on the challenges and core issues discussed below:

The field of XAI gained considerable attention in recent years by providing model improvement and debugging processes, and infuses greater trust in the decisions and predictions made by AI models. The authors discusses the significance of XAI in cybersecurity, specifically focusing on its applications in the IoT and anomaly-based IDS. This research examines how XAI techniques are utilized for identifying anomalies in IoT networks and highlighting the need for interpretable justifications and model confidence in cyber defense. It also investigates the use of machine learning and deep learning models in security applications, with a focus on highlighting deep learning's effectiveness in understanding security events. Furthermore, the research delves into the convergence of XAI, anomaly-based IDS, and IoT, addressing current challenges and potential solutions. The primary objective is to utilize XAI models to assist decision-makers in comprehending security incidents in compromised IoT networks and to uncover innovative cybersecurity applications. (Moustafa, Koroniotis, Keshk, Zomaya, & Tari, 2023)

According to (Khan, et al., 2022), conducted a comprehensive study on the adoption of Internet of Medical Things (IoMT) in healthcare, highlighting the lack of cybersecurity attention in IoMT implementation. The paper addresses challenges in securing IoMT networks and the limitations of conventional machine learning and deep learning methods in cybersecurity. To tackle these issues, the authors introduce the XSRU-IoMT model, which employs XAI techniques to enhance trust by offering explanations for predictive decisions. The model incorporates bidirectional simple recurrent units (SRU) with skip connections to address training difficulties in recurrent networks. Evaluation of the XSRU-IoMT model on the ToN_IoT dataset demonstrates its superior performance compared to state-of-the-art detection models, by recommending its suitability for practical deployment in real IoMT networks.

The authors (Rehman, Farrakh, & Khan, Explainable AI in Intrusion Detection Systems: Enhancing Transparency, 2023) focuses us to provide the secure network against all cyber-attacks and uses the XAI techniques to improve IDS readability. The authors propose a strategy combining NSL-KDD dataset training with XAI methods to explain post-modeling results and differentiate attacks from regular traffic. The LIME algorithm provides understandable explanations. The integrated XAI improves IDS interpretability in predicting attacks and recognizing regular traffic, increasing confidence in network security decisions.

Similarly, in another research work (Patil, et al., 2022), researchers have introduced an innovative IDS that utilized ML ensemble methods. By integrating attributes sourced from the CICIDS-2017 dataset and utilizing algorithms such as SVM, RF and decision trees, the system achieves excellent accuracy through the addition of an ensemble technique voting classifier. Furthermore, the model incorporates the XAI algorithm LIME to enhance the comprehensibility and insight into the dependable black-box methodology for effective intrusion detection. Experimental outcomes validate that the use of XAI LIME yields explanations that are more user-friendly and promptly responsive.

In (Asif, et al., 2022), authors presented the increasing importance of cybersecurity due to the expansion of the IoT and computer networks. It focuses on the limitations of traditional IDS in handling large volumes of data intelligently. In response, the authors put forth a solution called MR-IMID (MapReduce-Based Intelligent Model for Intrusion Detection), which merges machine learning and clustering methods to tackle this issue. MR-IMID effectively handles large datasets by utilizing standard hardware and detects intrusions in real-time from multiple network sources. The proposed model achieves the best detection accuracy during training and validation, outperforming previously published approaches.

(Medjahed, Istrate, Jerome, Baldinger, & Dorizzi, 2011) proposed an automated in-home healthcare monitoring system to address various needs and requirements. This telemonitoring platform integrates multiple sensors installed in the home to collect comprehensive data. The system comprises physiological and behavioral data from older adults, acoustical environment details, environmental factors, and medical expertise. Distinct algorithms are applied to each modality for processing and analysis. The data fusion technique is founded on fuzzy logic and medical recommendations, combines outputs from different subsystems. This multimodal fusion enhances the system's reliability by detecting distress situations, accounting for sensor malfunctions, environmental disturbances, and material limitations. The flexible fuzzy logic fusion allows for

easy combination of modalities or addition of other sensors. The model aims to enhance the overall quality of healthcare in domestic settings and support the well-being of elderly populations.

The emergence of the IoT has transformed cities into smart cities, but it also raises cybersecurity issues as a result of data sharing and continuous connectivity within IoT networks. DL-based models are complex and difficult to interpret, leading to a lack of understanding and trust in their decisions. To address this, XAI techniques are used to interpret and explain DL models' predictions. In this research (Houda, Brik, & Khoukhi, 2022), a fresh approach based on Explainable Artificial Intelligence (XAI) is developed to elucidate critical deep learning (DL) decisions within IoT-based Intrusion Detection Systems (IDS). The novel framework employs advanced deep neural networks for IoT IDS and integrates three key XAI techniques (RuleFit, LIME, and SHAP) to furnish both local and global explanations for DL-based choices. This framework enhances transparency and trust between the DL-based IDS model and cybersecurity professionals. By validating on NSL-KDD and UNSW-NB15 datasets, the usefulness of XAI framework is demonstrated in enhancing the interpretability of IoT IDS against prevalent IoT attacks and facilitating cybersecurity experts in comprehending IDS decisions..

The industries' varied characteristics and constrained resources render them vulnerable to an extensive range of cyber risks. It will lead it to financial losses, reputational damage, and data theft. To counter these risks, authors (Sridhar & Sivamohan, 2023) introduced a fresh approach to intrusion detection titled Bidirectional Long Short-Term Memory based Explainable Artificial Intelligence (BiLSTM-XAI).. This system efficiently detects network intrusions by utilizing data preprocessing, feature selection through the Krill herd optimization (KHO) algorithm, and explainable AI techniques such as SHAP and LIME. The experimental results using MATLAB 2016, Honeypot, and NSL-KDD datasets show the BiLSTM-XAI approach achieves superior performance, with a best classification accuracy in detecting intrusions, enhancing security and privacy in industrial networking systems.

The authors (Ardito, et al., 2020) provided an overview, highlighting that within the swiftly advancing E-Health sphere, ensuring the security of telemonitoring systems is becoming increasingly crucial due to emerging threats. This study presents a model named Cyberattack Detection System (CADS), which employs artificial intelligence methods to identify irregularities in telemonitoring systems independently, eliminating the requirement for a security analyst. The CADS model not only detects suspicious activities but also explains malicious conduct, providing healthcare staff with suspected attack

data for their insights. The system is designed for situations involving compromised remote patient health telemonitoring setups.

In the realm of network security, Intrusion Detection Systems (IDS) play a crucial role, and machine learning is a prominent approach for their implementation. But, latest researchs has unveiled the vulnerability of machine learning-based IDSs to adversarial attacks. To tackle this challenge, authors (Lee, Kim , cydenova, & Park, 2021) present a solution: an adversarial attack detection framework for machine learning-based explainable AI intrusion detection systems. This framework operates through two distinct phases: initialization and detection. During initialization, an IDS is trained using a support vector machine classification model, and explanations for normal data patterns are derived using LIME (local interpretable model-agnostic explanations). In the detection phase, these explanations are utilized to scrutinize the outcomes of the IDS classification process in order to identify adversarial attacks. The effectiveness of this proposed approach is assessed using the NSL-KDD dataset.

Table 1
Comparison of the Proposed Model with Previous Published Approaches

Author and title	Type of data	Decision Making	Predictive Model	Use of IDS	Use of XAI
(Sridhar & Sivamohan, 2023)	Industry 4.0	YES	YES	YES	YES
(Larriva-Novo, Villagrá, Vega-Barbas, Rivera, & Rodrigo, 2021)	Traffic Record	YES	YES	YES	NO
(Patil, et al., 2022)	Network traffic data CICIDS-2017	YES	YES	YES	YES
(Houda, Brik, & Khoukhi, 2022)	NSL-KDD and UNSW-NB15 datasets	YES	YES	YES	YES
(Ardito, et al., 2020)	E-Health SECTOR	YES	YES	YES	YES
(Kasongo, 2023)	NSL-KDD and the UNSW-NB15	YES	YES	YES	NO
(Ustun, Hussain, Yavuz, & Onen, 2021)	Smart grids security utilizing IEC 61850	YES	YES	YES	NO

Author and title	Type of data	Decision Making	Predictive Model	Use of IDS	Use of XAI
(Ho, Jufout, Dajani, & Mozumdar, 2021)	(CICIDS2017) dataset	YES	YES	YES	NO
(Wang, Zheng, Yang, & Wang, 2020)	NSL-KDD dataset	YES	YES	YES	YES
(Khan, et al., 2022)	Medical Records	YES	YES	YES	YES
(Kabir, Fida Hasan, Hasan, & Ansari, 2022)	Smart City Application	YES	YES	NO	YES
(Asif, et al., 2022)	NSL-KDD dataset	YES	YES	YES	NO
Proposed model	Medical record	YES	YES	YES	YES

In Table 1, we analyze diverse sectors, including healthcare organizations and the farming industry, as well as traffic management records. This comparison aims to highlight the benefits and contrasts of employing Explainable Artificial Intelligence (XAI) and Intrusion Detection Systems (IDS) for enhancing cybersecurity measures and fortifying network security.

CHAPTER 3

PROPOSED METHODOLOGY

In this chapter, a comprehensive explanation is furnished regarding the outlined framework, known as XAI-HCIDS (Explainable AI for Healthcare Intrusion Detection System). This framework is designed to achieve precise identification of intrusions within healthcare networks. As shown in Fig. 2, the entire architecture of the proposed XAI-HCIDS consists of four main streams: 1) preprocessing; 2) generative model training; 3) autoencoder training; and4) predictive model training. In this section, we describe the proposed methodology and each module (process) in detail.

3.1 RESEARCH DESIGN

Figure 2 presents that the arrangement of the proposed Healthcare system using Explainable Artificial Intelligence framework. It is being evaluated utilizing the process of training and validation consists of distinct stages. The training phase encompasses six layers specifically; Healthcare organization infrastructure, dataacquisition layer, preprocessing layer, Prediction layer, prediction analysis, Explainable Artificial Intelligence layer as well as XAI explanation.

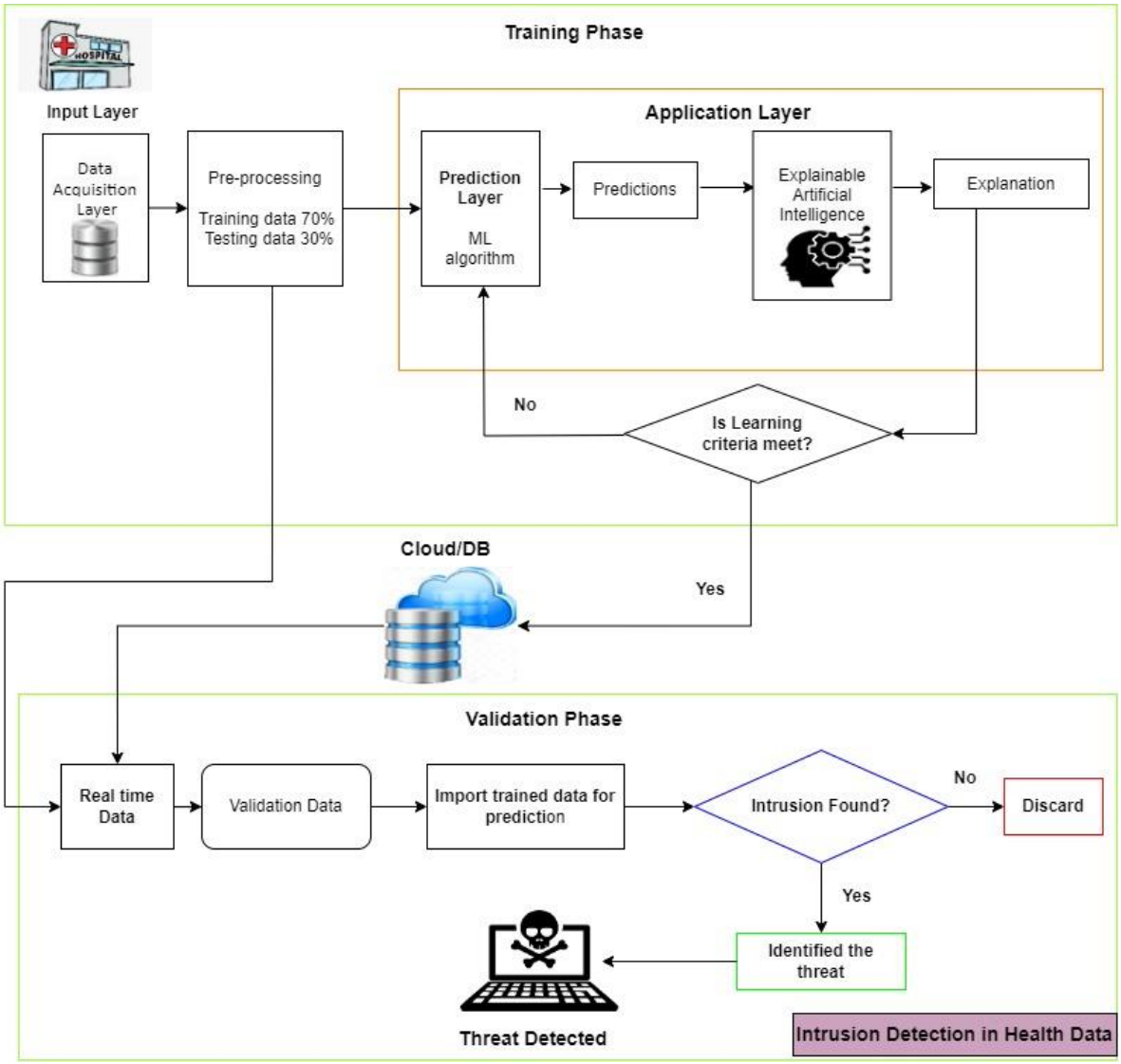


Figure 2: (XAI-HCIDS) Explainable AI for Healthcare Intrusion Detection System Framework

Figure 2 is the representation of proposed XAI-HCIDS model which is being evaluated utilizing the training and validation phases. The training phase layers specifically; the IoMT-enabled medical devices connected with the healthcare system are used to collect the data and send it to the preprocessing layer, which is playing a major role in mitigating the amount of noise generated due to the wireless medical devices communication. The preprocessed medical data is then promoted to the application layer where the intrusion intrusions are predicted using ML approach and forwarded to the XAI approach that is responsible for predicting explained patterns.

XAI methods use predictions and medicl data is utilized to generate explanations in verifying a specific decision making taken by a ML technique that is playing a tremendously important role in smooth intrusion detection in

healthcare data. The primary goal of XAI is to build trust in the proposed model. After the XAI, it is checked, if the learning criteria are found or not. In the case of "No", the ML will be retrained, and so on, but in the case of "Yes", the predicted outcome is stored on a cloud database. In the Validation stage, the trained patterns are imported from the cloud data set and real-time data from the medical devices input layer to check if the intrusion is predicted or not. In the case of no, the process goes exit, whereas in the case of yes the message is shown that smart intrusion in healthcare data is predicted.

3.2 IOMT ENABLED HEALTHCARE DATA INPUT LAYER

The IoMT (Internet of Medical Things) technology enables the collection and analysis of real-time healthcare data from various sources, such as medical sensors, integrated into patients and healthcare infrastructure. This technology serves as a medical data input layer for a larger IoMT system, which can include data analysis, decision-making, and medical communication. By gathering accurate and timely data on healthcare patterns and patient conditions, the IOMT technology can help improve healthcare management, reduce intrusion, and enhance overall healthcare efficiency.

Some of the technologies that can enable IoMT-based healthacre data collection using medical sensors. These technologies can facilitate the communication between patient and healthcare infrastructure in real-time, allowing for more efficient healthcare management and better decision-making by healthcare professionals and healthcare service providers. The IoMT-based medical data collection technology has the potential to revolutionize the way to approach healthcare management and improve the quality of life for patients and healthcare service provider by reducing intrusion in healthcare data and improving patient monitoring.

3.2.1 Data Collection

Data collection is the first step of the intrusion detection system, where the dataset are collected and processed for training and testing the Machine learning procedures. Data collection and processing phases characterizes network traffic and biometric data using wide range of features such as source byte, Destination byte, total packet count, heart rate, pulse rate etc.

Dataset: The data collection and processing phases using wide range of features such as such as duration, prtottypes, service flag, src_bytes, dst_bytes, land, wrong_fragment, urgent, dst_host_srv_count etc. The

XAI-HCIDS framework's data collection includes 42 functions for each record. The table displays all the data features and their corresponding data types used for training and testing in this research. The given table provides a detailed description of the functions.

Table 2
XAI-HCIDS Framework Dataset Structure

Sr.	Features	Datatype	Sr.	Features	Datastype
1	Duration	Integer	**22**	is_guest_login	Integer
2	protocol_type	Object	**23**	Count	Integer
3	Service	Object	**24**	srv_count	Integer
4	Flag	Object	**25**	serror_rate	Float
5	src_bytes	Integer	**26**	srv_serror_rate	Float
6	dst_bytes	Integer	**27**	rerror_rate	Float
7	Land	Integer	**28**	srv_rerror_rate	Float
8	wrong_fragment	Integer	**29**	same_srv_rate	Float
9	Urgent	Integer	**30**	diff_srv_rate	Float
10	Hot	Integer	**31**	srv_diff_host_rate	Float
11	num_failed_logins	Integer	**32**	dst_host_count	Integer
12	logged_in	Integer	**33**	dst_host_srv_count	Integer
13	num_compromised	Integer	**34**	dst_host_same_srv_rate	Float
14	root_shell	Integer	**35**	dst_host_diff_srv_rate	Float
15	su_attempted	Integer	**36**	dst_host_same_src_port_rate	Float
16	num_root	Integer	**37**	dst_host_srv_diff_host_rate	Float
17	num_file_creations	Integer	**38**	dst_host_serror_rate	Float
18	num_shells	Integer	**39**	dst_host_srv_serror_rate	Float
19	num_access_files	Integer	**40**	dst_host_rerror_rate	Float
20	num_outbound_cmds	Integer	**41**	dst_host_srv_rerror_rat	Float
21	is_host_login	Integer	**42**	Class	Object

This dataset, that was obtained from Kaggle, was created to record raw TCP/IP dump data for a network that simulated the LAN environment of a healthcare facility. The LAN was carefully designed to resemble an actual situation and was attacked by several cybercriminals. Inside this framework, a connection is characterized as a sequence of TCP packets with discrete beginning and ending times that enable data transfer between a source and a destination IP address inside a clearly defined protocol. Every connection is categorized as either normal or an attack, and each attack type is given a unique designation. The size of each connection record is roughly 100 bytes.

For every TCP/IP connection, 41 features, encompassing both quantitative and qualitative data, are extracted from both normal and attack instances (3 qualitative and 38 quantitative features). The connection type is indicated by the class variable, which falls into one of two categories: "Normal" or "Anomalous." The dataset is represented in Figure 3.

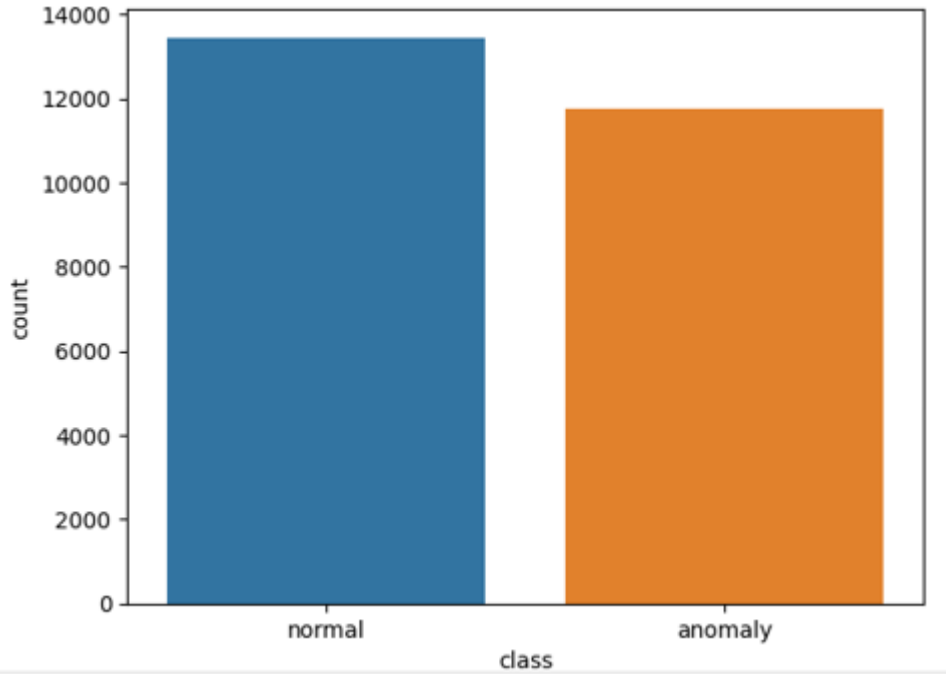


Figure 3: Dataset Representation

3.3 PREPROCESSING LAYER

The preprocessing layer is a critical component of data analysis that involves preparing and cleaning data for further analysis. In the context of IoMT-enabled healthacre data, the preprocessing layer is responsible for transforming the raw data collected from various medical sources, such as medical sensors, into a format that is suitable for analysis.

The preprocessing layer involves various techniques such as data cleaning, data integration, data transformation, and data reduction. Data cleaning involves removing any erroneous or missing data, while data integration involves combining data from multiple sources into a single dataset. Data transformation involves converting the data into a more usable form, such as changing the scale or normalizing the data. Data reduction involves selecting a subset of relevant data to reduce the size of the dataset and improve analysis performance.

By properly preprocessing healthcare data, the subsequent analysis will yield more accurate results, allowing healthcare professionals to make better-informed decisions to improve healthcare data flow and reduce intrusion.

The following steps describe the steps followed to preprocess the network flow metrics and biometric data:

1. Handling null values: In the dataset, there are few null values. We used a Jupyter notebook to check if there are null values in our dataset. The columns are transformed by deleting the null values and converting string to integer.

2. Handling zeros: There are a lot of zeros generated in the dataset especially in the attack traffic. The zeros values are left untouched as they are valid values produced during attack when the patient's biometric data are not able to be transported to the monitoring systems from the gateway.

3. Cross-validation folds: This includes splitting the original dataset into k equal parts (folds). It takes out one fold aside, and performs training over the rest k-1 folds and measures the performance. Then, repeats the process k times by taking different fold each time. Cross-validation is used to overcome the problem of overfitting and makes the predictions more general. Each fold is used for testing and for training.

3.4 APPLICATION LAYER

In the application layer of the this proposed research work, the utilization of XAI and ML techniques is expected to be a key driver of innovation and progress. XAI will be vital for ensuring that AI-powered systems are safe, reliable, and trustworthy, while ML techniques will enable these systems to analyze and make decisions based on vast amounts of data. To implement XAI in the IoMT, healthcare professionals need an AI system that is capable of providing transparent and understandable explanations of their decision-making processes. This require a deep understanding of the underlying algorithms and data, as well as the ability to present this information in a user-friendly format.

ML techniques is used extensively in the application layer of the IoV to analyze the vast amounts of data generated by sensors in vehicles and infrastructure. By training algorithms on this data, developers can create systems that are capable of making accurate predictions and decisions in real-time. To ensure that these decisions are transparent and understandable, healthcare professionals can incorporate XAI techniques such as feature importance analysis, visualizations, and other tools to help users understand the decision-making process. This is particularly important in the context of autonomous healthcare systems, where the decisions made by AI systems have significant impacts on safety and security of patients.

Similarly, the classification of data and the evaluation of result accuracy rely on a diverse array of machine learning models and algorithms. The ensuing section provides an in-depth exploration of certain models, elucidating their role in the testing and assessment of results.

The machine learning pipeline involves crucial stages, including data loading, preprocessing, training, performance evaluation with the implementation of XAI techniques to identify threats, and model implementation. Figure 4 visually represents these key phases of machine learning, incorporating the integration of the XAI ensemble technique of the voting classifier. A comprehensive explanation of each step in the implemented methodology is provided below.

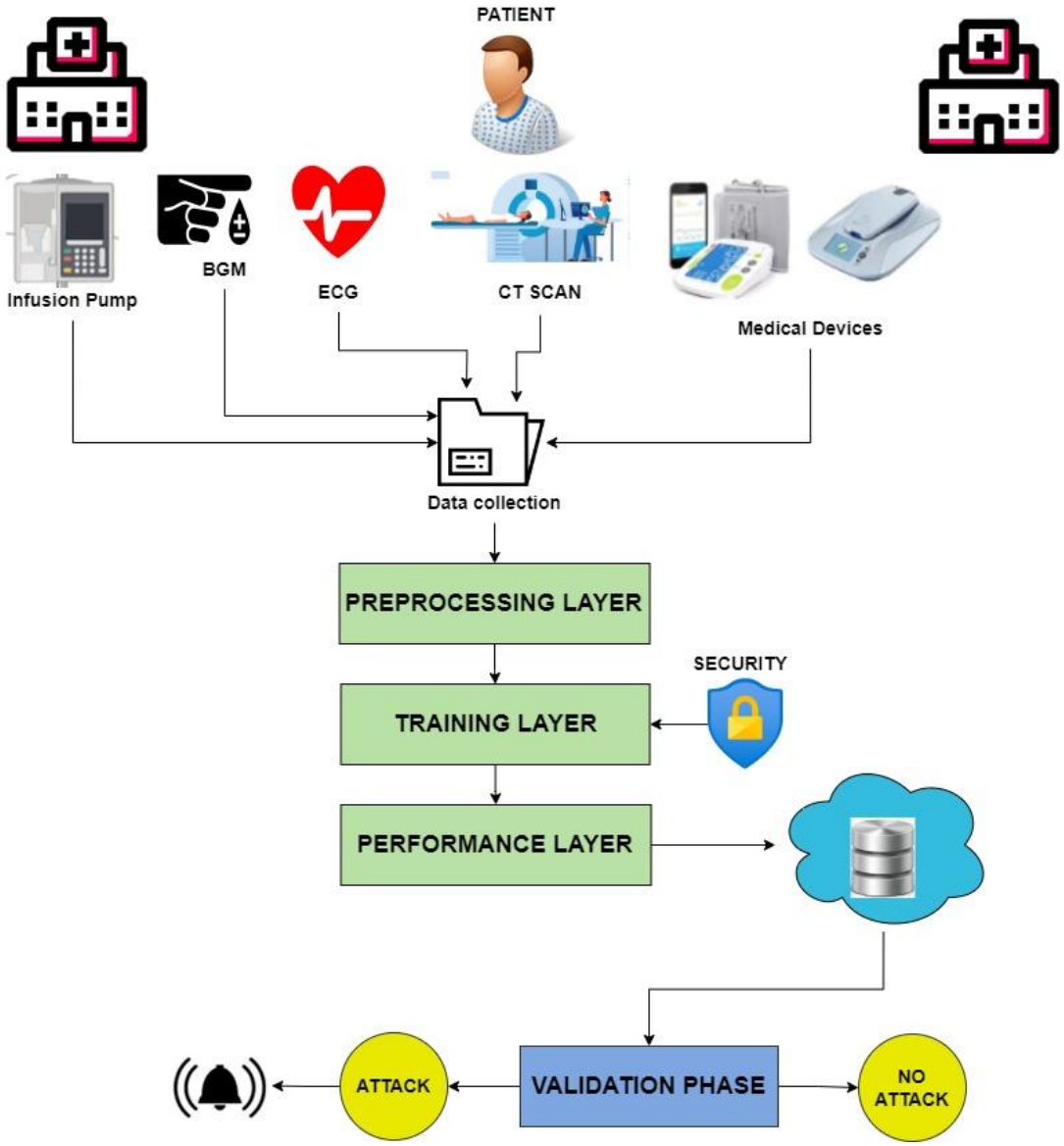


Figure 4: Pipeline Model Framework

3.4.1 Neural network

A particular class of machine learning algorithm called a neural network draws inspiration from the structure and functioning of the human brain. It is comprised of layer-organized, networked processing nodes.

3.4.2 Support-Vector Machine

For classification and regression applications, support vector machines (SVMs) are a kind of machine learning algorithm. By analyzing the data, they determine which hyperplane maximally divides the classes.

3.4.3 Decision Tree

Using a tree-like model to predict outcomes based on the correlations between characteristics in a dataset, decision trees are a particular kind of machine learning technique. The method it uses to arrive at a conclusion regarding the target variable for each group is to divide the data into progressively smaller groups according to the feature values.

3.4.4 Random Forest

A random forest is a form of ensemble learning technique that trains numerous decision trees on random subsets of data and then combines their predictions to generate a final prediction. When compared to employing a single decision tree, this can enhance the model's performance.

3.4.5 Nearest-Neighbor Chain Algorithm

Nearest neighbor is a form of machine learning algorithm that predicts a sample by finding the most similar samples in the training set and utilizing their labels to create a prediction.

3.4.6 Linear Regression

Linear regression is a machine learning approach that is used to simulate the linear connection between one or more independent variables and a dependent variable. It works by fitting the data to a straight line (or hyperplane

in higher dimensions) that minimizes the distance between the points and the line. This line can then be used to create predictions based on new data.

3.4.7 Logistic Regression

Logistic regression is a classification task-specific machine learning technique. It employs a logistic function to predict the likelihood that a sample belongs to a specific class. A logistic function is a mathematical function that converts any input value into a number between 0 and 1. This is important for classification problems since it allows us to interpret the model's output as a probability of belonging to a specific class. For example, if we want to know if an email is spam or not, the logistic regression model output may be read as the likelihood that the email is spam. If the likelihood exceeds a particular level, the email might be classified as spam.

3.5 AI MODELS

AI models are responsible for predicting patterns based on multiple ML algorithms like ANN, support vector machines, etc. In this proposed encryption based framework, decision tree algorithm is being used for the prediction. As we know that the line equation

Then the output of AI models is fed forward for the predictions. XAI methods use predictions and health data to generate explanations.

3.6 EXPLAINABLE ARTIFICIAL INTELLIGENCE (XAI)

XAI aids medical specialists in interpreting black-box models and their decision-making procedure to verify a specific decision taken by a ML model that is extremely important in the medical field. The primary goal of XAI is to built trust in through ML the importance of local and global variables using post-hoc explanations.

XAI is a research field that goals to make AI systems results more understandable to medical specialists. XAI focuses on the challenge of deciphering the mysteries of the black boxes, but it also implies Responsible AI because it can aid in creating transparent models. A human being can explain an interpretable system, and this is closely related to the idea of explainability. Enabling explainability in ML aims to make it easier for end-users and other stakeholders to understand the reasoning behind algorithmic decisions.

The model-agnostic interpretation methods, the current advancements in ML, are used to explain the complex models whereas retaining a good prediction performance. The model-agnostic interpretation is more flexible than the model-specific explanation method as it differentiates the model from the explanations. There are two types of model agnostic interpretation methods: local and global explanations.

Using LIME as a local explanation method is the most common. While PDP and SHAP are the most widely used interpretable approaches globally. Using LIME as a local explanation method, the local surrogate models are trained to be able to explain the complex models. First, LIME introduces a new dataset by perturbing the original data. Using the new dataset to train, it then builds an interpretable model, such as a decision tree. Finally, the black box model's prediction accuracy is compared to the interpretable model's accuracy. LIME is distinct as follows:

$$\gamma(\mathrm{x}) = \mathrm{argminL}\,(\mathrm{f}, \mathrm{g}, \pi\mathrm{x}) + \Omega(\mathrm{g})\mathrm{g}\epsilon\mathrm{G} \tag{3.1}$$

$$\widehat{f_{x_s}}(x_s) = \frac{1}{n}\sum_{i=1}^{n} \widehat{f_{x_s}}\left(x_s, x_c^i\right) \tag{3.2}$$

Where $\widehat{f_{x_s}}(x_s)$ is the partial function showing the global association of an input feature through the predicted result. s is a feature set comprising only one or two features, x_s represents the set of features is to be plotted by $\widehat{f_{x_s}}(x_s), x_c$ states the real feature standards from the dataset for the features we are not involved in. n is the number of instances of the dataset.

The Shapley values measure the impact of features on a complex model using SHAP. Shapley values are defined as the average of the marginal contributions. Concerning all possible coalitions, this feature value significantly impacts the prediction. Shapley value is represented as:

$$\varphi_j\,(x) = \sum_{s \subseteq \{x_1, x_2, \ldots, x_m\} \setminus \{x_j\}} \frac{|s|!\,(m - |s| - 1)!}{m!} (val(s \cup \{x_j\} - val(s))) \tag{3.3}$$

where $\varphi_j\ (x)$ is the Shapley value of x_j , x_j signifies a feature value, s is a feature subclass of the model, m shows the number of features, val is the prediction for feature standards in set s.

After predicting through the XAI, it is checked if the information concerning disease prediction is found or not. In the case of “No”, the AI models will be retrained, and so on, but in the case of “Yes”, the information concerning intrusion prediction is stored on a cloud database.

In the Validation phase, 30% data is tested from the XAI enabled dataset for validation. There are four phases implemented firstly the real time data is imported from trained data for prediction after that it passes through the validation layer to check the difference between the training and testing data. After that the data it import trained data for prediction to check that either the intrusion detection is founded or not. In the case of no, the process goes exit, whereas in the case of yes, the message is shown that intrusion detection is founded.

CHAPTER 4

SIMULATIONS AND RESULTS

In the current healthcare industry, formidable challenges are reshaping the industry's operational fabric and impacting patient care. Foremost among these challenges is the relentless threat of cybersecurity breaches, demanding heightened measures to protect the integrity and confidentiality of patient data. Simultaneously, concerns over data privacy loom large, navigating a delicate balance between compliance with stringent regulations and facilitating effective data-sharing for enhanced patient outcomes. The industry also grapples with interoperability issues, impeding seamless data exchange across healthcare entities and hindering collaborative initiatives. Resource constraints, both in terms of finances and personnel, contribute to operational inefficiencies, limiting the sector's ability to adopt advanced technologies. Moreover, global health crises, such as pandemics, amplify existing challenges, testing the resilience and adaptability of healthcare systems worldwide. These multifaceted challenges underscore the imperative for innovative solutions and strategic advancements to fortify the healthcare sector against the evolving threats and uncertainties it faces.

This transformative research introduced an advanced approach leveraging Explainable Artificial Intelligence (XAI) to fortify intrusion detection systems within healthcare cybersecurity. This innovative model serves as a sign, combining the expertise of healthcare professionals with the embedded knowledge within the XAI framework. Using decision tree as the driving force, the system is rigorously tested on a dataset sourced from a Kaggle repository, comprising 22544 instances. Segmented into training (70% or 14986 samples) and validation sets (30% or 7558 samples), this approach employs SVM to bolster intrusion detection capabilities. Evaluation metrics, computed with varying parameters delineated in equations 4.1 to 4.9, meticulously assess the model's performance. This pioneering method not only fortifies cybersecurity in healthcare but also exemplifies XAI's potential in combatting intricate threats, setting the stage for transformative advancements in healthcare cybersecurity infrastructure.

$$Sensitivity = \frac{\sum True\ Positive}{\sum Condition\ Positive} \tag{4.1}$$

$$Specificity = \frac{\sum True\ Negative}{\sum Condition\ Negative} \tag{4.2}$$

$$Accuracy = \frac{\sum True\ Positive + \sum True\ Positive}{\sum Total\ Population} \tag{4.3}$$

$$Miss - Rate = \frac{\sum False\ Negative}{\sum Condition\ Positive} \tag{4.4}$$

$$Fallout = \frac{\sum False\ Positive}{\sum Condition\ Negative} \tag{4.5}$$

$$Likelihood\ Positive\ Ratio = \frac{\sum True\ Positive\ Ratio}{\sum False\ Positive\ Ratio} \tag{4.6}$$

$$Likelihood\ Negative\ Ratio = \frac{\sum True\ Positive\ Ratio}{\sum False\ Positive\ Ratio} \tag{4.7}$$

$$Positive\ Predictive\ Value = \frac{\sum True\ Positive}{\sum\ Predicted\ Condition\ Positive} \tag{4.8}$$

$$Negative\ Predictive\ Value = \frac{\sum True\ Negative}{\sum\ Predicted\ Condition\ Negative} \tag{4.9}$$

Table 3
Training of the Proposed Intrusion Detection Model for IoMT using KNN

Proposed Model Training			
Input	**Total number of Samples (14986)**	**Result (output)**	
	Expected output	Predicted Positive	Predicted Negative
	11080 Positive	True Positive (T.P.)	False Positive (F.P.)
		10980	100
	3906 Negative	False Negative (F.N.)	True Negative (TN)
		6	3900

Table 3 is showing the proposed model in order to predict the intrusion using KNN during the training period. Throughout training, a total of 14986 samples are used, which are separated into 11080,3906 positive in addition negative samples, correspondingly. 10980 true positives are positively predicted, and no intrusion is identified, but 100 records are mistakenly

predicted as negatives, representing disease. Likewise, 3906 samples are acquired, with negative presenting intrusion in addition positive presenting no intrusion, with 3900 samples appropriately recognized as negative presenting intrusion also 6 samples mistakenly predicted as positive, demonstrating no intrusion in spite of the existence of the intrusion.

Table 4
Validation of the Proposed Intrusion Detection Model for IoMT using KNN

Proposed Model Validation			
Input	**Total number of samples (7558)**	**Result (output)**	
	Expected output	Predicted Positive	Expected output
	3498 Positive	True Positive (T.P.)	3498 Positive
		3435	63
	4060 Negative	False Negative (F.N.)	4060 Negative
		65	3995

Table 4 is showing the proposed model in order to predict the intrusion using KNN during the training period. Throughout training, a total of 7558 samples are used, which are separated into 3498,4060 positive in addition negative samples, correspondingly. 3435 true positives are positively predicted, and no intrusion is identified, but 63 records are mistakenly predicted as negatives, representing disease. Likewise, 4060 samples are acquired, with negative presenting intrusion in addition positive presenting no intrusion, with 65 samples appropriately recognized as negative presenting intrusion also 3995 samples mistakenly predicted as positive, demonstrating no intrusion in spite of the existence of the intrusion.

Table 5
Training of the Proposed Intrusion Detection Model for IoMT using LR

Proposed Model Training			
Input	**Total number of samples (14986)**	**Result (output)**	
	Expected output	Predicted Positive	Predicted Negative
	10200 Positive	True Positive (T.P.)	False Positive (F.P.)
		10100	100
	4786 Negative	False Negative (F.N.)	True Negative (TN)
		886	3900

Table 5 is showing the proposed model in order to predict the intrusion using LR during the training period. Throughout training, a total of 14986 samples are used, which are separated into 10200,4786 positive in addition

negative samples, correspondingly. 10100 true positives are positively predicted, and no intrusion is identified, but 100 records are mistakenly predicted as negatives, representing disease. Likewise, 4786 samples are acquired, with negative presenting intrusion in addition positive presenting no intrusion, with 3900 samples appropriately recognized as negative presenting intrusion also 886 samples mistakenly predicted as positive, demonstrating no intrusion in spite of the existence of the intrusion.

Table 6
Validation of the Proposed Intrusion Detection Model for IoMT using LR

Proposed Model Validation			
Input	**Total number of samples (7558)**	**Result (output)**	
	Expected output	Predicted Positive	Expected output
	3498 Positive	True Positive (T.P.)	3498 Positive
		3127	371
	4060 Negative	False Negative (F.N.)	4060 Negative
		210	3850

Table 6 is showing the proposed model in order to predict the intrusion using LR during the training period. Throughout training, a total of 7558 samples are used, which are separated into 3498,4060 positive in addition negative samples, correspondingly. 3127 true positives are positively predicted, and no intrusion is identified, but 371 records are mistakenly predicted as negatives, representing disease. Likewise, 4060 samples are acquired, with negative presenting intrusion in addition positive presenting no intrusion, with 3850 samples appropriately recognized as negative presenting intrusion also 210 samples mistakenly predicted as positive, demonstrating no intrusion in spite of the existence of the intrusion.

Table 7
Training of the Proposed Intrusion Detection Model for IoMT using DT

Proposed Model Training			
Input	**Total number of samples (14986)**	**Result (output)**	
	Expected output	Predicted Positive	Predicted Negative
	10200 Positive	True Positive (T.P.)	False Positive (F.P.)
		10100	100
	4786 Negative	False Negative (F.N.)	True Negative (TN)
		886	3900

Table 7 is showing the proposed model in order to predict the intrusion using DT during the training period. Throughout training, a total of 14986 samples are used, which are separated into 10200,4786 positive in addition negative samples, correspondingly. 10100 true positives are positively predicted, and no intrusion is identified, but 100 records are mistakenly predicted as negatives, representing disease. Likewise, 4786 samples are acquired, with negative presenting intrusion in addition positive presenting no intrusion, with 3900 samples appropriately recognized as negative presenting intrusion also 886 samples mistakenly predicted as positive, demonstrating no intrusion in spite of the existence of the intrusion.

Table 8
Validation of the Proposed Intrusion Detection Model for IoMT using DT

Proposed Model Validation			
Input	**Total number of samples (7558)**	**Result (output)**	
	Expected output	Predicted Positive	Expected output
	3498 Positive	True Positive (T.P.)	3498 Positive
		3484	14
	4060 Negative	False Negative (F.N.)	4060 Negative
		26	4034

Table 8 is showing the proposed model in order to predict the intrusion using DT during the training period. Throughout training, a total of 7558 samples are used, which are separated into 3498,4060 positive in addition negative samples, correspondingly. 3484 true positives are positively predicted, and no intrusion is identified, but 14 records are mistakenly predicted as negatives, representing disease. Likewise, 4060 samples are acquired, with negative presenting intrusion in addition positive presenting no intrusion, with 4034 samples appropriately recognized as negative presenting intrusion also 26 samples mistakenly predicted as positive, demonstrating no intrusion in spite of the existence of the intrusion.

Table 9
Proposed Model Performance using Different Statistical Measures

		Accuracy	Sensitivity TPR	Specificity TNR	Miss-Rate (%) FNR	Fall-out FPR	LR+	LR-	PPV (Precision)	NPV
KNN	**Training**	99.29	99.95	97.50	0.71	2.50	39.98	0.0072	99.10	99.85
	Validation	98.31	98.14	98.45	1.69	1.55	63.31	0.0171	98.20	98.40
LR	**Training**	93.42	91.94	97.50	6.58	2.50	36.77	0.0674	99.02	81.49
	Validation	92.31	93.71	91.21	7.69	8.79	10.66	0.0843	98.39	94.83
DT	**Training**	99.73	99.95	99.13	0.27	0.88	113.57	0.0027	99.68	99.85
	Validation	99.47	99.26	99.65	0.53	0.35	283.6	0.0053	99.60	99.36

Table 9 is the comparison of three algorithms K-Nearest Neighbors (KNN), Logistic Regression (LR), and Decision Trees (DT) Decision Trees consistently outperform the others in terms of performance and accuracy. Across both the training and validation datasets, Decision Trees exhibit the highest accuracy, sensitivity, specificity, and precision. The results indicate that Decision Trees are more adept at correctly classifying instances, making them a robust choice for the given dataset.

K-Nearest Neighbors (KNN) performs well, particularly in training, but exhibits a slight decrease in performance on the validation set. Logistic Regression (LR) shows reasonable accuracy but falls short compared to Decision Trees in key metrics. Decision Trees, with their ability to create a tree-like model, capture intricate patterns in the data, leading to superior performance.

The comparative analysis suggests that for this specific dataset, Decision Trees provide a more accurate and reliable solution. However, it's essential to consider the nature of the dataset, the complexity of the problem, and other factors when selecting the most suitable algorithm for a given task.

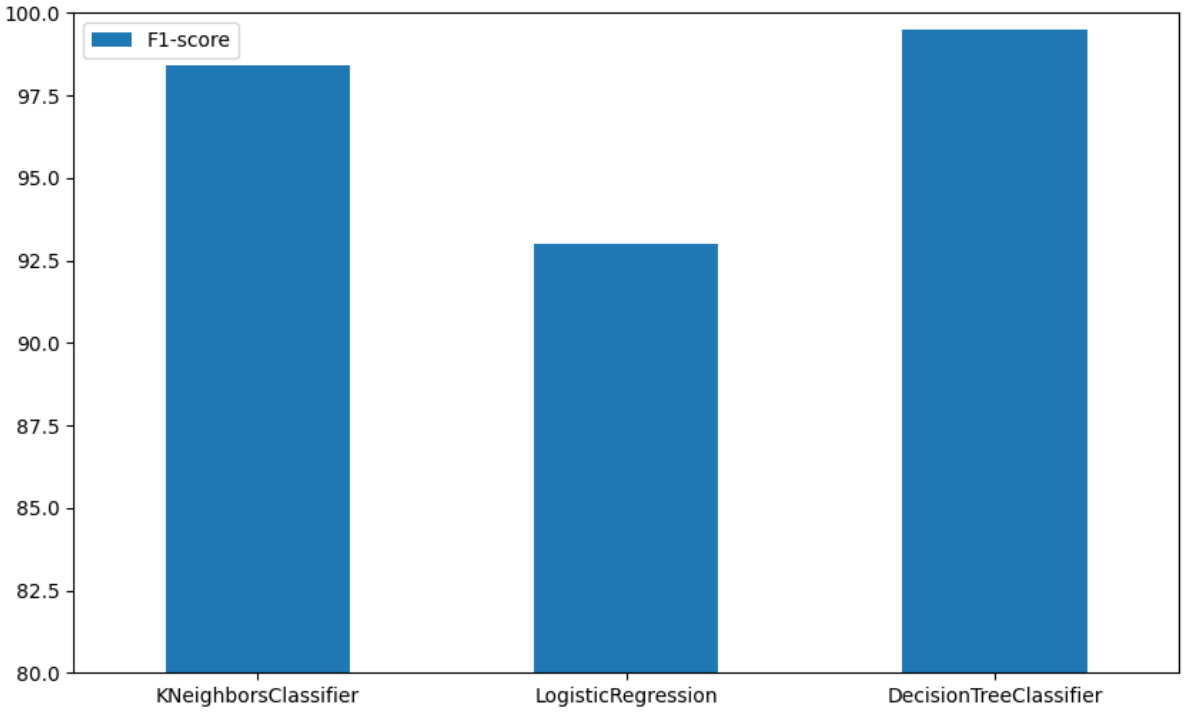


Figure-5: F1-Score Representation

Figure 5 is showing the analysis of three distinct algorithms, namely K-Nearest Neighbors (KNN), Logistic Regression (LR), and Decision Trees (DT), which were systematically applied to discern their performance in a comprehensive evaluation. Noticeably, the Decision Trees algorithm emerged

as the clear frontrunner, showcasing the highest F1-score among the trio. This result signifies that the Decision Trees algorithm excelled in stunning a balance between precision and recall, offering superior performance compared to both KNN and LR. The robustness and efficacy of Decision Trees, as highlighted by the elevated F1-score, underscore its potential as the algorithm of choice for the specific characteristics.

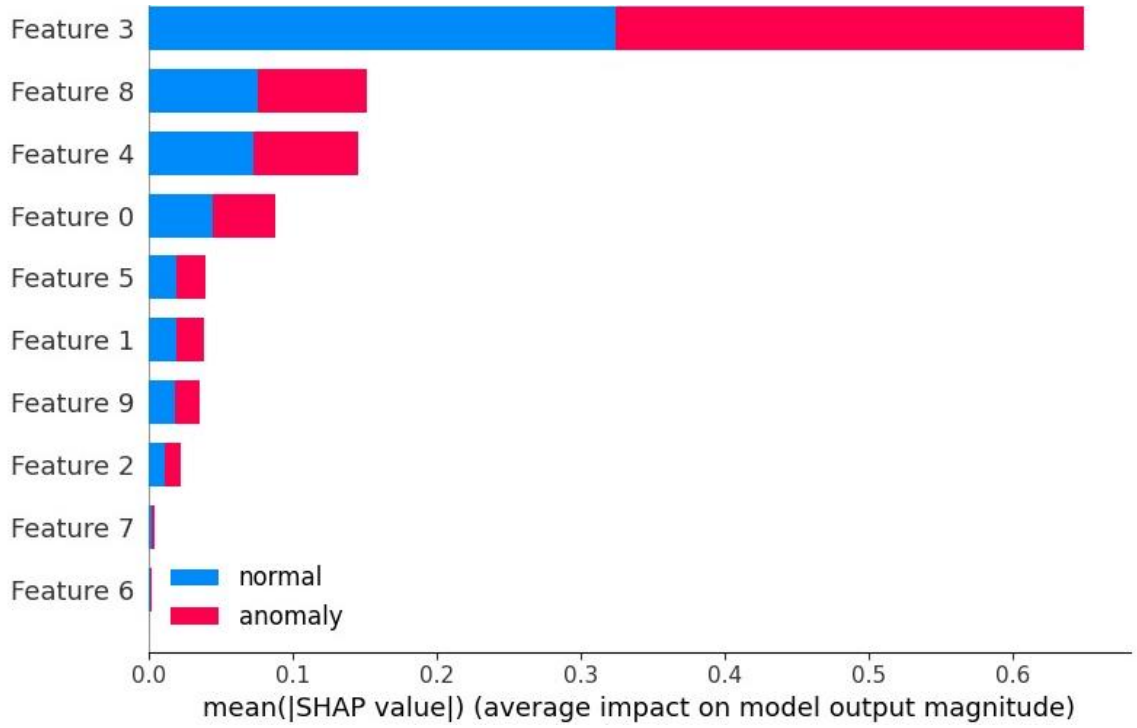


Figure 6: Graphical Representation of SHAP Value

The SHAP (SHapley Additive exPlanations) value graph in figure 6 provides valuable insights into the interpretability of machine learning models. By focusing on a specific prediction instance, SHAP values elucidate the impact of each feature on deviating from the baseline or average prediction. Positive and negative SHAP values indicate the direction and magnitude of influence, with high values signifying significant contributions. Identification of consistently impactful features across multiple predictions reveals their importance in driving model predictions. Exploration of interaction effects sheds light on synergies or conflicts between features. These global insights, coupled with real-world implications, enhance our understanding of the model's decision-making process. However, acknowledging limitations in SHAP value interpretation is crucial, and while providing valuable insights, these values may not capture all intricacies of complex models. In summary, SHAP values offer a powerful tool for interpreting machine learning models, facilitating meaningful insights for model users and stakeholders.

CHAPTER 5

CONCLUSION

In a time of rapid healthcare technological metamorphosis, the (XAI-HCIDS) framework emerges as symbol of trust, flexibility, and security. By increasing the clarity and transparency of AI-driven decisions, this research lays a robust foundation for the digital landscape of healthcare. Within the intricate challenges posed by cybersecurity in healthcare, the proposed approach covers the way for reinforcing defenses, encouraging trust, and securing the welfare of both patients and healthcare professionals.

REFERENCES

[1] Ahmad, I., Emad Ul Haq, Q., Imran, M., O. Alassafi, M. and A. AlGhamdi, R. (2022). An Efficient Network Intrusion Detection and Classification System. *Mathematics*, 10, 530.

[2] Alder, S. (2020). *Healthcare Data Breach Report*.

[3] Ardito, C., Di Noia, T., Sciascio, E., Lofù, D., Vitulano, F. and Pazienza, A. (2020). *An Artificial Intelligence Cyberattack Detection System to Improve Threat Reaction in e-health.* European Union, Horizon 2020 research and innovation programme, through ECHO project.

[4] Ashoor, A.S. and Gore, S. (2011). Importance of intrusion detection system (IDS). *International Journal of Scientific and Engineering Research*, 2(1), 1-4.

[5] Asif, M., Abbas, S., Khan, M.A., Fatima, A., Khan, M.A. and Lee, S.W. (2022). MapReduce based intelligent model for intrusion detection using machine learning technique. *Journal of King Saud University-Computer and Information Sciences*, 34(10), 9723-9731.

[6] Bandari, V. (2023). Enterprise data security measures: a comparative review of effectiveness and risks across different industries and organization types. *International Journal of Business Intelligence and Big Data Analytics*, 6(1), 1-11.

[7] Bhosale, K.S., Nenova, M. and Iliev, G. (2021). A study of cyber attacks: In the healthcare sector. *In 2021 Sixth Junior Conference on Lighting (Lighting)* (pp. 1-6). IEEE.

[8] Castonguay, C. (2021). *Why Study Healthcare Management Now?*

[9] Du, X., Chen, B., Ma, M. and Zhang, Y. (2021). *Research on the Application of Blockchain in Smart Healthcare: Constructing a Hierarchical Framework.* Journal of Healthcare Engineering.

[10] Du, X., Chen, B., Ma, M. and Zhang, Y. (2021). Research on the application of blockchain in smart healthcare: constructing a hierarchical framework. *Journal of Healthcare Engineering*, Available at: doi: 10.1155/2021/6698122

[11] Gopalan, S.S., Raza, A. and Almobaideen, W. (2021). *IoT Security in Healthcare using AI: A Survey.* 2020 International Conference on Communications, Signal Processing, and their Applications (ICCSPA), Sharjah, United Arab Emirates, 1-6.

[12] Ho, S., Jufout, S., Dajani, K., & Mozumdar, M. (2021). *A Novel Intrusion Detection Model for Detecting Known and Innovative Cyberattacks Using Convolutional Neural Network.* IEEE Open Journal of the Computer Society, 2, 14-25.

[13] Houda, Z.A., Brik, B. and Khoukhi, L. (2022). *"Why Should I Trust Your IDS?"*: An Explainable Deep Learning Framework for Intrusion Detection Systems in Internet of Things Networks. *IEEE Open Journal of the Communications Society*, 3, 1164-1176.

[14] Kabir, M., Fida Hasan, K., Hasan, M.K. and Ansari, K. (2022). *Explainable Artificial Intelligence for Smart City Application: A Secure and Trusted Platform. Explainable Artificial Intelligence for Cyber Security.* Studies in Computational Intelligence, vol 1025. Springer, Cham.

[15] Kasongo, S.M. (2023). A deep learning technique for intrusion detection system using a Recurrent Neural Networks based framework. *ELSEVIER Computer Communications,* 199, 113-125.

[16] Khan, I.A., Moustafa, N., Razzak, I., M. Tanveer, Tanveer, M., Pi, D., Ali, B.S. (2022). *XSRU-IoMT: Explainable simple recurrent units for threat detection in Internet of Medical Things networks.* Elsevier Future Generation Computer Systems, 181-193.

[17] Khan, I.A., Moustafa, N., Razzak, I., Tanveer, M., Pi, D., Pan, Y. and Ali, B.S. (2022). XSRU-IoMT: Explainable simple recurrent units for threat detection in Internet of Medical Things networks. *Future generation computer systems*, 127, 181-193.

[18] Khan, M.A., Abbas, S., Rehman, A., Saeed, Y., Zeb, A., Uddin, M.I., Nasser, N. and Ali, A. (2020). *A Machine Learning Approach for Blockchain-Based Smart Home Networks Security.* IEEE Network.

[19] Kumaar, M., Samiayya, D., Vincent, P.R., Srinivasan, K., Chang, C.Y. and Ganesh, H. (2022). *A Hybrid Framework for Intrusion Detection in Healthcare Systems Using Deep Learning.* Frontier. Public Health.

[20] Larriva-Novo, X., Villagrá, V., Vega-Barbas, M., Rivera, D. and Rodrigo, M.S. (2021). *An IoT-Focused Intrusion Detection System Approach Based on Preprocessing Characterization for Cybersecurity Datasets.* MDPI Sensors 21, 2, 656.

[21] Lee, C., Kim , T.W., cydenova, E. and Park, J.H. (2021). *Detection of Adversarial Attacks in AI-Based Intrusion Detection Systems Using Explainable AI.* Human-centric Computing and Information Sciences.

[22] Maseno, E., Wang, Z. and Xing, H. (2022). *A Systematic Review on Hybrid Intrusion Detection System.* Security and Communication Networks, Article ID 9663052, 23 pages.

[23] Medjahed, H., Istrate, D., Jerome, B., Baldinger, J.L. and Dorizzi, B. (2011). *A pervasive multi-sensor data fusion for smart home healthcare monitoring.* IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011), Taipei, Taiwan, 1466-1473.

[24] Moustafa, N., Koroniotis, N., Keshk, M., Zomaya, A. and Tari, Z. (2023). *Explainable Intrusion Detection for Cyber Defences in the Internet of Things: Opportunities and Solutions.* IEEE Communications Surveys & Tutorials.

[25] Otoum, Y. and Nayak, A. (2021). AS-IDS: Anomaly and Signature Based IDS for the Internet of Things. *Journal of Network and Systems Management,* 29, Article number: 23.

[26] Patil, S., Varadarajan, V., Mazhar, S.M., Sahibzada, A., Ahmed, N., Sinha, O., Kumar, S., Shaw, K. and Kotecha, K. (2022). Explainable artificial intelligence for intrusion detection system. *Electronics,* 11(19), 3079.

[27] Rehman, A. and Farrakh, A. (2023). Improving Clinical Decision Support Systems: Explainable AI for Enhanced Disease Prediction in Healthcare. *International Journal of Computational and Innovative Sciences,* 2(2), 9-23.

[28] Rehman, A., Farrakh, A., & Khan, S. (2023). Explainable AI in Intrusion Detection Systems: Enhancing Transparency and Interpretability. *International Journal of Advanced Sciences and Computing*, 2(1), 7-20.

[29] Sridhar, S. and Sivamohan, S. (2023). An optimized model for network intrusion detection systems in industry 4.0 using XAI based Bi-LSTM framework. *Neural Comput & Applic* 35, 11459–11475.

[30] Tsantikidou, K. and Sklavos, N. (2021). Vulnerabilities of Internet of Things, for Healthcare Devices and Applications. *8th NAFOSTED Conference on Information and Computer Science (NICS)*, Hanoi, Vietnam, 498-503.

[31] USTUN, T.S., Hussain, S., Yavuz, L. and Onen, A. (2021). Artificial Intelligence Based Intrusion Detection System for IEC 61850 Sampled Values Under Symmetric and Asymmetric Faults. *IEEE Access* 9, 56486-56495.

[32] Wang, M., Zheng, K., Yang, Y. and Wang, X. (2020). An Explainable Machine Learning Framework for Intrusion Detection Systems. *IEEE Access*, 8, 73127-73141.