People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

University of 8 May 1945-Guelma-

Faculty of Mathematics, Computer Science and Science of Matter

Department of Computer Science

Master Thesis

Specialty: Computer Science

Option: Science and Technology of Information and Communication

Theme

Enhancing Security in the Internet of Medical Things: A Machine Learning Approach for Intrusion Detection Systems

Presented by:

Manar Bensaada

Supervised by:

Dr Karima Benhamza

Jury Members:

Dr Djalila Boughareb

Dr Soumia Felkaoui

June 2024

Acknowledgments

Above all, I am profoundly grateful to God, the source of all wisdom and strength. It is through His grace that doors of opportunity have been opened and paths of knowledge have been illuminated. For the countless blessings and the unwavering guidance throughout my academic journey.

To my esteemed supervisor ***Dr. Benhamza Karima*** *I extend my heartfelt gratitude for her steadfast belief in my capabilities and her inspiring drive for excellence. Her mentorship has been both a privilege and an honor, profoundly shaping my academic journey and personal development.*

I would like to express my sincere gratitude to the members of the jury for dedicating their time and expertise to evaluate my work. Your insights and feedback are invaluable to the completion and improvement of this thesis.

To all those who have played a part, no matter how big or small, in shaping this journey, I extend my heartfelt appreciation.

Dedication

A heartfelt thank you to my extraordinary parents, the architects of my dreams and the guardians of my ambitions. Your unwavering support, endless sacrifices, and boundless love have been the cornerstone of my journey.

Mom, your nurturing spirit and unwavering determination have been my guiding light, while Dad, your wisdom and strength have been my pillars of strength. From sleepless nights to endless encouragement, you've been my rock through it all.

To my wonderful sister, thank you for patiently listening to my endless ramblings about my thesis, even when it made no sense to you. Your confused nods were my inspiration. I promise, no more machine learning talk—at least for now!

To my friends and classmates, thank you for making this year so memorable and full of laughter. From the endless jokes and memes to those epic Loup Garou games that seemed to never end, you've turned even the toughest days into something fun

. And to my friends, who offered words of wisdom like "Have you tried turning it off and on again?"—your technical support was... invaluable—we've made it a year to remembe.

Abstract

The Internet of Medical Things (IoMT) revolutionizes healthcare by enabling real-time monitoring and remote care but introduces significant security vulnerabilities. This thesis proposes an intrusion detection system (IDS) using the XGBoost algorithm to address these vulnerabilities. Utilizing the WUSTL-EHMS-2020 dataset, which integrates network flow metrics and patient biometric data, our model demonstrates superior performance with 99.11% accuracy, 98.05% recall, and a prediction time of 0.02 seconds. Comparative analysis with other machine learning models underscores the effectiveness of our approach. The results highlight the potential of ensemble learning methods in enhancing IoMT security, ensuring the protection of sensitive medical data and patient safety in interconnected healthcare environments

Keywords: Intrusion Detection, IoMT, WUSTL-EHMS-2020, XGBoost, Machine learning.

Résumé

L'Internet des Objets Médicaux (IoMT) révolutionne les soins de santé en permettant une surveillance en temps réel et des soins à distance, mais introduit d'importantes vulnérabilités de sécurité. Cette thèse propose un système de détection d'intrusion (IDS) utilisant l'algorithme XGBoost pour répondre à ces vulnérabilités. En utilisant le jeu de données WUSTL-EHMS-2020, qui intègre des métriques de flux réseau et des données biométriques de patients, notre modèle démontre une performance supérieure avec une précision de 99,11 %, un rappel de 98,05 %, et un temps de prédiction de 0,02 secondes. Une analyse comparative avec d'autres modèles d'apprentissage automatique souligne l'efficacité de notre approche. Les résultats mettent en évidence le potentiel des méthodes d'apprentissage ensembliste pour améliorer la sécurité de l'IoMT, assurant la protection des données médicales sensibles et la sécurité des patients dans des environnements de soins interconnectés.

Mots-clés: Détection d'intrusion, IoMT, WUSTL-EHMS-2020, XGBoost, Apprentissage automatique

ملخص

إنترنت الأشياء الطبية (IoMT) تُحدث ثورة في مجال الرعاية الصحية من خلال توفير الرصد في الوقت الفعلي والرعاية عن بُعد، لكنها تعرض النظام الصحي لثغرات أمنية هامة. تقدّم هذه الرسالة نظامًا لكشف الاختراقات (IDS) باستخدام خوارزمية XGBoost لمعالجة هذه الثغرات. من خلال استخدام مجموعة بيانات WUSTL-EHMS-2020، التي تجمع بين بيانات تدفق الشبكة وبيانات حيوية للمرضى، يُظهر نموذجنا أداءً متميزًا بدقة تبلغ 99.11٪، واسترجاع يبلغ 98.05٪، ووقت تنبؤ يبلغ 0.02 ثانية. يُسلط التحليل المقارن مع نماذج التعلم الآلي الأخرى الضوء على فعالية نهجنا. تُبرز النتائج الإمكانات الكامنة لطرق التعلم الجماعي في تعزيز أمان إنترنت الأشياء الطبية، وضمان حماية البيانات الطبية الحساسة وسلامة المرضى في بيئات الرعاية المتصلة.

الكلمات الرئيسية: كشف الاختراق، إنترنت الأشياء الطبية، مجموعة بيانات WUSTL-EHMS-2020، XGBoost، التعلم الآلي

Table of Contents

Table of Contents

List of Figures

List of Tables

List of Abbreviations

IoMT Internet of Medical Things
ML Machine Learning
IDS Intrusion Detection System
AI Artificial Intelligence
EHR Electronic Health Record
DL Deep Learning
DoS Denial of Service
DDoS Distributed Denial of Service
TPR True Positive Rate
FPR False Positive Rate
AUC Area Under the Curve
MLP Multi-Layer Perceptron
MITM Man-In-The-Middle
SVM Support Vector Machine
DT Decision Tree
RF Random Forest
NB Naive Bayes
ANN Artificial Neural Network
KNN k-Nearest Neighbor
CNN Convolutional Neural Network
IoT Internet of Things
LSTM Long Short-Term Memory
MLP Multi-Layer Perceptron
DNN Deep Neural Network
PCA Principal Component Analysis
LR Logistic Regression
RNN Recurrent Neural Network

Introduction

The Internet of Medical Things (IoMT) has emerged as a transformative technology in healthcare, offering real-time patient monitoring and remote data collection. It holds immense potential for managing chronic diseases, providing remote patient care, and facilitating early disease detection. For example, insulin pumps can automatically regulate blood sugar levels in diabetic patients, and pacemakers can deliver life-saving electrical shocks during heart arrhythmias. IoMT applications also include fall detection for the elderly, performance monitoring for athletes, and improved access to healthcare in underserved areas.

Moreover, IoMT generates vast amounts of detailed medical data, enabling more efficient treatments, reducing medical errors, and facilitating earlier disease identification. This translates to quicker interventions, improved patient outcomes, and cost savings for stakeholders such as insurance companies and healthcare providers. Remote access to medical data by healthcare professionals and families further enhances care delivery by eliminating unnecessary hospital visits and optimizing resource utilization.

However, despite the significant benefits of IoMT, its widespread adoption faces a major hurdle: security vulnerabilities. The reliance on wireless communication for data transmission from sensors to servers creates pathways for attackers to exploit. Compromised data confidentiality, integrity, and availability can lead to incorrect treatments and jeopardize patient safety. Conventional security solutions designed for typical IT networks are often inadequate for IoMT due to the inherent limitations of these devices, including resource constraints, diverse device types, and complex network configurations. While initial IoMT security efforts have employed

cryptography, authentication, and trust-based techniques, these methods struggle to address evolving and sophisticated threats targeting IoMT systems.

This growing concern for IoMT security necessitates the development of robust intrusion detection systems (IDS). Traditional signature-based IDS approaches are limited in their ability to detect novel attacks. Machine learning, particularly ensemble learning methods, offers a promising avenue for intrusion detection in IoMT networks. Ensemble learning models can learn complex data patterns and identify anomalies indicative of malicious activity.

The aim of this project is to propose and implement an IDS for IoMT networks using XGBoost, a powerful ensemble learning algorithm. We investigate the effectiveness of XGBoost in identifying and mitigating security threats within the IoMT ecosystem.

This thesis is structured into three main chapters:

Chapter 1 provides an overview of the Internet of Medical Things, including its architecture, challenges, and limitations and security threat it may face.

Chapter 2 reviews related work, examining the strengths and weaknesses of existing approaches.

Chapter 3 outlines the methodology and implementation details of our proposed machine learning-based intrusion detection system. We describe the dataset used for training and evaluation, discuss the preprocessing steps employed, and present the ensemble learning algorithms utilized. Furthermore, we evaluate and compare the performance of these algorithms, providing insights into their predictive capabilities and potential for clinical adoption.

Finally, the **conclusion** summarizes the key findings and proposes avenues for future research.

Chapter 1
Internet of Medical Things
(IoMT)

1.1 Introduction

The number of connected devices has been constantly increasing due to the arrival of new paradigms like the Internet of Things (IoT) and the evolution of communications systems. Different IoT applications and environments have impacted our daily activities. In this chapter, we focus on Internet of Medical Things (IoMT), providing a definition and their types and architecture, as well as the challenges and limitations and finally the security threat these devices.

1.2 What is IoMT

The Internet of Medical Things (IoMT) is a network of interconnected medical devices and sensors that can transmit real-time data to healthcare providers. By integrating IoMT, smart hospitals can collect and analyze patient data from various sources, including wearable devices, smart sensors, and medical equipment such as electrocardiography, electromyography, and electroencephalography. This data can help healthcare providers make informed decisions, improve patient outcomes, optimize clinical workflows, and even save lives [1].

1.3 The raise of IoMT

The healthcare landscape has undergone a digital revolution, transforming how we manage and utilize health data. The journey began in the 1990s with Healthcare 1.0, where doctors transitioned from paper records to digital entries stored in specialized systems. Healthcare 2.0 saw hospitals integrate data across individual doctors' computers. Healthcare 3.0 introduced the Electronic Health Record (EHR), creating a centralized repository of a patient's medical history.

Today, we're in the era of Healthcare4.0, witnessing the integration of artificial intelligence, big data analytics, and the Internet of Medical Things (IoMT), including sensors and wearables. This data empowers healthcare professionals with more accurate diagnoses, improved treatment decisions, and cost control insights. This ongoing evolution promises a future where healthcare is increasingly connected, efficient, and patient-centered, paving the way for the anticipated Healthcare 5.0, which aims to integrate emotional intelligence into patient care [3].

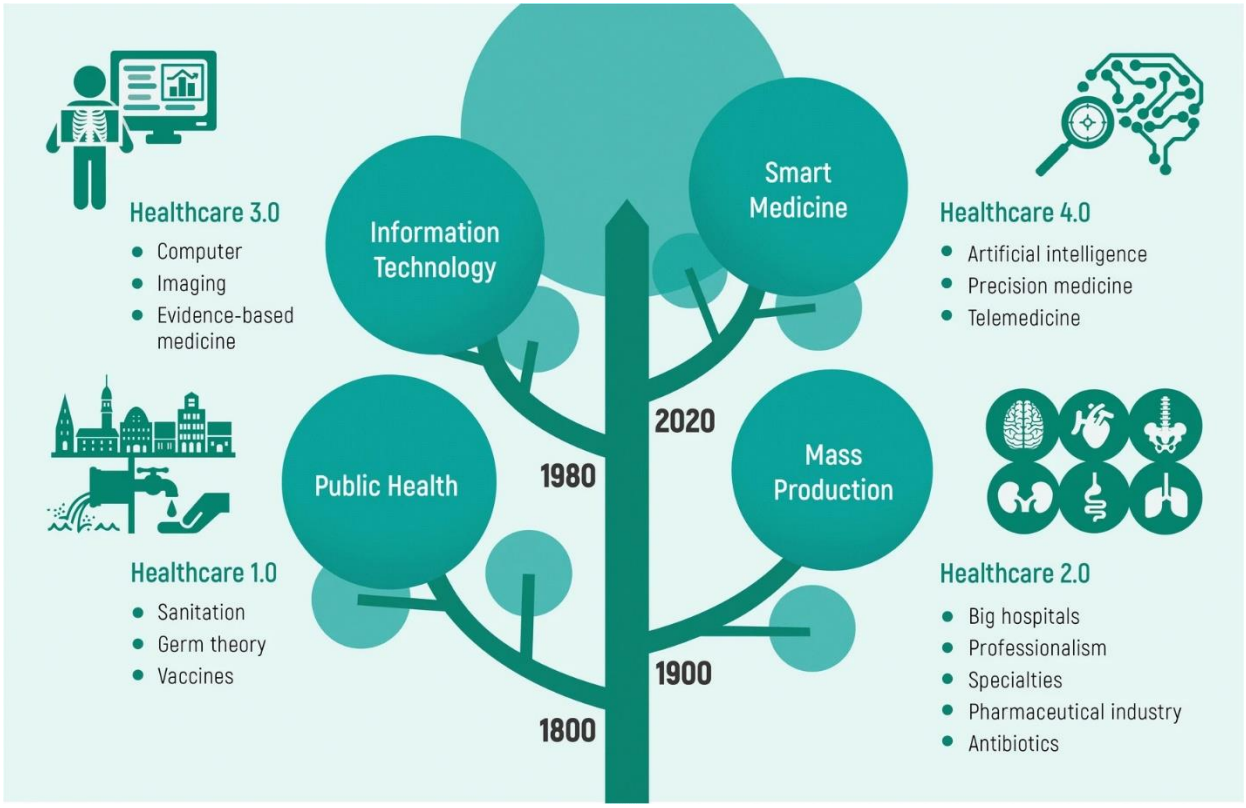

Figure 1. 1: *The groing trend of healthcare [3].*

1.4 IoMT Ecosystems

Smart healthcare systems are organized into various ecosystems, each characterized by specific types of IoMT devices [2,4,5], as illustrated in Figure 1.2 [6].

- The On-Body ecosystem is made up of wearable technologies such as fitness trackers and smartwatches, as well as implantable devices like glucose monitors, and even smart clothing with embedded sensors.
- In-Home ecosystem consists of portable medical equipment like blood pressure monitors, telemedicine devices for remote consultations, and smart medication dispensers.
- In-Clinic ecosystem pertains to outpatient medical services with devices like portable electrocardiogram monitors and patient check-in kiosks.
- In-Hospital ecosystem includes a collaborative environment of patients, healthcare professionals, and a diverse range of medical machinery, such as network-connected defibrillators, surgical tables, electrocardiogram machines, and smart beds that adjust settings for patient comfort and alert staff to patient needs.

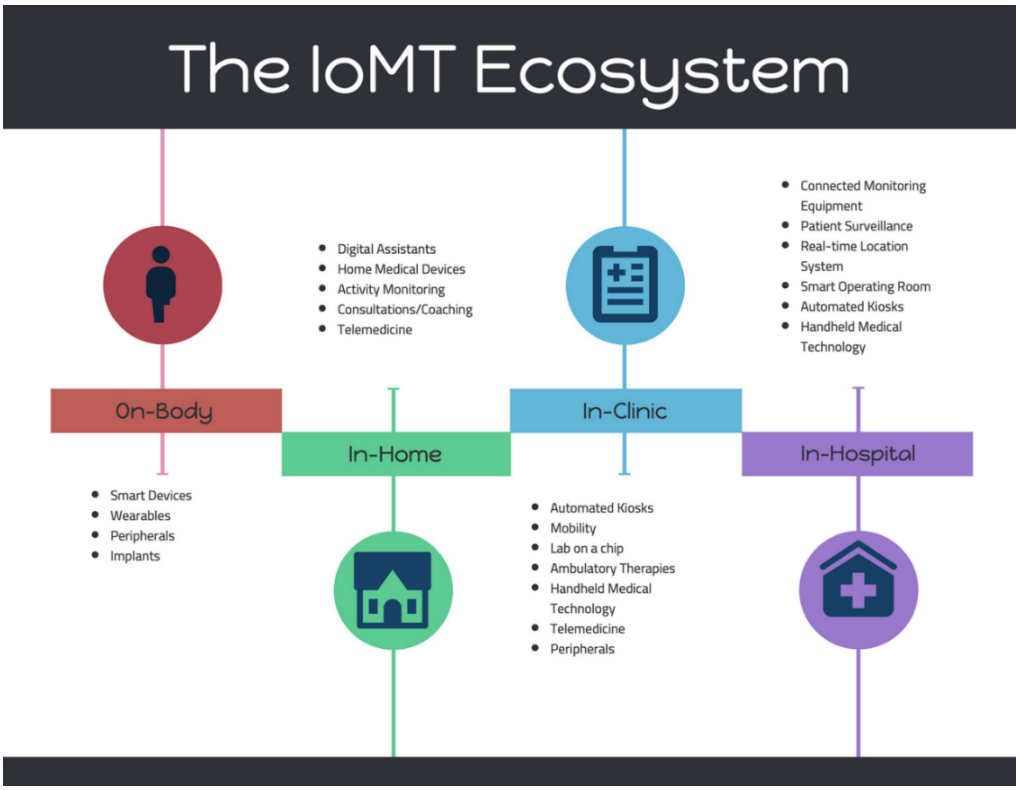

Figure 1. 2: *IoMT Ecosystem [6].*

1.5 Architecture of IoMT

The architecture of the IoMT can be described as a multi-layered framework, with each layer serving distinct functions and responsibilities [7-8], as depicted in Figure 1.3.

1.5.1 Perception Layer

This is the first layer, also known as the sensor layer. It consists of various medical devices and sensors that directly interact with the environment to collect data. Examples include wearable devices like heart rate monitors, glucose meters, and other health monitoring sensors.

1.5.2 Network Layer

The second layer is responsible for the transmission of the collected data. It uses communication protocols and network technologies to send the sensor data to the processing units. This layer ensures that data is securely and efficiently transmitted to the next layer for further processing.

1.5.3 Application Layer

The third layer is where the data is utilized to provide valuable insights and services. It includes applications and services that process and analyze the data to support healthcare decisions and patient care. This layer also encompasses the user interface through which healthcare providers and patients interact with the IoMT system.

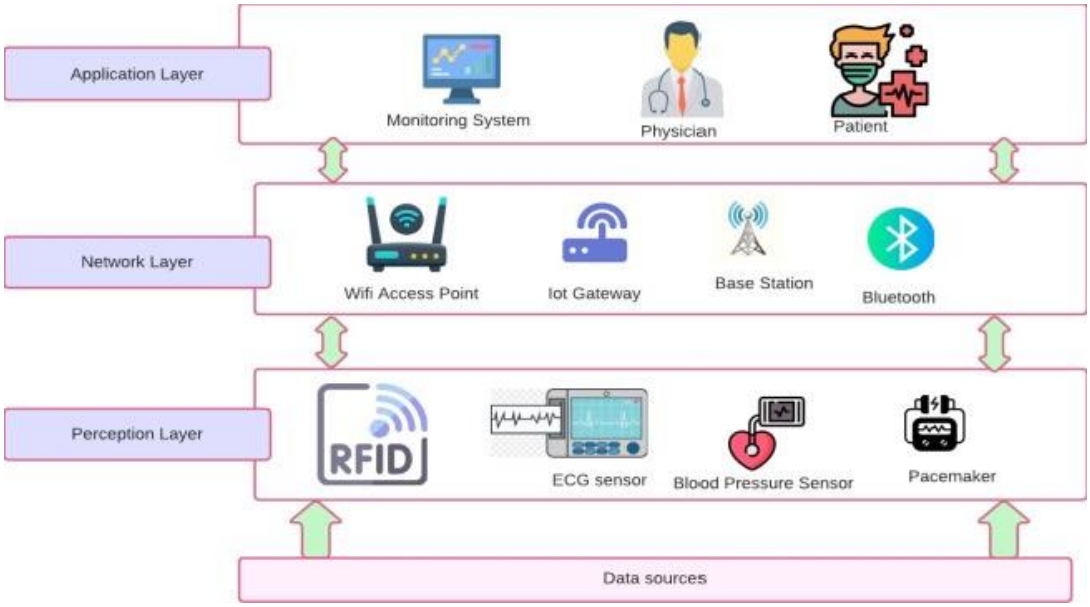

Figure 1. 3: *IoMT Architecture [8].*

1.6 Challenges and Limitations

Currently, the Internet of Medical Things (IoMT) is a rapidly growing field, but it does come with several limitations and challenges [36,37]:

- **Security:** Security in the Internet of Medical Things (IoMT) is a paramount concern due to the highly sensitive nature of health-related data. The primary security challenges include the risk of data breaches, which could lead to the exposure of personal health information, and the potential for device tampering, which can compromise the integrity of medical data and device functionality
- **Resource Limitations:** Unlike traditional computers, IoMT devices often have limited processing power and storage capacity. This restricts the implementation of robust security measures like complex encryption algorithms.
- **Interoperability:** Connecting devices from different vendors can be difficult. There's a need for a standardized approach to ensure seamless communication between various IoMT devices.
- **Scalability:** As the number of connected devices grows, the existing infrastructure may struggle to support the increased load. This includes limitations in computation capability and communication protocols.
- **Regulatory and Legal Issues**: The IoMT ecosystem involves many stakeholders, which can complicate the legal and regulatory landscape. Compliance with various laws and regulations is essential but can be difficult to navigate.

1.7 Security threat

Internet of Medical Things (IoMT) environments are susceptible to various security threats These threats target the core principles of cybersecurity [38], often referred to as the CIA triad Figure1.4:

Figure 1. 4: *CIA triad [9].*

- ✓ Confidentiality: Ensuring that sensitive patient data is accessed only by authorized individuals. Breaches of confidentiality can lead to privacy violations and misuse of personal health information.
- ✓ Integrity: Maintaining the accuracy and reliability of data. Attacks on integrity can result in altered patient records, leading to misdiagnoses or inappropriate medical interventions.
- ✓ Availability: Keeping IoMT services and data accessible to authorized users, especially for critical healthcare operations. Attacks that disrupt availability can have life-threatening consequences.

Understanding these threats is crucial for developing effective security measures and safeguarding the IoMT ecosystem. Here's a breakdown of common intrusion types that can face them:

- o **Man-in-the-Middle (MitM) Attacks:** An attacker positions itself between two communicating devices so he has the capability to surreptitiously monitor and record all the data exchanged. Furthermore, they possess the ability to manipulate the communication by injecting fraudulent messages, potentially altering the information relayed or decisions made based on that data [10].
- o **False Data Injection Attack:** This type of assault targets the integrity of data within IoMT networks. It involves the deliberate insertion of erroneous or manipulated data into the system. The aim is to disrupt the normal functioning of medical sensors and data analytics, leading to flawed operational decisions. Such attacks can have serious implications, potentially resulting in incorrect medical

interventions or diagnoses. [11]

- **Denial of Service (DoS) Attack:** This type of assault is executed with the

intent to disrupt the normal operations of a targeted system, server, or network by overwhelming it with a flood of internet traffic. The attacker's goal is to render the service inaccessible to its intended users. A Distributed Denial of Service (DDoS) Attack amplifies this strategy by launching the offensive from a multitude of sources, making it more challenging to mitigate and trace back to the origin [7].

- **Physical Tampering**: This can include modifying the device's hardware to

disrupt its functionality, exploiting firmware vulnerabilities to install malware, or altering the device's configuration settings gaining command over the device's functions [12]

- **Ransomware:** It refers to a type of malicious software designed to encrypt

sensitive data like patient records, rendering it inaccessible then demand a ransom payment in exchange for providing the decryption key to unlock the encrypted data. Ransomware attacks targeting IoMT systems can disrupt critical medical services.[12]

- **Eavesdropping**: also known as passive interception or sniffing, occurs when

an unauthorized party intercepts and listens to the data being transmitted between IoMT devices and the network. This can include sensitive patient information, authentication credentials, or other confidential communications. The attacker typically uses this information for malicious purposes, such as identity theft, fraud, or further network infiltration [12].

- **Jamming Attacks:** Jamming attacks specifically target the wireless

communication channels used by IoMT devices. By flooding the channel with interference or noise, attackers can disrupt the transmission of data, leading to a loss of service [7].

1.8 Conclusion

In this chapter, we provided a general introduction to IoMT, including its definition, the evolution of healthcare system, Ecosystem and architectures of IoMT, and some of their challenges and limitations and finally the security threat these devices. In recent years, machine learning has emerged as a promising tool for intrusion detection in medical IOT. In the next chapter, we will review the literature on security techniques for IoMT.

Chapter 2
Related Works

2.1 Introduction

The burgeoning field of the Internet of Medical Things (IoMT) necessitates sophisticated security measures to protect against cyber threats. This chapter gives an overview of available datasets that are essential for evaluating intrusion detection systems and then reviews existing security techniques for IoMT.

2.2 Available Datasets for IoMT

In IoMT datasets are collections of data that reflect real-world medical settings. To test IDS systems in healthcare, we need datasets that include a variety of medical devices and the unique types of data and sensors used in IoMT. These datasets should also address the special security and privacy issues that IoMT systems face. In this section, we explore various datasets available for IoMT [2].

- The WUSTL-EHMS-2020 dataset is a comprehensive collection created using an Enhanced Healthcare Monitoring System testbed, which includes network flow metrics and patients' biometrics available on [39]. It's designed to address the lack of datasets combining these elements [25].
- The ECU-IoHT dataset is a specialized collection designed to simulate various cyberattacks within an IoMT environment available on [40]. It was created to aid the healthcare security community in analyzing attack behavior and developing

robust countermeasures against vulnerabilities [24].

- The IoT-Flock framework features an open-source data generator tool for creating realistic IoT healthcare scenarios with both normal and malicious device traffic. It facilitates the generation of datasets for cybersecurity research, particularly for developing machine learning models to detect and prevent cyber-attacks in healthcare systems [41].

2.3 Literature Review on Security Techniques for IoMT

This thesis followed a structured research methodology involving a comprehensive database search, including IEEE Xplore, Springer, ScienceDirect, Elsevier, and MDPI. Keywords such as 'intrusion detection', 'IoMT security', 'IoMT security techniques', 'attack detection', 'machine learning', 'deep learning', 'smart healthcare', 'Internet of Medical Things', 'IoMT', 'Internet of Health Things', 'IoHT', and 'MIoT' were used. Publications from 2019 to 2024 that met the inclusion criteria were selected.

The literature review is categorized into five major security approaches: encryption and cryptography-based approaches, access control-based approaches, authentication-based approaches, intrusion detection-based approaches, machine learning-based approaches, and deep learning-based approaches.

2.3.1 Encryption

Encryption is a critical security measure in the IoMT realm, serving as the first line of defense against unauthorized access to sensitive medical data. It involves converting data into a coded format that can only be accessed and deciphered by individuals with the correct encryption key, ensuring that intercepted data remains unreadable and secure from potential cyber threats.

In this work, a novel security architecture for IoMT is introduced, leveraging ciphertext policy attribute-based encryption (CP-ABE). This architecture addresses the challenge of dynamic encryption in response to changing patient conditions during remote monitoring. New components hosted on IoMT gateways facilitate encryption and decryption, ensuring the security of medical data in e-healthcare systems [53].

2.3.2 Access control

The process of restricting access to IoMT devices and data based on user roles and permissions. This ensures that only authorized users can access specific data or perform certain actions on IoMT devices. A common access control model is Role-Based Access Control (RBAC), which assigns permissions based on pre-defined user roles (e.g., doctor, nurse, patient) within the healthcare system.

In this work [13] authors introduce D2DAC-IoMT, a certificate-based D2D access control scheme. Utilizing elliptic curve cryptography, D2DAC-IoMT offers enhanced security without compromising efficiency, as validated through formal and informal analyses.

2.3.3 Authentication

The process of verifying the legitimacy of a user or device attempting to connect to the IoMT network. This prevents unauthorized access and impersonation attacks. Multi-factor authentication (MFA) is a common approach, requiring users to provide multiple verification factors (e.g., password, fingerprint) to gain access. Digital certificates can also be used to establish trust between devices and servers within the IoMT network.

In the study conducted by Deebak et al. [14], an authentic-based privacy preservation protocol for smart e-healthcare systems in the Internet of Things (IoT) is proposed. The research introduces a Secure and Anonymous Biometric Based User Authentication Scheme (SAB-UAS) designed to ensure secure communication within healthcare applications. This scheme aims to address the privacy preservation issues in the IoMT by providing a solution that not only authenticates users but also maintains their anonymity.

2.3.4 Intrusion detection System

An Intrusion Detection System (IDS) is a security mechanism designed to monitor and analyze network traffic or system activities for signs of potential intrusions, which are unauthorized actions that threaten the confidentiality, integrity, or availability of information systems. It's a critical layer of defense in today's ever-evolving cybersecurity landscape [15-16].

a. Types of IDS in IoMT

- **Network-Based IDS (NIDS):** These systems monitor network traffic for suspicious activity and are typically deployed at strategic points within the network to monitor traffic to and from all devices on the network.

- **Host-Based IDS (HIDS):** These are installed on individual devices or hosts and monitor the inbound and outbound packets from the device only, providing protection to individual IoMT devices.

a. Detection Methods

- **Signature-Based Detection:** This method relies on predefined signatures of known threats. It's effective against known attacks but can't detect new, unknown threats.
- **Anomaly-Based Detection:** It creates a baseline of normal activity and uses machine learning or statistical methods to detect deviations from this norm, which could indicate an intrusion.

2.3.5 Machine Learning-Based IDS

Machine Learning-Based IDS is a system that leverage the power of machine learning algorithms to identify and respond to potential security threats within a network.

These systems leverage various machine learning techniques, such as supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning, and ensemble learning, to train predictive models capable of distinguishing between normal network behavior and anomalous or suspicious activities.

a. Supervised learning is a machine learning approach where a model is trained to make predictions or decisions using labeled data. In this method, the algorithm is given a dataset comprising input variables "features" and their corresponding output variables "labels". By examining the relationships in this labeled data, the model learns to associate the input features with the output labels [17]. The objective of supervised learning is to develop a model capable of accurately predicting or classifying new, unseen data based on the patterns learned from the training data. Common supervised learning algorithms used in ML-IDS include:

- **Support Vector Machines (SVM):** is primarily used for classification problems but can also be applied to regression models. It is a linear model that provides solutions for both linear and nonlinear problems. SVM operates on the concept of margin calculation. By creating hyperplanes that effectively separate the dataset into different classes [18].
- **Decision Trees (DT):** A decision tree is a flexible algorithm used for both classification and regression tasks. It creates a tree-like structure where each node

represents a decision based on a specific feature, and branches show the possible outcomes. This continues until a final decision or prediction is made at the leaf nodes. Decision trees are easy to understand and interpret, making them a popular choice for many machine learning applications [19].

- **Logistic Regression (LR):** is employed for binary classification tasks, wherein the outcome variable possesses two potential categories. It establishes a model that correlates the input variables with the likelihood of belonging to a specific class [20].

b. Unsupervised learning: is a machine learning approach where algorithms discern patterns and relationships within unlabeled data without predefined output labels or specific guidance. In contrast to supervised learning, no target variables or labels are provided to the algorithm. Instead, it autonomously explores the data, identifying inherent structures or patterns, grouping similar items, and minimizing dataset dimensionality. The objective is to uncover latent patterns and arrange the data into meaningful clusters without explicit instructions on what to search for like Clustering techniques [17].

c. Semi-supervised learning combines elements of both supervised and unsupervised learning, using a small amount of labeled data in conjunction with a larger unlabeled dataset to improve model performance.

d. Reinforcement learning (RL) is a branch of machine learning where an agent learns to make decisions by performing actions in an environment to achieve a reward. The agent improves its actions based on the feedback of rewards or penalties it receives, aiming to maximize the total reward over time [21].

e. Ensemble learning: Is a technique that combines multiple individual models, called base learners or weak learners, to create a more accurate and robust predictive model. The idea behind ensemble learning is that by combining the predictions of multiple models using a combination rule to obtain a single prediction, the ensemble model can overcome the limitations of individual models and achieve better performance [22]. Common ensemble techniques include:

- **Bagging (Bootstrap Aggregating):** Bagging involves training multiple models in parallel on different subsets of the training data, which are generated by random

sampling with replacement. The final prediction is made by averaging (for regression) or majority voting (for classification) across all models [22].

Random Forest (RF) is a classic example of a bagging method. It creates an ensemble of decision trees, each trained on a different subset of the data, and combines their predictions to produce the final output.

- **Boosting**: trains models sequentially, where each model attempts to correct the errors of its predecessor. Models are added iteratively, and each new model focuses on the examples that were misclassified by previous models Misclassified samples get more weight, causing the base learner to focus on such samples. The final model is a weighted sum of all individual models [22].

  Common Algorithms: XGBoost, AdaBoost, LightGBM.

- **Stacking:** involves training multiple different models (first-level models) and then using another model (second-level or meta-model) to combine their predictions. The meta-model is trained on the outputs of the first-level models to produce the final prediction [22].

- **Voting**: Voting is technique where multiple classifiers, each providing a prediction for the same input sample, have their predictions combined through a voting process to form a final output. The voting can be either 'hard', where the majority class label is selected, or 'soft', where the average probability across all classifiers determines the class label [23].

Several studies have applied these techniques to IoMT, leveraging machine learning to enhance security measures. The following are some important works:

In a recent paper [24], the authors developed a dataset named ECU-IoHT designed to analyze attacks targeting the Internet of Health Things (IoHT). This dataset was created by performing several types of attacks targeting a healthcare environment containing devices such as temperature sensor, blood pressure sensor and heart rate sensor.

In [25], Hady et al. designed an Enhanced Healthcare Monitoring System (EHMS) testbed that monitors the patients' biometrics data and collects network flow metrics. This system helped them to collect a dataset of 16 thousand records of normal and attack healthcare data.

In a study referenced in [26], the authors describe their development of a network intrusion detection model that employs a tree classifier.

Oladimeji [27], developed an intrusion detection system for the Internet of Medical Things, employing machine learning to differentiate between normal and attack traffic.

In this work [28], the authors propose an efficient and effective Anomaly-based Intrusion Detection System (AIDS) for IoMT networks. The system leverages host-based and network-based techniques, considering computational costs. By using machine learning algorithms, it identifies abnormalities and malicious incidents in the IoMT network.

In [29], Kilincer et al. proposed recursive feature elimination (RFE) and multilayer perceptron (MLP) model with 10-fold cross validation on different datasets.

In this study [30], the authors conducted a comparative analysis of various machine learning techniques for detecting intrusions in smart healthcare systems. They evaluated their proposed model using the UNSW-NB15 network intrusion benchmark dataset.

In [31], the authors propose a machine learning-based anomaly detection system and evaluate it using the TON IoT dataset. The study implements several machine learning algorithms, including Random Forest, Decision Tree, Logistic Regression, Support Vector Machine (SVM), and K-Nearest Neighbor, with both binary and categorical classifications.

2.3.6 Deep Learning-Based IDS

A Deep Learning-based Intrusion Detection System (IDS) utilizes neural networks and deep learning techniques to detect anomalies or cyber threats within network traffic or system behavior. These systems learn patterns from large amounts of data and can identify deviations from normal behavior, making them effective for securing complex environments like the Internet of Medical Things (IoMT). Now, let's proceed with the previous works in this category.

Ravi et al. [32] In their paper, the authors present a deep learning-based approach for intrusion detection in IoMT systems, utilizing features from network flows and patient biometrics.

In this research [33], the authors applied Principal Component Analysis (PCA) for feature reduction and utilized a multi-layer perceptron to effectively classify cyber-attacks on IoT-based healthcare devices.

Chaganti et al. [34] proposed a novel intrusion detection system that combines particle

swarm optimization with deep learning. Their methodology, termed PSO-DNN, was designed to implement an effective and accurate IDS within IoMT environments.

[35 In this paper, the authors propose SafetyMed, a novel Intrusion Detection System (IDS) that combines Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to protect Internet of Medical Things (IoMT) devices from cyber-attacks.

2.4 Evaluation metrics of IDS

When evaluating the effectiveness of Intrusion Detection Systems (IDS), several metrics are commonly used to assess their performance.

2.4.1 Prediction time

The duration an IDS takes to classify data, crucial for IoMT where immediate threat detection and response can be life-saving. It reflects the system's efficiency and suitability for real-time applications.

2.4.2 Confusion matrix

The confusion matrix is the best way to represent classification results so we can calculate the metrics

– True positive (TP): Intrusions that are successfully detected by the IDS.

– False positive (FP): Normal/non-intrusive behavior that is wrongly classified as intrusive by the IDS.

– True Negative (TN): Normal/non-intrusive behavior that is successfully labeled as normal/non-intrusive by the IDS.

– False Negative (FN): Intrusions that are missed by the IDS, and classified as normal/non-intrusive

ACTUAL	PREDICTED	
	Attack	Normal
Attack	TP	FN
Normal	FP	FN

Table 2. 1: *Confusion matrix.*

2.4.3 Accuracy

Measures the proportion of correctly identified instances (both attacks and normal activities) out of the total instances

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TP} + \boldsymbol{TN}}{\boldsymbol{TP} + \boldsymbol{TN} + \boldsymbol{FP} + \boldsymbol{FN}} \quad (2.1)$$

2.4.4 Recall

Recall can be defined as the proportion of true positives with respect to all the positives that exist in the ground truth.

$$\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP} + \boldsymbol{FN}} \quad (2.2)$$

2.4.5 Precision

Indicates the proportion of true positive alerts (correctly identified attacks) out of all alerts generated.

$$\boldsymbol{Precision} = \frac{\boldsymbol{TP}}{\boldsymbol{TP} + \boldsymbol{FP}} \quad (2.3)$$

2.4.6 F1-Score

Defined as a harmonic mean of precision and recall.

$$\boldsymbol{F1 - Score} = \frac{\boldsymbol{2 \times Precision \times Recall}}{\boldsymbol{Precision + Recall}} \quad (2.4)$$

A notable gap is the limited use of ensemble learning techniques, which are underutilized in the medical domain despite their potential benefits. This observation paves the way for the exploration of such methods, which may offer a promising avenue for enhancing the robustness and accuracy of intrusion detection systems. The introduction of an ensemble learning model in this research aims to fill this gap, potentially setting a new benchmark for performance in the field.

2.5 Conclusion

In this chapter, we delve into the existing research related to our topic—the enhancement of security in the Internet of Medical Things (IoMT). By conducting a comprehensive comparative analysis of various methodologies, we arrive at a compelling conclusion: machine learning-based intrusion detection systems (IDS) offer the most effective approach. Building upon this insight, the subsequent chapter will introduce our proposed model.

Chapter 3
Methodology and Implementation

3.1 Introduction

This chapter This chapter details the methodology for developing a machine learning-based IDS for IoMT. It covers dataset selection, exploratory analysis, the rationale for choosing XGBoost, system architecture, implementation details, performance evaluation, a comparative analysis and a discussion of the findings.

3.2 Dataset Description

The dataset utilized in this study, known as WUSTL-EHMS-2020, is a specialized collection of data designed for cybersecurity research within the Internet of Medical Things (IoMT) domain. It was generated from a real-time Enhanced Healthcare Monitoring System (EHMS) testbed, which is a composite system that captures both network flow metrics and patient biometrics. This integration is particularly rare and valuable due to the scarcity of datasets that encompass both types of data. Data flow starts from the sensors, passes through the gateway, and finally reaches the server for visualization. However, an attacker could potentially intercept this data before it reaches the server so we are in the transmission part [39].

Among the various IoMT datasets available, the WUSTL-EHMS-2020 dataset was chosen for its unique combination of network flow metrics and patient biometric data. This comprehensive integration provides a more realistic and holistic view of the network's security posture, making it particularly suitable for research in detecting sophisticated cyber threats.

Below are the key characteristics and statistical details of the WUSTL-EHMS-2020 dataset:

- **Features:**

  Network flow metrics (35 features).

  Patients' biometric features (8 features) [39].

- **Type of Attacks:** The dataset focuses on man-in-the-middle attacks [26], specifically:
  - Spoofing: Involves sniffing packets between the gateway and the server, compromising patient data confidentiality.
  - Data injection: Modifies packets on-the-fly, violating data integrity.

Measurement	Value
Dataset size	4.4 MB
Number of normal samples	14,272 (87.5%)
Number of attack samples	2,046 (12.5%)
Total number of samples	16,318

Table 3. 1: *Dataset statical informations [39].*

3.3 Exploratory Data Analysis (EDA)

The Exploratory Data Analysis (EDA) phase of our study was instrumental in uncovering the underlying structure and relationships within the "wustl-ehms-2020" dataset. This section details the methodologies employed in the EDA and the insights gleaned from this comprehensive analysis.

- Our initial focus was on count analysis to understand the distribution of data across various features. This included:
  - Total Count: The number of observations within each feature.
  - Uniqueness: The number of unique values present in each feature.
  - Top Value and Frequency: The most common values and their occurrence rates, for example looking for which addresses suffered the most attacks.
- Biometric Data Analysis: A key part of our EDA was the analysis of biometric data. We

scrutinized features such as heart rate, blood oxygen and temperature levels to:

- Understand the normal operating ranges and detect any deviations that might indicate anomalies.
- Investigate any relationships between biometric readings and the occurrence of attacks, which could suggest physiological responses to security breaches or system malfunctions.

The EDA not only informed our preprocessing and model selection but also provided a foundation for hypothesis generation.

3.4 Proposed model

3.4.1 Model selection

Initially, deep learning models were considered for the intrusion detection system due to their success in complex pattern recognition tasks. However, the limited size of the "wustl-ehms-2020" dataset, comprising 16,000 samples, posed a significant challenge. Deep learning models generally require large datasets to train effectively and avoid overfitting. Our experiments confirmed that with the available data, deep learning models were unable to achieve the desired level of performance.

Given the constraints of our dataset and due to the imbalance of the dataset, we pivoted to explore machine learning methods. Especially Ensemble Methods that are inherently better at handling imbalanced data. Among these methods, XGBoost emerged as a particularly promising candidate.

- **Why XGBoost?**

  XGBoost, or eXtreme Gradient Boosting, offers several compelling advantages that make it suitable for our application [42]:

  - **Regularization**: It includes regularization techniques that help prevent overfitting, which is particularly important when working with limited data
  - **Performance:** XGBoost has been shown to outperform other algorithms on tabular data, providing high accuracy and execution speed.
  - **Flexibility:** The algorithm allows for extensive hyperparameter tuning, enabling us to optimize the model specifically for our dataset.

- **Interpretability:** Unlike deep learning models, which can be seen as 'black boxes', XGBoost models are often more interpretable, allowing for a better understanding of feature importance and decision-making processes.

- **How XGBoost works?**

XGBoost is an ensemble learning method, which means it combines the predictions from multiple models (decision trees) to make a final prediction. At its core, XGBoost uses the gradient boosting framework. It starts with a base model and sequentially adds new models (trees) that correct the errors made by the previous ones. The process continues until a stopping criterion is met, such as the number of trees reaching a specified limit or the error reduction falling below a threshold [43-44]. shown in the Figure 3.1.

The XGBoost algorithm predicts the output as the sum of all tree results:

$$\hat{y}_i = \sum_{k=1}^{n} f_k(x_i), \quad f_k \in F \tag{3.1}$$

$\hat{y}_i$: is the predicted value of i[th] instance xi.

f_k : is the prediction of the k[th] tree.

x_i :is the feature vector of the i[th] instance.

The model's objective function is a combination of a differentiable loss function $L(\theta)$ and a regularization term $\Omega(\theta)$ that penalizes the complexity of the model, which helps to prevent overfitting:

$$Obj(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{3.2}$$

where $\hat{y}_i$ is the prediction and $y_i$is the target.

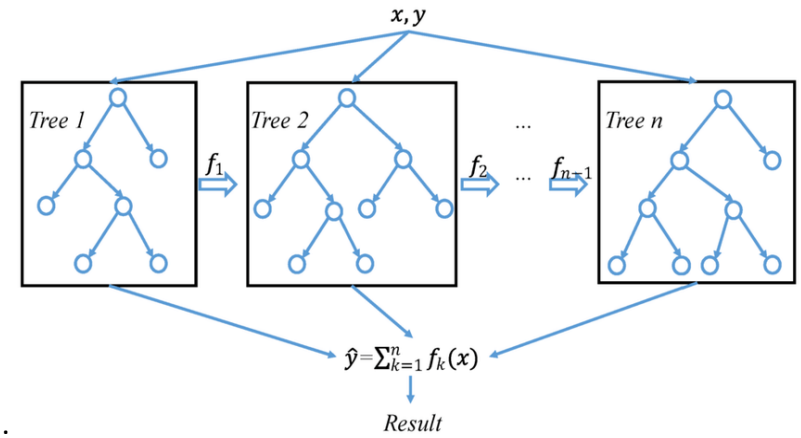

***Figure 3. 1**: XGBoost Architecture [44].*

3.4.2 Model architecture

The proposed model classifies input data into 'Attack' and 'normal' categories. To enhance performance and ensure reliability, feature scaling and selection techniques were employed, which streamline the processing by reducing the number of features and thus the overall detection time. The XGBoost model was finely tuned through hyperparameter optimization to achieve the best estimator, ensuring optimal performance. This approach ensures that the model operates effectively with limited computational resources, making it suitable for deployment at both node and network levels within the IoMT system.

3.5 Implementation

3.5.1 Runtime environment

The experiments were conducted on a system with the following specifications:

- Operating System: Windows 10 Professional.
- Processor: Intel Core i5-7440HQ CPU @ 2.80GHz.
- Memory: 16GB RAM.
- Storage: 256GB SSD.

3.5.2 Libraries

NumPy: (Numerical Python) is an open-source Python library extensively used in science and engineering. It provides multidimensional array data structures, like the homogeneous, N-dimensional ndarray, along with a comprehensive collection of functions that operate efficiently on these structures. [45].

Pandas: is an open-source Python library designed for efficient data manipulation and analysis. It offers robust data structures, such as DataFrames, to manage large and complex datasets. The library provides a variety of tools for data cleaning, transformation, and exploration, making it particularly useful for handling tabular data often found in scientific fields like social science and bioinformatics. [46].

Scikit-learn: or sklearn, a renowned machine learning library since its inception in 2007, offers a comprehensive suite of algorithms for various machine learning tasks, such as classification, regression, dimensionality reduction, and clustering. Beyond its algorithms, scikit-learn provides modules for essential tasks like data preprocessing, feature extraction, hyperparameter tuning, and model evaluation [47].

XGBoost: stands as a powerful library that advances the field of machine learning by offering an efficient and scalable implementation of gradient boosting algorithms. It is renowned for its performance and speed, which stems from its ability to execute parallel tree boosting. Often referred to by its acronyms GBDT or GBM, this library is adept at tackling a vast array of data science challenges with precision and rapidity. Its versatility is further highlighted by its compatibility with major distributed systems like Hadoop, SGE, and MPI, enabling it to handle extensive datasets that can reach into the billion [50].

TensorFlow is a powerful open-source machine learning framework developed by Google Brain for building and training deep learning models. It provides a comprehensive ecosystem of tools, libraries, and resources that facilitate the creation of neural networks, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs) [52].

Matplotlib: is a Python library that specializes in producing static, animated, and interactive visualizations in two dimensions. It's an essential tool for data analysis, allowing users to create a wide range of graphs and plots with high customizability, facilitating clear and detailed data representation [48].

Seaborn: is a Python visualization library based on Matplotlib that provides a high-level interface for drawing attractive and informative statistical graphics. It simplifies the process of creating complex visualizations and is particularly suited for exploring and understanding data patterns [49].

To implement the IDS based on XGBoost, we followed these steps:

3.5.3 Preprocessing

In order to build a highly accurate model, the preprocessing of the "wustl-ehms-2020" dataset was a critical step. The dataset, while rich in potential insights, required meticulous preparation to ensure the integrity and quality of the data fed into the machine learning models: Data Cleaning; Feature Selection and Data Splitting.

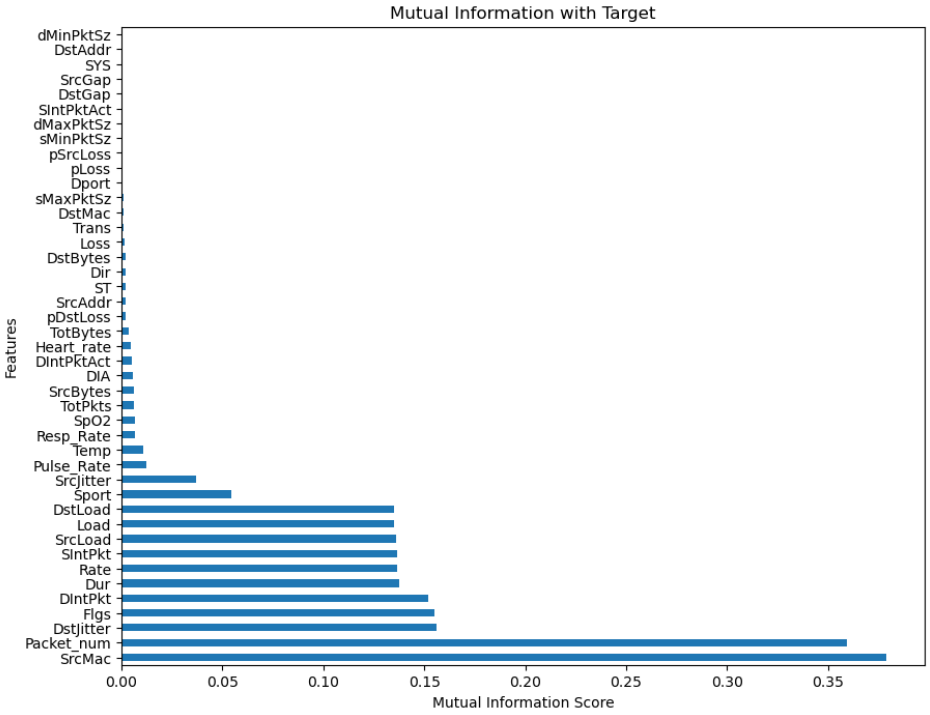

Figure 3. 2: *Mutal information with target.*

These preprocessing steps ensured that the dataset was clean, appropriately transformed, and ready for the machine learning models. This thorough preprocessing not only improves model performance but also ensures the reliability of the results.

3.5.4 Model building

1. The XGBoost model was initialized with default parameters using the XGBoost library [50] The model was then trained on the preprocessed training set, with iterative improvements made based on initial performance evaluation

PARAMETER	VALUES
Learning Rate	0.01, 0.1, 0.2
Maximum Depth	3, 6, 9
Subsample Ratio	0.8, 1.0
Column Subsample	0.8, 1.0
Regularization	1.0, 0.1
Regularization	0.0, 0.1
Number of Estimators	200, 300, 350

Table 3. 2: *Hyperparameters values for parameters model tuning.*

We Used different k values (3, 5, 7, 10) to perform k-fold cross-validation, ensuring a robust evaluation of the model's performance across different subsets of the data. This provided insights into the model's ability to generalize to unseen data. Finally, fitting the XGBoost model to ensure the model's robustness and generalizability.

3.6 Evaluation

To ensure a comprehensive assessment of our proposed XGBoost-based IDS, we will employ the evaluation metrics outlined in Section 2.3. These metrics, which include accuracy, precision, recall, F1-score, and time prediction, are standard benchmarks in the field of machine learning for classification tasks.

3.7 Results

The performance of the stacking ensemble model was further assessed on the test set to evaluate its generalization ability and robustness.

Figure 3.4 shows the confusion matrix, providing a detailed breakdown of the model's classification performance.

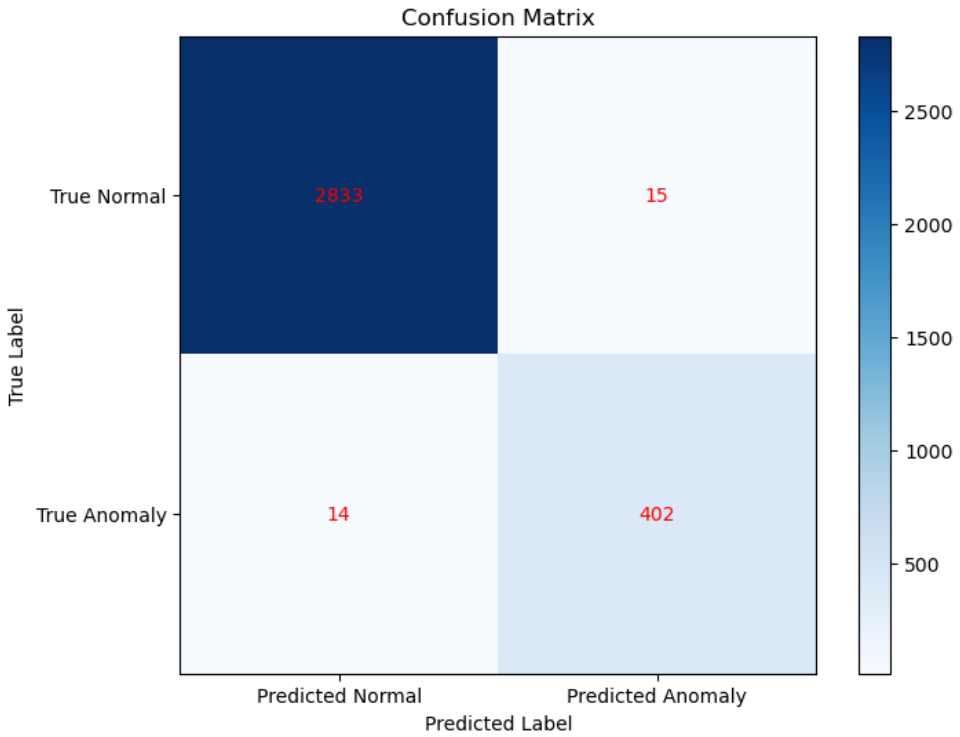

Figure 3. 3: *Confusion matrix.*

The following Table3.4 summarizes the performance metrics of the model.

METRICS	ACURACY	RECALL	PRECISION	F1-SCORE	TIME PREDICTION	TRAINING TIME
Values	99.11	98.05	97.95	98.00	0.02s	1.8s

Table 3. 3*: Performance metrics of proposed model*

These metrics collectively demonstrate the model's high accuracy and precision, along with its capability to recall a significant proportion of positive instances. The quick training and prediction times underscore the model's suitability for real-time applications.

Furthermore, the robustness of this model was validated during my participation in the AI 24 Day event organized by the University of Guelma, where I presented it as a workshop speaker [51]. The model garnered positive feedback from the academic community, reinforcing its potential for practical, real-world application. This external validation serves as a testament to the model's reliability and effectiveness in the field of IoMT security.

Additionally, a research paper describing this work and model was accepted at the 3RD INTERNATIONAL CONFERENCE ON FRONTIERS IN ACADEMIC RESEARCH [54], marking a significant milestone and demonstrating international recognition for the conducted research. The first paper of research paper is included in the annexes.

The Precision-Recall curve, shown in Figure 3.5, provides a comprehensive view of the model's performance across different thresholds

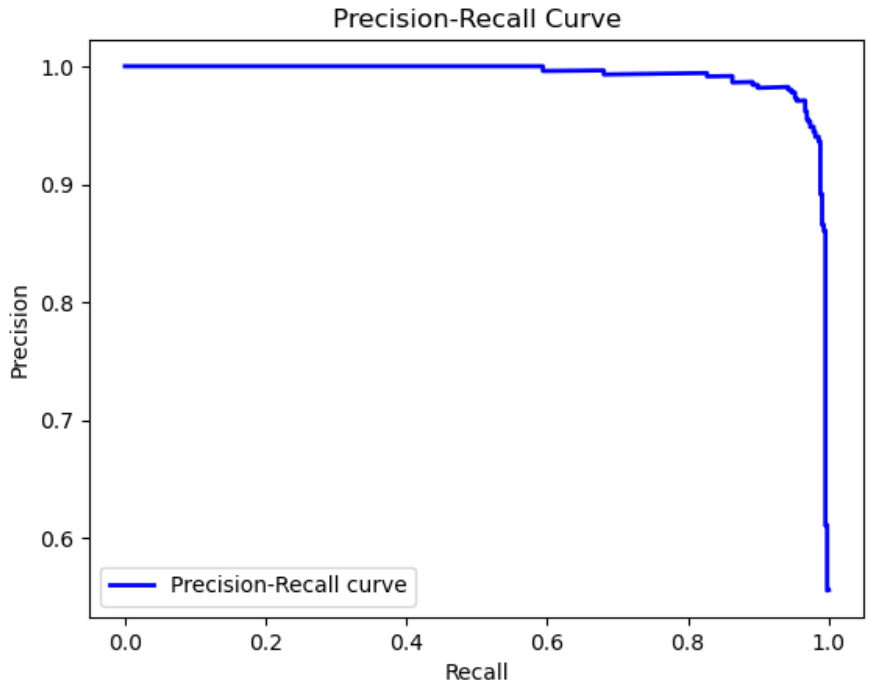

***Figure 3. 4**: Precision-Recall curve.*

The curve maintains a high level of precision across most recall levels, indicating that the model can identify true positives without a significant increase in false positives. A higher area under the curve (AUC) indicates better performance, with the model achieving high precision and recall simultaneously.

3.8 Conclusion

In this chapter, we detailed the methodology and implementation of an IDS for IoMT using the WUSTL-EHMS-2020 dataset. Our exploration revealed that machine learning methods, particularly XGBoost, are highly effective for intrusion detection in IoMT environments. Through meticulous preprocessing, model building, and evaluation, we achieved a high-performing IDS that balances accuracy, recall, and efficiency. The comparative analysis further underscored the robustness of our proposed model against other popular algorithms. This study demonstrates the critical role of advanced machine learning techniques in enhancing cybersecurity for interconnected medical devices, paving the way for safer and more reliable healthcare systems.

Conclusion

The Internet of Medical Things (IoMT) presents a transformative opportunity in healthcare, offering real-time patient monitoring and remote data collection. However, the widespread adoption of IoMT is hindered by significant security vulnerabilities, threatening patient safety and data integrity. Our research addresses this critical challenge by proposing an Intrusion Detection System (IDS) tailored to IoMT networks. Leveraging machine learning techniques, particularly ensemble learning with XGBoost, we developed a robust framework capable of identifying and mitigating security breaches in real-time.

The consistency of this thesis is a testament to its structured approach, where each chapter seamlessly builds upon the preceding one, contributing to a coherent narrative. Beginning with the exploration of IoMT's potential and vulnerabilities, the thesis progresses methodically to summarize previous work before delving into the detailed presentation of our solution.

Looking to the horizon, our future perspectives include refining the model to adapt to the ever-evolving landscape of cyber threats, expanding the dataset to encompass a broader spectrum of IoMT devices, and ultimately, deploying the IDS in real-world scenarios to validate its efficacy in real-time environments. The potential for this research to influence future developments in IoMT security is immense, and it is our hope that it will serve as a catalyst for further innovation in the field.

In conclusion, our thesis underscores the imperative of advanced machine learning techniques in safeguarding patient data and ensuring the integrity of IoMT networks. By proposing innovative solutions and identifying areas for future research, we contribute to the ongoing quest for secure and resilient healthcare systems in the digital age.

Bibliography

[1] Yuehong, Y. I. N., Zeng, Y., Chen, X., & Fan, Y. (2016). The internet of things in healthcare: An overview. Journal of Industrial Information Integration, 1, 3-13.

[2] Hernandez-Jaimes, Mireya Lucia & Martinez-Cruz, Alfonso & Ramírez Gutiérrez, K.A. & Feregrino, Claudia. (2023). Artificial intelligence for IoMT security: A review of intrusion detection systems, attacks, datasets and Cloud-Fog-Edge architectures. Internet of Things. 23. 1-82. 10.1016/j.iot.2023.100887.

[3] Chen, C., Loh, EW., Kuo, K.N. et al. (2020). The Times they Are a-Changin' – Healthcare 4.0 Is Coming!. J Med Syst 44, 40. https://doi.org/10.1007/s10916-019-1513-0

[4] https://ordr.net/article/iot-healthcare-examples Last access to the website: 08/05/2024

[5] https://www.helpwire.app/blog/internet-of-medical-things-healthcare/ Last access to the website: 08/05/2024

[6] Caldwell, Z.B. (2022). The Case for a Security Metric Framework to Rate Cyber Security Effectiveness for Internet of Medical Things (IoMT). In: Hudson, F.D. (eds) Women Securing the Future with TIPPSS for Connected Healthcare. Women in Engineering and Science. Springer, Cham. https://doi.org/10.1007/978-3-030-93592-4_4

[7] si-ahmed, Ayoub & Al-Garadi, Mohammed & Boustia, Narhimene. (2022). Survey of Machine Learning Based Intrusion Detection Methods for Internet of Medical Things.

[8] Das, Sunayana & Samal, Tushar & Mohanta, Bhabendu & Nayak, Aishwarya. (2023). Emerging Cyber Threats in Healthcare: A Study of Attacks in IoMT Ecosystems. 77-82. 10.1109/I-SMAC58438.2023.10290147.

[9] https://medium.com/@yadavtejas249/cia-and-dad-triad-ef84a94f9aee Last access to the website: 10/05/2024.

[10] Sanei, S., Jarchi, D. and Constantinides, A.G. (2020). Quality of Service, Security, and Privacy for Wearable Sensor Data. In Body Sensor Networking, Design and Algorithms https://doi.org/10.1002/9781119390060.ch13

[11] Ferrag, M. A., Shu, L., & Choo, K. R. (2021). Fighting COVID-19 and future pandemics with the Internet of Things: Security and privacy perspectives. IEEE/CAA Journal of Automatica Sinica, 8(9), 1477-1499. https://doi.org/10.1109/JAS.2021.1004087

[12] What security threats are targeting IoMT devices (and how to prevent being hacked) | Nuspire Last access to the website: 10/05/2024.

[13] Chaudhry, S. A., Irshad, A., Nebhen, J., Bashir, A. K., Moustafa, N., Al-Otaibi, Y. D., & Zikria, Y. B. (2021). An anonymous device to device access control based on secure certificate for internet of medical things systems. Sustainable Cities and Society, 75, 103322. https://doi.org/10.1016/j.scs.2021.103322.

[14] Deebak, B. D., Al-Turjman, F., Aloqaily, M., & Alfandi, O. (2019). An authentic-based privacy preservation protocol for smart e-healthcare systems in IoT. IEEE Access, 7, 135632-135649. https://doi.org/10.1109/ACCESS.2019.2941575

[15] What is an Intrusion Detection System (IDS)? | IBM. Qu'est-ce qu'un système de détection d'intrusion (IDS) ? | IBM Last access to the website: 20/05/2024.

[16] Intrusion Detection System (IDS) – GeeksforGeeks. Intrusion Detection System (IDS) - GeeksforGeeks Last access to the website: 20/05/2024

[17] Baştanlar, Y., Özuysal, M. (2014). Introduction to Machine Learning. In: Yousef, M., Allmer, J. (eds) miRNomics: MicroRNA Biology and Computational Analysis. Methods in Molecular Biology, vol 1107. Humana Press, Totowa, NJ. https://doi.org/10.1007/978-1-62703-748-8_7

[18] Jakkula, V. (2006). *Tutorial on support vector machine (SVM)*. School of EECS, Washington State University.

[19] https://www.ibm.com/topics/decision-trees Last access to the website: 08/06/2024.

[20] David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant. (2013). Applied logistic regression, volume 398. John Wiley & Sons.

[21] AlMahamid, F., & Grolinger, K. (2021). Reinforcement learning algorithms: An overview and classification. 2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), 1-7. https://doi.org/10.1109/CCECE53047.2021.9569056.

[22] Mienye, I. D., & Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. IEEE Access, 10, 99129-99149. https://doi.org/10.1109/ACCESS.2022.3207287

[23] Shiplu, A.I., Rahman, M.M., Watanobe, Y. (2024). A Robust Ensemble Machine Learning Model with Advanced Voting Techniques for Comment Classification. In: Sachdeva, S., Watanobe, Y. (eds) Big Data Analytics in Astronomy, Science, and Engineering. BDA 2023. Lecture Notes in Computer Science, vol 14516. Springer, Cham. https://doi.org/10.1007/978-3-031-58502-9_10.

[24] Ahmed, M., Byreddy, S., Nutakki, A., Sikos, L. F., & Haskell-Dowland, P. (2021). ECU-IoHT: A dataset for analyzing cyberattacks in Internet of Health Things. Ad Hoc Networks, 122, 102621. https://doi.org/10.1016/j.adhoc.2021.102621.

[25] Ahady, Anar & Ghubaish, Ali & Salman, Tara & Ünal, Devrim & Jain, Raj. (2020). Intrusion Detection System for Healthcare Systems Using Medical and Network Data: A Comparison Study. IEEE Access. PP. 10.1109/ACCESS.2020.3000421.

[26] Gupta, K., Sharma, D. K., Datta Gupta, K., & Kumar, A. (2022). A tree classifier based network intrusion detection model for Internet of Medical Things. Computers & Electrical Engineering, 102, 108158. https://doi.org/10.1016/j.compeleceng.2022.108158.

[27] Deborah Oladimeji. (2021). An intrusion detection system for internet of medical things. Dalhousie University. URL: https://dalspace.library.dal.ca/handle/10222/80561.

[28] Zachos, G., Essop, I., Mantas, G., Porfyrakis, K., Ribeiro, J. C., & Rodriguez, J. (2021). An anomalybased intrusion detection system for internet of medical things networks. Electronics, 10(21), 2562. https://doi.org/10.3390/electronics10212562

[29] Firat Kilincer, I., Ertam, F., Sengur, A., Tan, R.-S., & Rajendra Acharya, U. (2023). Automated detection of cybersecurity attacks in healthcare systems with recursive feature elimination and multilayer perceptron optimization. Biocybernetics and Biomedical Engineering, 43(1), 30–41. https://doi.org/10.1016/j.bbe.2022.11.005.

[30] Basharat, Asma & Mohamad, Mohd & Khan, Attiya. (2022). Machine Learning Techniques for Intrusion Detection in Smart Healthcare Systems: A Comparative Analysis. 29-33. 10.1109/ICSSA54161.2022.9870973.

[31] Reji, Alan & Pranggono, Bernardi & Marchang, Jims & Shenfield, Alex. (2023). Anomaly Detection for the Internet-of-Medical-Things. 1944-1949. 10.1109/ICCWorkshops57953.2023.10283523.

[32] Ravi, Vinayakumar & Pham, Tuan & Alazab, Mamoun. (2023). Deep Learning-Based Network Intrusion Detection System for Internet of Medical Things. IEEE Internet of Things Magazine. 6. 10.1109/IOTM.001.2300021.

[33] Judith, A., Kathrine, G. J. W., Silas, S., & J, A. (2023). Efficient deep learning-based cyber-attack detection for Internet of Medical Things devices. *Engineering Proceedings, 59*, 139. https://doi.org/10.3390/engproc2023059139

[34] Chaganti, R., Mourade, A., Ravi, V., Vemprala, N., Dua, A., & Bhushan, B. (2022). A particle swarm optimization and deep learning approach for intrusion detection system in the internet of medical things. Sustainability, 14(19), 12828. https://doi.org/10.3390/su141912828

[35] F Faruqui, N., Yousuf, M. A., Whaiduzzaman, M., Azad, A., Alyami, S. A., Liò, P., Kabir, M. A., & Moni, M. A. (2023). SafetyMed: A novel IoMT intrusion detection system using CNN-LSTM hybridization. Electronics, 12(35), 3541. https://doi.org/10.3390/electronics12173541

[36] https://cifs.health/backgrounds/internet-of-medical-things-challenges-and-adoptions/ Last access to the website: 29/05/2024.

[37] Shafiq, M., Choi, J. G., Cheikhrouhou, O., & Hamam, H. (2023). Advances in IoMT for healthcare systems. Sensors (Basel), 24(1), 10. https://doi.org/10.3390/s24010010.

[38] Kumar, G., Singh, O.P., & Saini, H. (Eds.). (2021). Cybersecurity: Ambient Technologies, IoT, and Industry 4.0 Implications (1st ed.). CRC Press. https://doi.org/10.1201/9781003145042

[39] https://www.cse.wustl.edu/~jain/ehms/index.html Last access to the website: 01/06/2024.

[40] "ECU-IoHT" by Mohiuddin Ahmed, Surender Byreddy et al. https://ro.ecu.edu.au/datasets/48/. Last access to the website: 01/06/2024.

[41] Hussain, F., Abbas, S. G., Shah, G. A., Pires, I. M., Fayyaz, U. U., Shahzad, F., Garcia, N. M., & Zdravevski, E. (2021). A framework for malicious traffic detection in IoT healthcare environment.

Sensors, 21(9), 3025. https://doi.org/10.3390/s21093025.

[42] https://www.geeksforgeeks.org/xgboost/ Last access to the website: 10/06/2024.

[43] Chen, Tianqi & Guestrin, Carlos. (2016). XGBoost: A Scalable Tree Boosting System. 785-794. 10.1145/2939672.2939785

[44] Wang, Yuanchao & Pan, Z. & Zheng, J. & Qian, L. & Mingtao, Li. (2019). A hybrid ensemble method for pulsar candidate classification. Astrophysics and Space Science. 364. 10.1007/s10509-019-3602-4.

[45] https://numpy.org/doc/stable/user/whatisnumpy.html Last access to the website: 12/06/2024.

[46] McKinney, W. et al. (2010). Data structures for statistical computing in python. In Proceedings of the 9th Python in Science Conference, 51-56. Austin, TX.

[47] Hackeling, G. (2014). Mastering machine learning with scikit-learn. Packt Publishing.

[48] Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3), 90-95.

[49] Nongthombam, K., & Sharma, D. (2021). Data analysis using python. International Journal of Engineering Research & Technology (IJERT).

[50] https://xgboost.readthedocs.io/en/stable/ Last access to the website: 18/06/2024.

[51] Bensaada, M. (2024). Intrusion detection for Internet of Medical Things using machine learning. Paper presented at AI 24 DAY, University 08 May 1945, Guelma.

[52] TensorFlow Last access to the website: 18/06/2024.

[53] Ghanbarafjeh, M., Barati, M., Rana, O., & Ranjan, R. (2022). Developing a Secure Architecture for Internet of Medical Things Using Attribute-Based Encryption. In *2022 IEEE/ACM 15th International Conference on Utility and Cloud Computing (UCC)* (pp. 157-162). Vancouver, WA, USA. doi: 10.1109/UCC56403.2022.00028.

[54] https://www.icfarconf.com/