



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

AI-Driven Intrusion Detection Framework for Healthcare Cybersecurity

by

Mahabub Alam Tuhin

Supervisor: Dr Dipo Dunsin

Project submitted as part of the requirements for the award of MSc in Cybersecurity & Digital Forensics

February 2026

DECLARATION

I, Mahabub Alam Tuhin declare that I am the sole author of this Project that all references cited have been consulted; that I have conducted all work of which this is a record, and that the finished work lies within the prescribed word limits.

This work has not previously been accepted as part of any other degree submission.

Signed: Mahabub Alam Tuhin

Date: 02/02/2026

FORM OF CONSENT

I Mahabub Alam Tuhin here by consent that my Project, submitted in candidature for the MSc In Cybersecurity & Digital Forensics if successful, may be made available for inter-library loan or photocopying (subject to the law of copyright), and that the title and abstract may be made available to outside organizations.

Signed Mahabub Alam Tuhin

Date 02/02/2026

Abstract

Healthcare organizations increasingly rely on interconnected digital systems to support clinical operations, patient data management, and medical decision-making. While this digital transformation improves efficiency and quality care, it also exposes healthcare systems to escalating cybersecurity threats. Cyberattacks such as ransomware, network intrusions, and exploitation of Internet of Medical Things (IoMT) devices pose serious risks to patient safety, service continuity, and data confidentiality.

This dissertation investigates the application of artificial intelligence–driven intrusion detection systems to enhance cybersecurity within healthcare environments. A supervised machine learning–based framework is designed, implemented, and experimentally evaluated using the CICIDS2017 intrusion detection dataset. Two widely adopted algorithms - Random Forest and Support Vector Machine - are implemented and compared using healthcare-relevant performance metrics, with particular emphasis on recall and false negative rates.

Experimental results demonstrate that Random Forest outperforms Support Vector Machine in terms of recall, computational efficiency, and interpretability, making it more suitable for healthcare cybersecurity contexts. The findings highlight the importance of prioritizing patient safety, transparency, and feasibility when deploying AI-driven security solutions in regulated environments. This research contributes a practical and ethically compliant framework for healthcare intrusion detection and provides recommendations for future research and real-world deployment.

Contents

CHAPTER 1: INTRODUCTION	10
1.1 Background and Context.....	10
1.2 Problem Statement.....	11
1.3 Aim of the Research	12
1.4 Research Objectives	12
1.5 Scope of Research.....	13
1.6 Significance of Research.....	14
1.7 Dissertation Structure	14
CHAPTER 2: LITERATURE REVIEW.....	15
2.1 Digital Transformation of Healthcare Systems	16
2.2 Healthcare Cyber Threat Landscape	17
2.3 Impact of Cyberattacks on Patient Safety and Clinical Outcomes.....	18
2.4 Legacy Systems and IoMT Security Challenges.....	18
2.5 Limitations of Traditional Cybersecurity Approaches in Healthcare	19
2.6 Artificial Intelligence in Cybersecurity	20
2.7 Machine Learning Paradigms for Intrusion Detection	21
2.7.1 Supervised Learning Approaches.....	21
2.7.2 Unsupervised and Anomaly Detection Approaches	22
2.7.3 Hybrid and Semi-Supervised Approaches	22
2.8 Evaluation Metrics for Healthcare Intrusion Detection Systems	22
2.9 Intrusion Detection Datasets for Healthcare Research	23
2.10 Ethical, Legal, and Regulatory Considerations	23
2.11 Research Gaps and Chapter Summary	24
2.12 Explainability, Trust, and AI Adoption in Healthcare Cybersecurity.....	24
2.13 Operational Constraints and Deployment Challenges in Healthcare.....	25
2.14 Comparative Analysis of Existing Healthcare IDS Research.....	26

2.15 Summary of Research Gaps	26
2.16 Comparative Review of AI-Based Intrusion Detection Studies in Healthcare	27
2.17 Critical Justification for Model Selection: Random Forest and Support Vector Machine.....	28
2.18 Synthesis of Literature and Linkage to Research Methodology	29
2.19 Extended Chapter Conclusion	30
CHAPTER 3: RESEARCH METHODOLOGY.....	31
3.1 Research Paradigm and Philosophical Positioning.....	31
3.2 Research Approach.....	31
3.3 Research Design.....	32
3.4 Dataset Selection and Justification.....	33
3.4.1 CICIDS2017 Dataset.....	33
3.4.2 Dataset Subset Selection	33
3.5 Experimental Environmental and Tools.....	33
3.6 Data Preprocessing.....	34
3.6.1 Data Cleaning	34
3.6.2 Label Transformation.....	35
3.6.3 Feature Scaling	35
3.6.4 Handling Class Imbalance.....	35
3.7 Machine Learning Model Selection.....	36
3.7.1 Random Forest Classifier	36
3.7.2 Support Vector Machine (SVM)	36
3.8 Model Training and Validation.....	36
3.9 Evaluation Metrics	36
3.10 Ethical Considerations.....	37
3.11 Reliability and Validity	37
CHAPTER 4: DESIGN PROCESS AND EXPERIMENTAL FRAMEWORK	38
4.1 Design Objectives	38

4.2 Overall System Architecture	39
4.3 Data Acquisition Layer.....	39
4.4 Data Preprocessing and Feature Engineering Layer	40
4.4.1 Data Cleaning and Integration	40
4.4.2 Feature Selection and Scaling.....	41
4.4.3 Label Encoding and Class Balancing	42
4.5 Machine Learning Detection Layer	42
4.5.1 Random Forest Detection Module	42
4.5.2 Support Vector Machine Detection Module.....	43
4.6 Training and Testing Workflow.....	44
4.7 Evaluation Framework Design.....	45
4.8 Experimental Environment Architecture	46
4.9 Security, Ethical, and Compliance Considerations in Design	46
4.10 Design Limitations	46
CHAPTER 5: IMPLEMENTATION (EXPERIMENTAL EVALUATION).....	47
5.1 Implementation Environment.....	47
5.2 Dataset Preparation and Loading	48
5.2.1 Dataset Files Used	48
5.2.2 Data Loading Process	48
5.3 Data Preprocessing Implementation.....	49
5.3.1 Cleaning and Feature Filtering	49
5.3.2 Label Encoding	49
5.3.3 Feature Scaling	50
5.3.4 Handling Class Imbalance.....	50
5.4 Dataset Splitting Strategy	50
5.5 Random Forest Model Implementation.....	50
5.5.1 Model Configuration	50

5.5.2 Training Process	51
5.5.3 Prediction and Output Generation	51
5.6 Support Vector Machine Implementation	51
5.6.1 Model Configuration	51
5.6.2 Training Process	51
5.6.3 Prediction and Output Generation	52
5.7 Performance Measurement Implementation	52
5.8 Experimental Reproducibility	53
5.9 Implementation Challenges	53
5.10 Ethical and Security Considerations in Implementation	53
CHAPTER 6: EVALUATION OF RESULTS	54
6.1 Evaluation Strategy	54
6.2 Evaluation Metrics	54
6.3 Random Forest Model Performance	55
6.3.1 Overall Performance	55
6.3.2 Confusion Matrix Analysis	55
6.3.3 Discussion of Results	57
6.4 Support Vector Machine Model Performance	57
6.4.1 Overall Performance	57
6.4.2 Confusion Matrix Analysis	57
6.4.3 Discussion of Results	58
6.5 Comparative Analysis of Models	59
6.5.1 Accuracy and F1-Score Comparison	59
6.5.2 Recall and False Negative Analysis	60
6.5.3 Computational Efficiency	60
6.6 Alignment with Research Objectives	61
6.7 Comparison with Existing Literature	62

6.8 Limitations of Evaluation	62
CHAPTER 7: CONCLUSIONS AND RECOMMENDATIONS	63
7.1 Summary of Research.....	63
7.2 Achievement of Research Objectives	64
7.3 Key Findings and Contributions.....	64
7.4 Practical Implications.....	65
7.5 Limitations of the Research	66
7.6 Recommendations for Future Research	66
7.7 Final Conclusion.....	67
CHAPTER 8: REFLECTION	68
8.1 Reflection on Research Planning and Methodology	68
8.2 Technical Skills Development	69
8.3 Challenges Encountered and Problem-Solving	69
8.4 Ethical Awareness and Professional Responsibility	70
8.5 Reflection on Learning Outcomes and Professional Growth.....	70
8.6 Reflection on Limitations and Areas for Improvement.....	71
8.7 Final Reflective Summary.....	71
REFERENCES	72
Appendix	74

List of figures and table

Figure 4.1: Mounting Google Drive and Defining Dataset Paths

Figure 4.2: Loading and Combining CICIDS2017 CSV Files

Figure 4.3: Handling Missing and Infinite Values

Figure 4.4: Selecting Numeric Features and Applying Standard Scaling

Figure 4.5: Random Forest Training and Prediction

Figure 4.6: SVM Training and Prediction

Figure 4.7: Train/Test Split with Stratification

Figure 4.8: Computing Evaluation Metrics

Figure 6.1: Confusion Matrix for the Random Forest intrusion detection model

Figure 6.2: Confusion Matrix for the Support Vector Machine intrusion detection model

Figure 6.3: Performance comparison of Random Forest and Support Vector Machine Models Using the CICIDS2017 Dataset

CHAPTER 1: INTRODUCTION

1.1 Background and Context

The digital transformation of healthcare systems has significantly reshaped the delivery, management, and monitoring of clinical services. Healthcare organizations increasingly rely on interconnected digital infrastructures, including Electronic Health Records (EHRs), Hospital Information Systems (HIS) cloud-based healthcare platforms, telemedicine technologies, and the Internet of Medical Things (IoMT). These systems enable real-time access to patient data, improved coordination among clinical teams and data-driven decision-making that enhances patient outcomes and operational efficiency [1],[2].

However, the rapid integration of digital technologies has also expanded the cyber-attack surface of healthcare environments. Healthcare systems process and store large volumes of highly sensitive data, including personally identifiable information (PII), protected health information (PHI), medical imaging data, and genomic records. This data is highly valuable to cybercriminals, making healthcare organizations frequent targets of cyberattacks [3]. Recent studies show that the healthcare sector consistently experiences higher rates of cyber incidents than many other industries, including ransomware attacks, data breaches, and network intrusions [4], [5].

Unlike other sectors, cybersecurity incidents in healthcare can have direct consequences on patient safety. Cyberattacks may disrupt access to EHR systems, delay diagnostic procedures, disable medical devices, and interrupt clinical workflows. Empirical studies have demonstrated that ransomware attacks on hospitals are associated with increased emergency department delays, reduced service availability, and statistically significant increases in patient mortality rates [6], [7]. These findings highlight that cybersecurity in healthcare is not merely an information technology issue, but a critical patient safety concern.

Traditional cybersecurity mechanisms, such as firewalls, antivirus software, and signature-based intrusion detection systems (IDS), are increasingly inadequate in protecting modern healthcare infrastructures. Healthcare networks are highly heterogeneous, often combining legacy systems,

modern IT infrastructure, cloud services, and resource-constrained IoMT devices. Many medical devices operate on outdated operating systems and cannot be easily patched due to regulatory restrictions and the need for uninterrupted clinical operation [8], [9]. As a result, static and rule based security approaches struggle to detect sophisticated and evolving cyber threats.

Artificial Intelligence (AI) and Machine Learning (ML) techniques have emerged as promising solutions to address these limitations. AI-driven cybersecurity systems can analyze large-scale network traffic data, learn complex behavioral patterns, and adapt to evolving threat landscapes. ML-based intrusion detection systems have demonstrated strong potential in detecting both known and unknown cyber threats by learning from historical data rather than relying solely on predefined signatures [10], [11]. These capabilities are especially relevant for healthcare environments, where early detection and minimization of false negatives are essential.

Despite increasing academic interest, many AI-driven cybersecurity solutions lack experimental validation in healthcare-relevant contexts or fail to consider ethical, regulatory, and operational constraints. There is therefore a clear need for structured research that evaluates the feasibility and effectiveness of AI-driven intrusion detection systems within healthcare-like environments using realistic datasets and robust experimental methodologies.

1.2 Problem Statement

Healthcare organizations face a complex cybersecurity challenge: they must protect sensitive patient data and mission-critical systems while ensuring continuous availability and compliance with strict regulatory frameworks such as GDPR and HIPA [12], [13]. Cyberattacks such as ransomware, Distributed Denial of Service (DDoS) attacks, and unauthorized network intrusions can disrupt clinical services, compromise patient confidentiality, and undermine trust in healthcare institutions [14], [5].

Traditional intrusion detection systems rely heavily on static rules and predefined attack signatures, which limit their ability to detect novel or evolving threats. Furthermore, such systems often generate high false-positive rates, contributing to alert fatigue among security teams and reducing

operational effectiveness [15]. In healthcare settings, excessive false alarms may interfere with clinical workflows, while false negatives may allow malicious activity to persist undetected.

AI-driven intrusion detection systems offer a potential solution by leveraging machine learning algorithms to model normal network behavior and identify deviations indicative of malicious activity. However, the adoption of AI-based security solutions in healthcare is hindered by concerns related to reliability, explainability, ethical compliance, and computational feasibility. Many existing studies focus on generic enterprise networks and do not sufficiently address healthcare specific risks or evaluation criteria [4], [16].

The core problems addressed in this research are the lack of experimentally validated healthcare-oriented AI-driven detection frameworks that balance detection effectiveness with ethical and practical constraints. This research seeks to investigate whether supervised machine learning techniques can reliably detect cyber threats in healthcare-like network environments while minimizing false negatives and maintaining operational suitability.

1.3 Aim of the Research

The primary aim of this research is to design, implement, experimentally evaluate an AI-driven intrusion detection framework for cybersecurity in healthcare systems. The study aims to assess the feasibility and effectiveness of supervised machine learning techniques in detecting malicious network traffic within a simulated healthcare environment using realistic cybersecurity datasets.

1.4 Research Objectives

This Research is guided by the following objectives

- To analyze cybersecurity challenges and threat characteristics specific to healthcare systems.

- To critically review existing AI and machine learning approaches for intrusion detection in healthcare environments.
- To design a structured AI-driven intrusion detection framework suitable for healthcare systems while considering ethical and regulatory requirements.
- To implement supervised machine learning models for intrusion detection using a realistic cybersecurity dataset.
- To evaluate and compare the performance of selected machine learning models using standard cybersecurity metrics.
- To assess the suitability of AI-driven intrusion detection techniques for healthcare environments based on recall, accuracy, and false negative rates.

1.5 Scope of Research

This research focuses on experimental evaluation rather than real-time deployment. The study is limited to supervised machine learning models, specifically Random Forest and Support Vector Machine (SVM), due to their interpretability, robustness, and widespread use in intrusion detection research [10], [17], [18].

The experimental evaluation uses a subset of the CICIDS2017 dataset, a publicly available dataset developed to simulate realistic network traffic and cyberattack scenarios [19]. All experiments are conducted in an offline environment using de-identified data to ensure ethical compliance. Real patient data, live healthcare systems, and real-time detection architectures are beyond the scope of this MSc-level research.

1.6 Significance of Research

This research contributes to healthcare cybersecurity by providing empirical evidence on the effectiveness of AI-driven intrusion detection techniques in healthcare-like environments. By prioritizing recall and false negative analysis, the study addresses practical security concerns that are particularly critical in healthcare settings [7], [21].

The findings demonstrate how supervised machine learning techniques can enhance intrusion detection capabilities while maintaining feasibility and ethical compliance. The proposed framework provides a foundation for future research into real-time detection, deep learning models, federated learning, and explainable AI in healthcare cybersecurity.

1.7 Dissertation Structure

The remainder of this dissertation is organized as follows:

- **Chapter 2** presents a critical review of literature related to healthcare cybersecurity and AI-driven intrusion detection.
- **Chapter 3** describes the research Methodology, including data collection, Processing, and evaluation strategies.
- **Chapter 4** details the design process and experimental framework.
- **Chapter 5** discusses system implementation and experimental setup.
- **Chapter 6** evaluates and analyses the experimental results.
- **Chapter 7** concludes the research and presents recommendations for the future work.
- **Chapter 8** provides a reflective account for research processes and the learning outcomes.

CHAPTER 2: LITERATURE REVIEW

The increasing digitization of healthcare systems has fundamentally transformed clinical service delivery, organizational workflows, and patient engagement. Healthcare organizations now depend heavily on interconnected information systems, including Electronic Health Records (EHRs), Hospital Information Systems (HIS), cloud-based healthcare platforms, telemedicine infrastructures, and the Internet of Medical Things (IoMT). These technologies enable real-time access to patient data, improved coordination between clinical teams, and data-driven medical decision-making, which collectively contribute to improved healthcare efficiency and patient outcomes [1].

However, the same digital transformation that has modernized healthcare has also introduced complex and evolving cybersecurity risks. Healthcare systems have become attractive targets for cybercriminals due to the high value of medical data and the critical nature of healthcare services. Unlike other sectors, cybersecurity failures in healthcare can have immediate and severe consequences, including disruption of clinical services, compromise of patient safety, and loss of life [20]. As a result, cybersecurity has emerged as a fundamental concern for healthcare organizations, regulators, and policymakers.

This chapter gives a expansive and critical review of existing literature related to cybersecurity in healthcare systems, with a particular focus on AI-driven intrusion detection approaches. The chapter begins by examining the digital transformation of healthcare and the resulting cybersecurity challenges. It then analyses the evolving cyber threat landscape, the documented impact of cyberattacks on patient safety, and the vulnerabilities introduced by legacy systems and IoMT devices. Traditional cybersecurity approaches are critically evaluated, followed by an in-depth review of artificial intelligence and machine learning techniques applied to intrusion detection. Finally, the chapter identifies key research gaps that encourage the development of the AI-driven intrusion detection framework proposed in this research.

2.1 Digital Transformation of Healthcare Systems

Healthcare systems worldwide have undergone rapid and sustained digital transformation over the past two decades. The adoption of Electronic Health Records has replaced paper-based documentation, enabling structured data storage, improved accessibility, and enhanced interoperability across clinical departments. Studies have shown that EHR adoption improves care coordination, reduces medical errors, and supports evidence-based clinical decision-making [1].

Hospital Information Systems integrate administrative, financial, diagnostic, and clinical processes, enabling more efficient resource management and improved service delivery. In parallel, cloud computing has been increasingly adopted to address scalability and data storage challenges associated with large volumes of healthcare data. Cloud-based healthcare platforms enable flexible data access, remote collaboration, and advanced analytics while reducing on premise infrastructure costs [21].

Telemedicine has further accelerated healthcare digitization by enabling remote consultations, monitoring, and diagnosis. The COVID-19 pandemic significantly increased reliance on telehealth services, highlighting both the benefits and security risks associated with remote healthcare delivery [22]. While telemedicine improves accessibility and continuity of care, it also expands the attack surface by introducing new endpoints, communication channels, and third-party service dependencies.

The Internet of Medical Things represents one of the most transformative developments in healthcare digitization. IoMT devices include wearable sensors, infusion pumps, patient monitors, implantable devices, imaging systems, and remote diagnostic tools. These devices continuously generate and transmit sensitive health data, supporting real-time monitoring and personalized healthcare delivery [8]. The number of connected medical devices has increased exponentially, with estimates suggesting tens of billions of IoMT devices deployed globally [23].

Despite these advancements, healthcare digital transformation has significantly increased system complexity. Healthcare IT environments are highly heterogeneous, often integrating legacy clinical systems with modern IT infrastructure, cloud services, and third-party platforms. Many healthcare organizations continue to rely on legacy systems that were not designed with modern cybersecurity

principles in mind [9]. This heterogeneity complicates security management and increases the likelihood of vulnerabilities arising from misconfigurations, outdated software, and inconsistent security policies.

2.2 Healthcare Cyber Threat Landscape

Healthcare organizations have become one of the most targeted sectors for cyberattacks. Medical data is particularly valuable to cybercriminals due to its longevity and richness, containing personal identifiers, medical histories, insurance details, and financial information. Research indicates that medical records can be worth significantly more than financial data on underground markets, making healthcare institutions lucrative targets [3].

Ransomware attacks represent one of the most severe and prevalent threats facing healthcare systems. In a ransomware attack, attackers encrypt critical systems and demand payment for decryption keys. Healthcare organizations are especially vulnerable because system downtime can disrupt essential clinical services and endanger patient safety. Studies have documented a dramatic increase in ransomware attacks against hospitals, with attackers deliberately targeting healthcare due to the high likelihood of ransom payment [16].

Empirical evidence suggests that ransomware attacks often involve multi-stage attack chains, including phishing emails, credential theft, lateral movement within networks, and privilege escalation. These attacks exploit both technical vulnerabilities and human factors, such as lack of cybersecurity awareness among healthcare staff [5]. Once attackers gain access, they can disable EHR systems, laboratory information systems, and imaging platforms, severely impacting clinical operations.

Beyond ransomware, healthcare organizations face threats from Distributed Denial-of-Service (DDoS) attacks, unauthorized network intrusions, insider threats, and data breaches. Insider threats are particularly concerning due to the large number of employees requiring system access and the limited cybersecurity training typically provided to clinical personnel [16]. Human error, including phishing and credential compromise, remains a leading cause of healthcare security incidents [5].

The complexity of healthcare networks, combined with regulatory constraints and limited cybersecurity budgets, makes comprehensive threat detection and mitigation challenging. These factors underscore the need for intelligent and adaptive security mechanisms capable of detecting both known and emerging cyber threats in real time.

2.3 Impact of Cyberattacks on Patient Safety and Clinical Outcomes

A defining feature of healthcare cybersecurity is the direct relationship between cyber incidents and patient safety. Unlike many other sectors, where cyberattacks primarily result in financial loss or reputational damage, cybersecurity failures in healthcare can disrupt clinical services and lead to adverse patient outcomes.

Multiple empirical studies have demonstrated a link between cyberattacks and patient harm. Martin et al. reported that ransomware attacks on hospitals were associated with increased emergency department delays, cancellation of elective procedures, and diversion of patients to other facilities [6]. Similarly, research published in *JAMA Network Open* found statistically significant increases in mortality rates following ransomware incidents affecting hospital systems [7].

Cyber incidents can prevent clinicians from accessing EHRs, laboratory results, and diagnostic imaging, forcing reliance on incomplete information or manual processes. Network-connected medical devices may become unavailable or unreliable, increasing the risk of clinical error. In severe cases, hospitals may be forced to revert to paper-based workflows, which are slower, less accurate, and more prone to human error [16].

These findings have led to a growing recognition that cybersecurity is a patient safety issue rather than solely an IT concern. Regulatory bodies and healthcare organizations increasingly emphasize the integration of cybersecurity into clinical risk management and patient safety frameworks [12]. This shift in perspective has important implications for the design and evaluation of healthcare cybersecurity solutions, particularly intrusion detection systems.

2.4 Legacy Systems and IoMT Security Challenges

Legacy systems remain a significant source of vulnerability in healthcare environments. Many clinical systems operate on outdated operating systems and lack support for modern security

mechanisms such as encryption, strong authentication, and secure update processes [9]. Regulatory requirements and the need to maintain continuous clinical availability often delay system upgrades and patching, resulting in prolonged exposure to known vulnerabilities.

IoMT devices introduce additional security challenges due to hardware constraints, proprietary software, and long device lifecycles. He et al. highlights that many medical devices cannot be easily patched without regulatory approval, leaving them vulnerable for extended periods [8]. Furthermore, IoMT devices often communicate using unencrypted or weakly protected protocols, increasing the risk of interception and manipulation.

Compromised medical devices can serve as entry points for attackers to gain access to broader healthcare networks. Studies have demonstrated that attackers can exploit vulnerable IoMT devices to establish persistence, move laterally within networks, and exfiltrate sensitive patient data [24]. The scale and heterogeneity of IoMT deployments make comprehensive device-level security difficult, reinforcing the importance of network-based intrusion detection approaches.

2.5 Limitations of Traditional Cybersecurity Approaches in Healthcare

Traditional cybersecurity approaches remain widely deployed in healthcare environments, including firewalls, antivirus software, access control mechanisms, and signature-based intrusion detection systems (IDS). While these controls provide a baseline level of protection, extensive research has shown that they are increasingly inadequate for defending modern healthcare infrastructures against sophisticated cyber threats.

Signature-based IDS rely on predefined patterns or signatures derived from known attacks. These systems are effective at detecting previously identified threats but are inherently incapable of identifying zero-day attacks or novel variants that do not match existing signatures [10]. Given the rapid evolution of cyber threats and the frequent emergence of new attack techniques, reliance on static signatures leaves healthcare systems vulnerable to advanced and previously unseen attacks.

Anomaly-based traditional IDS attempt to identify malicious activity by detecting deviations from predefined baselines of normal behavior. However, these systems often rely on manually defined

thresholds and statistical assumptions that are poorly suited to the dynamic and heterogeneous nature of healthcare environments [11]. Clinical workflows, device usage patterns, and network traffic characteristics can vary significantly across departments and over time, making it difficult to define stable baselines. As a result, anomaly-based systems frequently generate high false positive rates.

High false-positive rates contribute to alert fatigue among security teams, reducing the effectiveness of intrusion detection systems and increasing the likelihood that genuine threats will be overlooked. In healthcare settings, excessive security alerts may also disrupt clinical workflows and divert attention away from patient care [15]. These limitations have prompted increasing interest in more intelligent and adaptive cybersecurity solutions.

In addition to detection limitations, traditional security systems often struggle to scale with the volume, velocity, and variety of data generated by modern healthcare infrastructures. The proliferation of IoMT devices and continuous data streams overwhelms rule-based systems, which lack the capacity to analyze complex patterns across large datasets [19]. Consequently, traditional cybersecurity approaches are increasingly viewed as insufficient for protecting contemporary healthcare systems.

2.6 Artificial Intelligence in Cybersecurity

Artificial Intelligence has been recognized as an important paradigm for enhancing cybersecurity capabilities across a wide range of domains. AI-driven security systems leverage machine learning algorithms to analyze large-scale datasets, identify complex patterns, and adapt to evolving threat landscapes. Unlike static rule-based systems, AI-based approaches can learn from historical data and improve detection performance over time [10].

In cybersecurity, AI techniques have been applied to intrusion detection, malware classification, anomaly detection, threat intelligence, and automated incident response. Buczak and Guven provide a comprehensive survey of machine learning methods used in cybersecurity, demonstrating their superiority over traditional approaches in detecting sophisticated and multistage attacks [10].

AI-based systems have shown promise for deployment in healthcare settings because of their ability to handle diverse types of data from different sources, including network traffic, system

logs, and device communications. This enables the systems to learn normal behavior patterns and detect abnormalities that could signal malicious activity, even when the traffic is encrypted [10].

Despite their potential, AI-based cybersecurity solutions face several challenges. These include data quality issues, class imbalance, computational requirements, and lack of explainability. In healthcare contexts, transparency and interpretability are especially important due to regulatory requirements and the need for clinician trust [25]. These challenges must be carefully addressed when designing AI-driven intrusion detection systems for healthcare environments.

2.7 Machine Learning Paradigms for Intrusion Detection

Machine learning-based intrusion detection systems (ML-IDS) can be classified into three categories: supervised learning, unsupervised learning, and semi-supervised learning. Each type has its own advantages and disadvantages when used in the healthcare cybersecurity field.

2.7.1 Supervised Learning Approaches

In supervised learning, algorithms are used with labelled data sets, which then help train the model. The model will then learn how to classify network traffic as either malicious or harmless. Some supervised learning algorithms that can be used for intrusion detection include Random Forest, Support Vector Machines, Decision Trees, k-Nearest Neighbours, Gradient Boosting, etc. [10].

Random Forest is an ensemble learning method that consists of a combination of decision trees and is commonly used for intrusion detection due to its strong performance and its ability to avoid overfitting during training [17]. Studies have demonstrated that Random Forest performs consistently well across a range of intrusion detection datasets, including those simulating healthcare environments.

Support Vector Machine is another widely used supervised learning algorithm for intrusion detection. SVM constructs optimal decision boundaries in high-dimensional feature spaces and has demonstrated strong classification performance in various cybersecurity applications [18].

However, SVM performance can be sensitive to parameter selection and feature scaling, and it may struggle with highly imbalanced datasets commonly encountered in intrusion detection scenarios.

Supervised learning approaches generally achieve high detection accuracy when sufficient labelled data is available. However, their reliance on labelled datasets limits their ability to detect previously unseen attack patterns, highlighting the need for complementary detection strategies.

2.7.2 Unsupervised and Anomaly Detection Approaches

Unsupervised learning approaches aim to detect anomalies without relying on labelled attack data. These methods model normal system behavior and flag deviations as potential intrusions. Common unsupervised techniques include clustering algorithms, Isolation Forest, and autoencoders [26].

Unsupervised approaches are particularly valuable for detecting zero-day attacks and insider threats, which may not be represented in labelled datasets. However, defining “normal” behavior in dynamic healthcare environments is challenging due to frequent changes in network configurations, device usage, and clinical workflows [15]. As a result, unsupervised methods may generate high false-positive rates if models are not continuously updated.

2.7.3 Hybrid and Semi-Supervised Approaches

Hybrid intrusion detection systems integrate supervised and unsupervised techniques to leverage the strengths of both approaches. Semi-supervised learning methods use limited labelled data alongside large volumes of unlabeled data, improving detection of novel attacks while maintaining classification accuracy [10].

Hybrid approaches have shown promise in healthcare cybersecurity research, as they can adapt to evolving threat landscapes while minimizing false positives. However, these systems are often more complex to implement and require careful tuning to ensure stability and reliability.

2.8 Evaluation Metrics for Healthcare Intrusion Detection Systems

The evaluation of intrusion detection systems requires careful selection of performance metrics, particularly in healthcare contexts. Accuracy alone is insufficient because of highly imbalanced nature of cybersecurity datasets, where benign traffic significantly outnumbers attack traffic [15].

Recall and false negative rates are particularly critical metrics in healthcare cybersecurity. A false negative represents an undetected attack, which may lead to system compromise, service disruption, or patient harm. Research consistently emphasizes that high recall is essential for healthcare intrusion detection systems [7], [4].

Precision is also important, as excessive false positives may overwhelm security teams and disrupt healthcare operations. Balanced evaluation based on accuracy, precision, recall, F1-score, and confusion matrices is therefore widely recommended in healthcare cybersecurity research [10].

2.9 Intrusion Detection Datasets for Healthcare Research

In the field of intrusion detection systems, publicly available data plays a vital role in the experimentation process. One of the most popular publicly available data collections for testing intrusion detection systems is the CICIDS2017 data set, which was developed by the Canadian Institute for Cybersecurity. [5].

CICIDS2017 contains realistic network traffic captured under control conditions, incorporating both benign activity and a wide range of attack scenarios. Several studies have validated its suitability for intrusion detection research, including healthcare-oriented simulations [7]. Due to its large size, many researchers use representative subsets of the dataset to ensure computational feasibility, an approach widely accepted in the literature.

2.10 Ethical, Legal, and Regulatory Considerations

The research in cybersecurity in the field of healthcare must adhere to several ethical and legislative guidelines, such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA). These regulations require that the confidentiality, integrity, and availability of patient information must be maintained [12].

Research involving real patient data requires ethical approval and rigorous data governance controls. As a result, simulation-based experimentation using publicly available datasets is widely

adopted in healthcare cybersecurity research. Such approaches enable realistic evaluation while ensuring ethical compliance and reproducibility [28].

2.11 Research Gaps and Chapter Summary

Although existing literature demonstrates the potential of AI-driven intrusion detection systems, several gaps remain. Many studies focus on generic enterprise networks rather than healthcare specific environments. Others prioritize detection accuracy without sufficient emphasis on false negatives, explainability, and patient safety implications.

Furthermore, there is a lack of experimentally validated frameworks that balance detection performance, ethical compliance, and computational feasibility in healthcare contexts. These gaps motivate the current research, which aims to design and evaluate an AI-driven intrusion detection framework tailored specifically to healthcare environments using realistic datasets and healthcare relevant evaluation metrics.

2.12 Explainability, Trust, and AI Adoption in Healthcare Cybersecurity

Even though artificial intelligence and machine learning approaches have shown promising performance for intrusion detection, the adoption of such approaches in healthcare cybersecurity not only depends on the performance of the approaches, but also other factors such as trust, transparency, and explainability. Healthcare environments are safety-critical, heavily regulated, and require interpretable security decisions.

One of the major drawbacks of most AI-powered security tools is that they are “black-box” systems. This is because complex models, especially deep learning models, are often able to make predictions with a high level of accuracy without being able to explain why they made those predictions. According to Doshi-Velez and Kim, this lack of interpretability is one of the reasons why AI systems have not been widely used in high-stakes applications like healthcare [25].

Explainable Artificial Intelligence (XAI) methods have been proposed to address this challenge. The proposed methods include feature importance ranking, SHAP (Shapley Additive Explanations), and LIME (Local Interpretable Model Agnostic Explanations) to improve the level of transparency in ML-based IDS [29]. In the context of healthcare cybersecurity, the importance of explainability is related to the ability to comply with regulations, perform forensic analysis, and increase the level of trust among medical professionals.

Studies suggest that interpretable models, such as Random Forest and Decision Trees, are often preferred in regulated environments due to their relative transparency compared to deep neural networks [17]. This consideration has influenced model selection in many healthcare cybersecurity studies and is a key factor motivating the use of Random Forest in the present research.

2.13 Operational Constraints and Deployment Challenges in Healthcare

Beyond algorithmic performance, practical deployment considerations significantly influence the feasibility of AI-driven intrusion detection systems in healthcare environments. Healthcare organizations operate under strict availability requirements, where system downtime can disrupt patient care. Security solutions must therefore operate passively and minimize performance overhead [16].

Computational constraints represent another challenge. Many healthcare organizations lack the infrastructure required to support resource-intensive AI models, particularly in smaller hospitals and clinics. Cloud-based deployment offers scalability but introduces additional concerns related to data sovereignty, latency, and regulatory compliance [21].

Integration with existing healthcare IT systems is also complex. Security solutions must coexist with legacy systems, vendor-specific medical devices, and heterogeneous network architectures. Studies indicate that intrusion detection systems designed for generic enterprise networks often fail to account for healthcare-specific workflows and constraints [16].

These operational challenges reinforce the importance of designing AI-driven intrusion detection frameworks that prioritize feasibility, modularity, and reproducibility rather than aggressive

automation. Experimental, simulation-based validation is therefore widely adopted as an appropriate research approach at MSc level.

2.14 Comparative Analysis of Existing Healthcare IDS Research

A review of existing healthcare intrusion detection research reveals several common trends and limitations. Many studies demonstrate high detection accuracy using benchmark datasets but fail to analyze false negatives in sufficient depth. In healthcare contexts, even a small number of undetected attacks can have severe consequences, making recall and false negative rates critical evaluation metrics [7].

Some studies focus exclusively on deep learning approaches, achieving high accuracy at the cost of interpretability and computational complexity. Others prioritize lightweight models but lack comprehensive evaluation across diverse attack scenarios. Furthermore, many proposed systems remain conceptual or lack experimental validation using realistic datasets [11].

Comparative analysis suggests that ensemble-based supervised learning models, like Random Forest, offer a favorable balance between detection performance, interpretability, and feasibility. However, there remains a lack of structured frameworks that integrate these models into healthcare-oriented security architectures while explicitly addressing ethical and operational constraints.

2.15 Summary of Research Gaps

Based on the critical review presented in this chapter, several key research gaps can be identified:

Healthcare-Specific Validation Gap:

Many intrusion detection studies focus on generic enterprise networks rather than healthcare-like environments.

Patient Safety Consideration Gap:

Existing research often prioritizes accuracy without sufficient emphasis on false negatives and patient safety implications.

Explainability Gap:

Limited attention is given to transparency and interpretability, despite their importance in regulated healthcare settings.

Feasibility and Resource Constraint Gap:

Few studies address computational and operational constraints faced by healthcare organizations.

Experimental Framework Gap:

There is a lack of structured, experimentally validated AI-driven intrusion detection frameworks tailored specifically to healthcare systems.

These gaps directly motivate the research presented in this dissertation, which aims to design, implement, and evaluate an AI-driven intrusion detection framework that balances detection effectiveness, ethical compliance, and practical feasibility in healthcare environments.

2.16 Comparative Review of AI-Based Intrusion Detection Studies in Healthcare

Research has increasingly been conducted to assess the potential of artificial intelligence and machine learning methods in the domain of intrusion detection in healthcare and healthcare-like environments. Though the results are quite promising, there is considerable variability in the experimental methodologies, data sets, and results, as well as the level of applicability of the methods examined in the research.

Several studies have focused on supervised learning approaches using benchmark intrusion detection datasets. For example, Sharafaldin et al. evaluated multiple machine learning classifiers, including Random Forest and Support Vector Machine, on the CICIDS2017 dataset and reported high detection accuracy across multiple attack types [19]. However, their analysis primarily emphasized accuracy and did not sufficiently explore false negative rates, which are critical in healthcare contexts.

Other researchers have applied deep learning models, like Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, to intrusion detection tasks. These models often achieve high performance because of their ability to learn complex feature representations [11]. Nevertheless, such approaches typically require substantial computational resources and large volumes of training data, limiting their feasibility for deployment in real-world healthcare environments, particularly in smaller institutions.

Healthcare-specific studies often highlight the challenges posed by IoMT devices and heterogeneous network traffic. Ullah et al. demonstrated that IoMT-focused intrusion detection systems must handle constrained devices, encrypted traffic, and non-standard communication protocols [24]. While their work emphasized the need for intelligent detection mechanisms, it remained largely conceptual and lacked comprehensive experimental validation.

Across literature, a common limitation is the lack of alignment between detection performance and healthcare operational requirements. Many studies demonstrate high accuracy in controlled environments but fail to consider explainability, deployment feasibility, and integration with existing healthcare systems. This gap reinforces the need for experimentally validated healthcare-oriented intrusion detection frameworks that balance performance with practical constraints.

2.17 Critical Justification for Model Selection: Random Forest and Support Vector Machine

Model selection is a critical decision in the design of AI-driven intrusion detection systems, particularly in regulated and safety-critical domains such as healthcare. While deep learning models have gained popularity, their adoption in healthcare cybersecurity remains limited due to concerns related to explainability, computational overhead, and regulatory acceptance.

Random Forest has been widely adopted in intrusion detection research due to its ensemble-based structure, which enhances robustness and reduces overfitting. One of its key advantages is interpretability: feature importance measures allow security analysts to understand which attributes contribute most significantly to detection decisions [17]. In healthcare environments, this transparency supports auditability and forensic investigation, making Random Forest particularly suitable.

Support Vector Machine offers complementary strengths. SVM performs well in high-dimensional feature spaces and is effective at constructing optimal decision boundaries for classification tasks [18]. In intrusion detection, SVM has demonstrated strong performance in detecting complex attack patterns, especially when appropriate kernel functions and feature scaling techniques are applied. However, SVM may be sensitive to class imbalance, a common characteristic of intrusion detection datasets.

Compared to deep learning models, both Random Forest and SVM offer lower computational complexity and greater interpretability. These characteristics align well with healthcare operational constraints, where system resources may be limited and explainability is essential. As a result, the selection of Random Forest and SVM in this research reflects a deliberate trade-off between detection performance, feasibility, and regulatory suitability.

2.18 Synthesis of Literature and Linkage to Research Methodology

The literature reviewed in this chapter provides a strong theoretical basis for the research methodology used in this research. Existing literature points to the limitations of conventional cybersecurity approaches, the promise of AI-based intrusion detection systems, and healthcare-specific factors that must be considered.

The prevalence of supervised learning approaches in intrusion detection research, particularly Random Forest and SVM, supports the experimental focus of this study. The use of benchmark datasets such as CICIDS2017 is consistent with best practices identified in the literature, enabling reproducible experimentation and meaningful comparison of results [19], [27].

Furthermore, the emphasis on recall and false negative analysis in healthcare cybersecurity literature directly informs the evaluation strategy adopted in this research. Rather than relying solely on accuracy, the study prioritizes metrics that reflect the real-world consequences of missed detections in healthcare environments.

By synthesizing these insights, this literature review directly informs the research design, data preprocessing strategy, model selection, and evaluation framework described in Chapter 3. The

proposed AI-driven intrusion detection framework is therefore grounded in established research while addressing identified gaps related to healthcare specificity, feasibility, and ethical compliance.

2.19 Extended Chapter Conclusion

In summary, this literature review has critically examined existing research on healthcare cybersecurity and AI-driven intrusion detection systems. It has demonstrated that while significant progress has been made, current approaches often fail to adequately address healthcare-specific requirements such as patient safety, explainability, and operational constraints.

The additional critical analysis presented in this chapter positions the current research within the broader academic landscape and justifies the methodological choices made. By addressing identified gaps and aligning with healthcare operational realities, this study aims to contribute meaningful and practically relevant insights to the field of healthcare cybersecurity.

CHAPTER 3: RESEARCH METHODOLOGY

The research methodology employed to develop, implement, and evaluate an artificial intelligence based framework to ensure cybersecurity in healthcare systems is discussed in this chapter. The objective of this chapter is to justify the research methodology, explain the experimental design, discuss the data sources and preprocessing, and highlight the evaluation methodology employed to assess the performance of the system. The methodology employed is crucial to ensure that the results obtained are valid, reliable, and replicable.

The methodology presented in this chapter is directly influenced by the research goals presented in Chapter 1 and the research gaps presented in Chapter 2. Considering the technical and performance-oriented nature of the field of intrusion detection, this research will employ a quantitative experimental methodology, which is recognized as the most suitable methodology to assess the performance of machine learning-based cybersecurity systems [30], [31].

3.1 Research Paradigm and Philosophical Positioning

Research paradigms define the underlying assumptions about knowledge generation and validation. This research is grounded in the positive paradigm, which assumes that reality is objective and can be measured by empirical observation and experimentation. Positivism is commonly adopted in computer science and cybersecurity research, where system performance is evaluated using quantifiable metrics such as accuracy, recall, and false negative rates [30]. The positivist paradigm supports hypothesis testing and objective comparison of alternative technical solutions. In the context of intrusion detection systems, this paradigm enables the direct evaluation of different machine learning models based on measurable performance outcomes. As a result, it is well suited to research that focuses on algorithmic effectiveness rather than subjective human interpretation.

3.2 Research Approach

As a quantitative research study, this research follows the positivist paradigm. Quantitative methods enable objective measurement and statistical comparison of machine learning models,

making them particularly suitable for cybersecurity experimentation [31]. Alternative qualitative approaches, such as interviews or surveys, were considered but deemed inappropriate for this research. The primary goal of the study is not to explore user perceptions or organizational practices, but to evaluate the technical performance of intrusion detection models. Quantitative experimentation therefore provides the most reliable and valid way to address the research objectives.

3.3 Research Design

The research adopts an experimental research design, which is widely used in artificial intelligence and cybersecurity studies [10], [28]. Experimental research provides the means for manipulating independent variables (machine learning algorithms) and observing the results on the dependent variables (intrusion detection performance metrics).

The experimental design followed in this study consists of the following stages:

- Selection of an appropriate intrusion detection dataset
- Data cleaning, preprocessing, and feature preparation
- Implementation of supervised machine learning models
- Training and testing models using controlled conditions
- Evaluation and comparative analysis of model performance

This structured design ensures methodological consistency, repeatability, and transparency, which are essential for academic rigor and examiner evaluation.

3.4 Dataset Selection and Justification

3.4.1 CICIDS2017 Dataset

The CICIDS2017 dataset, developed by the Canadian Institute for Cybersecurity, was chosen as the main dataset for the study. The CICIDS2017 dataset is a popular choice in intrusion detection system research, intended to mimic realistic network traffic patterns, including benign traffic and a range of different cyber-attacks [19].

The dataset includes comprehensive flow-based network features extracted using CICFlowMeter and incorporates modern attack scenarios such as Distributed Denial-of-Service (DDoS) and web-based attacks. Its realism, diversity, and public availability make it suitable for healthcare-oriented cybersecurity research without the ethical risks associated with real patient data.

3.4.2 Dataset Subset Selection

Due to the large size of the full CICIDS2017 dataset and computational constraints, a representative subset was selected for experimentation. The following files were used:

- Monday Working Hours. Cap ISCX csv file (benign traffic)
- Friday Working Hours Afternoon DDos pcap ISCX csv file (DDoS attacks)
- Thursday Working Hours Morning Web Attacks pcap ISCX csv file (web attacks)

The use of dataset subsets is common practice in intrusion detection research and does not compromise experimental validity when traffic diversity is preserved [27], [19]. This approach enables efficient experimentation while maintaining realistic attack representation.

3.5 Experimental Environmental and Tools

The experimental implementation was conducted using the following tools and technologies:

- Programming language: Python
- Execution environment: Google Colab
- Data processing libraries: Pandas, NumPy
- Machine learning framework: Scikit-learn
- Visualization tools: Matplotlib

Google Colab was selected due to its cloud-based infrastructure, ease of reproducibility, and ability to process large datasets without requiring specialized local hardware. This choice aligns with the practical constraints faced by many healthcare organizations and supports reproducible research.

3.6 Data Preprocessing

Raw network traffic data sets are generally associated with some level of noise, missing values, as well as redundant features, which might influence the performance of machine learning algorithms [32]. As such, extensive preprocessing was performed before the training of the machine learning model.

3.6.1 Data Cleaning

All selected CSV files were merged into a single dataset. Non-numeric attributes, such as timestamps and flow identifiers, were removed to ensure compatibility with supervised learning algorithms. Missing, infinite, and undefined values were handled using appropriate replacement and filtering techniques.

3.6.2 Label Transformation

The original multi-class labels in CICIDS2017 were transformed into a binary classification format, where:

- 0 = Benign traffic
- 1 = Malicious traffic

Binary classification simplifies the intrusion detection task and aligns with the primary objective of detecting malicious activity rather than identifying specific attack types.

3.6.3 Feature Scaling

Feature scaling has been achieved through the application of standardization techniques. This process is particularly critical for distance-based methods, for example, Support Vector Machine, since the performance of the algorithm depends on the scale of the features.

3.6.4 Handling Class Imbalance

The datasets used for intrusion detection systems are imbalanced by nature, meaning that benign traffic far exceeds attack traffic. Class imbalance can cause a bias in a machine learning algorithm, leading to poor detection of intrusions [15].

To address this issue, Random Over-Sampling (ROS) was applied to balance class distributions. Oversampling techniques improve minority-class detection and have been shown to enhance recall in intrusion detection systems [33].

3.7 Machine Learning Model Selection

After analyzing the literature and the research goals, two machine learning algorithms were chosen for implementation

3.7.1 Random Forest Classifier

Random Forest is an ensemble learning algorithm that combines multiple decision trees to improve classification accuracy and robustness. It is resistant to overfitting, performs well on high dimensional data, and provides feature importance measures, which enhance interpretability [17]. These characteristics make Random Forest suitable for healthcare cybersecurity, where transparency and reliability are essential.

3.7.2 Support Vector Machine (SVM)

Support Vector Machine is a margin-based classifier that constructs optimal decision boundaries in high-dimensional feature spaces. SVM has demonstrated strong performance in intrusion detection research but may be sensitive to class imbalance and parameter selection [18], [34]. Its inclusion enables comparative evaluation against an ensemble-based approach.

3.8 Model Training and Validation

The data set was divided into the training data set and the testing data set using the 80:20 ratio. This helps in the evaluation of the model using unseen data, which aids in the generalization of the model. This also helps in the avoidance of overfitting.

Both models have been trained with the same settings to ensure a fair comparison. Default and commonly used hyperparameters have been used to ensure methodological transparency and reproducibility.

3.9 Evaluation Metrics

Model performance was evaluated using multiple metrics:

- Accuracy

- precision
- Recall
- F1-score
- Confusion matrix

In healthcare cybersecurity, recall is particularly critical, as false negatives represent undetected attacks that may disrupt clinical services or compromise patient safety [7], [12], [34].

Consequently, recall and false negative rates were prioritized during evaluation.

3.10 Ethical Considerations

This study does not require human subjects, actual patient information, or an operational healthcare environment. All experiments were performed on publicly available datasets that do not contain identifiable information. Therefore, ethical approval was unnecessary.

This study follows the ethical standards set by the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA) by not using personal identifiable information and following proper data handling procedures [12], [13].

3.11 Reliability and Validity

Reliability was ensured through consistent experimental procedures, use of established datasets, and reproducible implementation in a controlled environment. Validity was supported by alignment between research objectives, methodology, and evaluation metrics. The use of widely accepted algorithms and datasets enhances the credibility and generalizability of the findings [31], [28].

CHAPTER 4: DESIGN PROCESS AND EXPERIMENTAL FRAMEWORK

This chapter presents the design process and experimental framework developed to implement an AI-driven intrusion detection system for healthcare cybersecurity. While Chapter 3 focused on the research methodology and experimental approach, this chapter concentrates on the architectural design, system workflow, and experimental structure underpinning the proposed solution.

The purpose of this chapter is to translate the theoretical foundations identified in the literature review and the methodological decisions outlined in Chapter 3 into a coherent system design. The chapter describes the overall system architecture, data flow, preprocessing pipeline, machine learning model integration, and evaluation framework. This structured design ensures that the proposed system is feasible, reproducible, and aligned with healthcare cybersecurity requirements.

4.1 Design Objectives

The design of the proposed intrusion detection framework was guided by several key objectives derived from the research gaps identified in Chapter 2 and the methodological constraints discussed in Chapter 3.

The primary design objectives are as follows:

- Accurate detection of malicious network traffic using supervised machine learning techniques.
- Minimization of false negatives, given the patient safety implications of undetected cyberattacks in healthcare environments [7].
- Interpretability and transparency, supporting trust, auditability, and regulatory compliance [25].
- Scalability and feasibility, ensuring the system can operate within resource-constrained healthcare settings.

- Reproducibility, enabling experimental validation using publicly available datasets.

These objectives informed all architectural and design decisions described in this chapter.

4.2 Overall System Architecture

The proposed AI-driven intrusion detection system follows a layered, modular architecture, which is commonly recommended in healthcare cybersecurity research due to its flexibility and maintainability [7], [10].

The system architecture consists of the following core components:

- Data acquisition layer
- Data preprocessing and feature engineering layer
- Machine learning detection layer
- Evaluation and reporting layer

This layered design enables separation of concerns, allowing each component to be developed, tested, and modified independently without affecting the overall system integrity.

4.3 Data Acquisition Layer

The data acquisition layer is responsible for sourcing network traffic data used to train and evaluate the intrusion detection models. In this research, network flow data was obtained from the CICIDS2017 dataset, which simulates realistic enterprise network traffic and attack scenarios [19]. The dataset includes both benign and malicious traffic flows, extracted using CICFlowMeter. Each flow is represented by a set of statistical features describing packet-level behavior, such as flow duration, packet size, and inter-arrival time. This flow-based representation aligns with modern intrusion detection practices and supports scalable analysis of high-volume network traffic [10].

Using publicly available datasets at this stage ensures ethical compliance, reproducibility, and alignment with best practices in cybersecurity experimentation [27].

Figure 4.1: Mounting Google Drive and Defining Dataset Paths

```
from google.colab import drive
drive.mount('/content/drive')

import os
DATA_DIR = "/content/drive/MyDrive/Dataset"
OUT_DIR = "/content/drive/MyDrive/CICIDS2017_results"
os.makedirs(OUT_DIR, exist_ok=True)
```

This listing shows how Google Colab connects to Google Drive to access CICIDS2017 CSV files and create an output folder for saving metrics and figures, supporting reproducibility.

4.4 Data Preprocessing and Feature Engineering Layer

Raw network traffic data cannot be directly used for machine learning due to noise, inconsistencies, and incompatible feature formats. Therefore, a dedicated preprocessing and feature engineering layer was designed to prepare the data for model training and evaluation.

4.4.1 Data Cleaning and Integration

Multiple CSV files representing different traffic scenarios were combined into a unified dataset. Non-numeric attributes, such as timestamps and flow identifiers, were removed to ensure compatibility with supervised learning algorithms. Records containing missing, infinite, or undefined values were identified and handled to prevent model bias and instability [32]. This step ensures that the dataset used in experimentation accurately reflects network behavior while maintaining data integrity.

Figure 4.2: Loading and Combining CICIDS2017 CSV Files

```
import pandas as pd

files = [
    "Monday-WorkingHours.pcap_ISCX.csv",
    "Friday-WorkingHours-Afternoon-DDos.pcap_ISCX.csv",
    "Thursday-WorkingHours-Morning-WebAttacks.pcap_ISCX.csv"
]

dfs = [pd.read_csv(os.path.join(DATA_DIR, f), low_memory=False) for f in files]
df = pd.concat(dfs, ignore_index=True)
df.columns = df.columns.str.strip()
```

This listing demonstrates loading multiple traffic scenarios and concatenating them into one dataset to ensure unified preprocessing and consistent model input features.

Figure 4.3: Handling Missing and Infinite Values

```
import numpy as np

x = df.replace([np.inf, -np.inf], np.nan)
x = x.dropna()
```

This snippet removes invalid values that can cause instability during model training and evaluation.

4.4.2 Feature Selection and Scaling

Feature selection was implicitly guided by the CICIDS2017 feature set, which has been validated in prior research as effective for intrusion detection tasks [19]. Rather than aggressively reducing dimensionality, all relevant numerical features were retained to preserve information richness.

Feature scaling was applied using standardization techniques to normalize feature distributions. This step is particularly important for algorithms such as Support Vector Machine, which rely on distance-based calculations and are sensitive to feature magnitude [18].

Figure 4.4: Selecting Numeric Features and Applying Standard Scaling

```
from sklearn.preprocessing import StandardScaler

x = df.drop(columns=["Label"]).select_dtypes(include=[np.number])

scaler = StandardScaler()
x_scaled = scaler.fit_transform(x)
```

This listing shows how only numeric CICIDS2017 flow features were retained and standardized to support algorithms such as SVM, which are sensitive to feature magnitudes.

4.4.3 Label Encoding and Class Balancing

The original multi-class attack labels were converted into a binary classification scheme, distinguishing between benign and malicious traffic. This design decision simplifies detection objectives and reflects real-world intrusion detection requirements, where the primary concern is identifying the presence of malicious activity.

To address class imbalance, which is inherent in intrusion detection datasets, Random Oversampling was applied to balance benign and attack classes. Oversampling techniques have been shown to improve minority-class detection and recall performance in cybersecurity applications [33].

4.5 Machine Learning Detection Layer

The machine learning detection layer represents the core intelligence of the proposed system. This layer is responsible for analyzing preprocessed network traffic and classifying each flow as benign or malicious.

4.5.1 Random Forest Detection Module

Random Forest was selected as the primary detection model due to its robustness, resistance to overfitting, and interpretability. The ensemble-based structure of Random Forest enables it to

capture complex relationships between features while maintaining stable performance across diverse attack types [17].

Feature importance metrics generated by Random Forest support explainability, allowing security analysts to understand which network characteristics contribute most significantly to detection decisions. This capability is particularly valuable in healthcare environments, where transparency and auditability are critical [25].

Figure 4.5: Random Forest Training and Prediction

```
from sklearn.ensemble import RandomForestClassifier

rf = RandomForestClassifier(n_estimators=200, random_state=42, n_jobs=-1)
rf.fit(X_train_scaled, y_train)
rf_pred = rf.predict(X_test_scaled)
```

This snippet shows how the Random Forest model is trained and used to predict unseen network traffic in the test dataset.

4.5.2 Support Vector Machine Detection Module

Support Vector Machine was implemented as a comparative detection model. SVM constructs optimal decision boundaries in high-dimensional feature spaces and has demonstrated strong classification performance in intrusion detection research [18], [34].

Including SVM enables comparative evaluation between ensemble-based and margin-based learning approaches. This comparison provides deeper insight into trade-offs between detection accuracy, recall, and computational complexity.

Figure 4.6: SVM Training and Prediction

```
from sklearn.svm import SVC

svm = SVC(kernel="rbf", C=1.0, gamma="scale")
svm.fit(X_train_scaled, y_train)
svm_pred = svm.predict(X_test_scaled)
```

This listing demonstrates implementation of SVM as a comparative model to evaluate trade-offs against Random Forest.

4.6 Training and Testing Workflow

The experimental workflow was designed to ensure fair and consistent evaluation of both machine learning models. The dataset was split into training and testing sets using an 80:20 ratio, ensuring that model performance was evaluated on previously unseen data. Both models were trained using identical data partitions and preprocessing steps. This controlled workflow minimizes experimental bias and supports valid comparison of detection performance [10]. Hyperparameter selection followed commonly recommended practices in intrusion detection research to balance performance and reproducibility.

Figure 4.7: Train/Test Split with Stratification

```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42, stratify=y
)
```

This ensures both benign and attack classes are proportionally represented in the training and test sets, supporting fair evaluation.

4.7 Evaluation Framework Design

The evaluation framework was designed to assess intrusion detection performance using metrics aligned with healthcare cybersecurity priorities. The following metrics were selected:

- Accuracy
- Precision
- Recall
- F1-score
- Confusion matrix

While accuracy provides an overall performance measure, recall was prioritized due to the severe implications of false negatives in healthcare environments [7], [12]. The confusion matrix was used to analyze misclassification patterns and understand trade-offs between detection sensitivity and false alarm rates.

This evaluation design ensures that model assessment reflects real-world healthcare security requirements rather than purely statistical performance.

Figure 4.8: Computing Evaluation Metrics

```
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score

accuracy = accuracy_score(y_test, rf_pred)
precision = precision_score(y_test, rf_pred, zero_division=0)
recall = recall_score(y_test, rf_pred, zero_division=0)
f1 = f1_score(y_test, rf_pred, zero_division=0)
```

This snippet computes standard intrusion detection metrics, with recall prioritized because false negatives are high risk in healthcare environments.

4.8 Experimental Environment Architecture

The experimental framework was implemented using a cloud-based execution environment (Google Colab), enabling scalable processing and reproducibility. Python was used as the primary programming language, supported by Scikit-learn for machine learning implementation. This environment supports modular experimentation, enabling rapid testing, parameter adjustment, and result visualization without dependency on local hardware constraints. Such flexibility aligns with resource limitations commonly faced by healthcare organizations [21].

4.9 Security, Ethical, and Compliance Considerations in Design

Security and ethical considerations were integrated into the design process from the outset. The system operates exclusively on offline, de-identified datasets, ensuring compliance with GDPR and HIPAA principles [12], [13].

No real patient data or live healthcare systems were involved in experimentation. The design avoids intrusive data collection and prioritizes passive analysis of network flow data, reducing operational risk and ethical concerns.

These considerations ensure that the proposed framework is suitable for academic research and aligns with regulatory expectations.

4.10 Design Limitations

While the proposed design offers several advantages, certain limitations must be acknowledged. The use of a simulated dataset may not capture all nuances of real healthcare network traffic. Additionally, binary classification does not distinguish between different attack categories.

CHAPTER 5: IMPLEMENTATION (EXPERIMENTAL EVALUATION)

This chapter presents the implementation of the AI-driven intrusion detection framework designed in Chapter 4. The objective of this chapter is to describe the practical realization of the proposed system, including data preparation, model implementation, training procedures, and experimental execution. Unlike the previous chapters, which focused on theoretical justification and system design, this chapter concentrates on the technical implementation and experimental execution of the proposed intrusion detection system.

The implementation was conducted using a controlled experimental environment to ensure reproducibility, consistency, and fairness in model evaluation. This chapter documents the tools, libraries, and procedures used to implement the Random Forest and Support Vector Machine models and prepares the foundation for the evaluation and analysis of results presented in Chapter 6.

5.1 Implementation Environment

The experimental implementation was carried out using a cloud-based computational environment to support scalability and reproducibility. Google Colab was selected as the primary execution platform due to its ability to handle large datasets, availability of pre-configured machine learning libraries, and ease of replication across different systems. The implementation environment consisted of the following components:

- Programming language: Python
- Execution platform: Google Colab
- Data processing libraries: Pandas, NumPy
- Machine learning framework: Scikit-learn

- Visualization tools: Matplotlib

The use of Python and Scikit-learn is consistent with established practices in intrusion detection research and supports rapid experimentation and transparent model implementation [10], [34].

5.2 Dataset Preparation and Loading

5.2.1 Dataset Files Used

The implementation utilized a subset of the CICIDS2017 dataset to ensure computational feasibility while maintaining realistic traffic diversity. The following CSV files were loaded into the experimental environment:

- Monday-WorkingHours.pcap_ISCX.csv (benign traffic)
- Friday-WorkingHours-Afternoon-DDos.pcap_ISCX.csv (DDoS attacks)
- Thursday-WorkingHours-Morning-WebAttacks.pcap_ISCX.csv (web-based attacks)

These files were selected to represent both normal network behavior and common cyberattack scenarios relevant to healthcare environments [19].

5.2.2 Data Loading Process

Each CSV file was imported using the Pandas library and combined into a single dataset. Merging the datasets enabled unified preprocessing and ensured consistent feature representation across all traffic types. Initial inspection was performed to verify dataset dimensions, feature availability, and label distributions.

This step ensured that all subsequent preprocessing and model training procedures were applied uniformly, reducing the risk of experimental bias.

5.3 Data Preprocessing Implementation

Data preprocessing was also implemented as an essential step in the process, which ensured the quality and reliability of machine learning training and evaluation. Data preprocessing can adversely impact intrusion detection performance, as discussed in [32].

5.3.1 Cleaning and Feature Filtering

Non-numeric attributes, including flow identifiers, timestamps, and protocol labels, were removed to ensure compatibility with supervised learning algorithms. Features containing missing, infinite, or undefined values were identified and handled through filtering and replacement techniques. This cleaning process ensured numerical stability and prevented runtime errors during model training.

5.3.2 Label Encoding

The original CICIDS2017 dataset includes multiple attack labels. For this implementation, labels were converted into a binary format to simplify intrusion detection objectives:

- 0 – Benign traffic
- 1 – Malicious traffic

Binary classification aligns with practical intrusion detection use cases, where the main goal is to identify the presence of malicious activity rather than classify specific attack categories [12].

5.3.3 Feature Scaling

Feature scaling was applied using standardization techniques to normalize feature distributions. This step was particularly important for the Support Vector Machine model, which relies on distance-based optimization and is sensitive to feature magnitude [18].

Scaling was applied consistently across training and testing datasets to prevent data leakage.

5.3.4 Handling Class Imbalance

The dataset exhibited significant class imbalance, with benign traffic instances outnumbering attack instances. To mitigate this issue, Random Over-Sampling (ROS) was implemented to balance class distributions.

Oversampling increases the representation of minority classes and has been shown to improve recall and detection sensitivity in intrusion detection systems [33]. This step was critical for healthcare cybersecurity, where undetected attacks can have severe consequences.

5.4 Dataset Splitting Strategy

After preprocessing, the data was split into training and testing sets using the 80:20 split ratio. This ensures that the evaluation of the model is done using new, unseen data, hence generalization of the model is achieved, and the possibility of overfitting is eliminated.

Both Random Forest and Support Vector Machine models were trained and tested using identical data partitions. This controlled approach ensures fair comparison of model performance [10].

5.5 Random Forest Model Implementation

5.5.1 Model Configuration

The Random Forest classifier was implemented using the Scikit-learn library. The model was configured with a predefined number of decision trees and standard parameter settings to balance performance and computational efficiency.

Random Forest was selected because of robustness, ability to handle high-dimensional data, and resistance to overfitting [17]. These characteristics make it particularly suitable for intrusion detection tasks involving complex network traffic patterns.

5.5.2 Training Process

The Random Forest model was trained using the preprocessed training dataset. During training, each decision tree learned a subset of features and samples, enabling ensemble-based decision making.

Training time and system resource usage were monitored to ensure feasibility within the experimental environment.

5.5.3 Prediction and Output Generation

After training, the Random Forest model was used to classify instances in the test dataset.

Predictions were generated in binary format, indicating benign or malicious traffic.

These predictions were stored for subsequent evaluation and comparison against ground truth labels.

5.6 Support Vector Machine Implementation

5.6.1 Model Configuration

The Support Vector Machine classifier was implemented using Scikit-learn with a radial basis function (RBF) kernel. The RBF kernel enables SVM to model non-linear decision boundaries, which are common in network intrusion detection tasks [18].

Parameter selection followed recommended practices to balance classification performance and computational cost.

5.6.2 Training Process

The SVM model was trained using the same preprocessed training dataset as the Random Forest model. Due to the computational complexity of SVM, training time was higher compared to Random Forest, particularly when handling large datasets.

This observation highlights important feasibility considerations for healthcare environments with limited computational resources.

5.6.3 Prediction and Output Generation

Following training, the SVM model was used to classify the test dataset. Predictions were generated in binary format and stored for evaluation.

Using identical test data for both models ensured consistency and fairness in experimental comparison.

5.7 Performance Measurement Implementation

Model performance was evaluated using multiple metrics implemented through Scikit-learn's evaluation modules. The following metrics were calculated:

- Accuracy
- Precision
- Recall
- F1-score
- Confusion matrix

Recall was prioritized due to its significance in healthcare cybersecurity, where false negatives represent undetected attacks that may disrupt clinical services or compromise patient safety [7], [12].

Confusion matrices were generated to visualize classification outcomes and identify patterns of misclassification.

5.8 Experimental Reproducibility

Reproducibility was a key consideration throughout the implementation process. All experiments were executed in a consistent environment using fixed dataset partitions and standardized preprocessing steps.

The use of publicly available datasets, open-source libraries, and a cloud-based execution platform supports reproducibility and transparency, which are essential for academic research [31].

5.9 Implementation Challenges

Several challenges were encountered during implementation. Handling the large size of the CICIDS2017 dataset required careful memory management and preprocessing optimization. Training time for the Support Vector Machine model was significantly higher than for Random Forest, highlighting computational trade-offs between different algorithms.

Additionally, class imbalance required careful handling to prevent biased model predictions.

These challenges informed design decisions and are further discussed in the evaluation chapter.

5.10 Ethical and Security Considerations in Implementation

The implementation adhered strictly to ethical research standards. No real patient data or live healthcare systems were used. All experiments were conducted offline using de-identified, publicly available datasets.

This approach ensures compliance with data protection regulations, including GDPR and HIPAA, and eliminates risks associated with experimentation on operational healthcare systems [12], [13].

CHAPTER 6: EVALUATION OF RESULTS

This chapter describes the evaluation and analysis of the experimental results achieved through the implementation of the AI-powered intrusion detection system framework discussed in Chapter 5. The main aim of this chapter is to evaluate the efficacy of the proposed system in identifying malicious network traffic in a healthcare-focused cybersecurity environment.

The evaluation process involves comparing the performance of the Random Forest and SVM models based on conventional intrusion detection system evaluation criteria. Special attention is given to the recall and false negative values, which are of utmost significance in healthcare cybersecurity, as unidentified threats could affect healthcare services or patient safety [7], [12]

6.1 Evaluation Strategy

The evaluation strategy was designed to provide a fair, consistent, and meaningful comparison between the implemented machine learning models. Both Random Forest and SVM were evaluated using identical datasets, preprocessing pipelines, and evaluation metrics. Performance evaluation was conducted on the test dataset, which consisted of 20% of the preprocessed data and was not used during model training. This approach ensures that evaluation results reflect the models' ability to generalize to unseen data, rather than memorizing training instances [10].

6.2 Evaluation Metrics

To comprehensively assess intrusion detection performance, multiple metrics were employed:

- Accuracy
- Precision
- Recall

- F1-score
- Confusion matrix

Accuracy provides an overall measure of correct classifications but may be misleading in imbalanced datasets [15]. Precision reflects the proportion of correctly identified attacks among all detected attacks, while recall measures the proportion of actual attacks correctly detected. In healthcare cybersecurity, recall is the most critical metric, as false negatives represent undetected intrusions that may lead to service disruption or patient harm [7], [12]. The F1-score was used to balance precision and recall, and confusion matrices were analyzed to understand misclassification patterns.

6.3 Random Forest Model Performance

6.3.1 Overall Performance

The Random Forest Model showed excellent overall performance on all evaluation criteria. High accuracy was attained, which indicated good discrimination between benign and malicious traffic. More importantly, the model attained a high recall rate, which indicated that most instances of malicious traffic were detected.

The ensemble-based structure of Random Forest enables robust classification by aggregating decisions from multiple decision trees. This characteristic contributes to improved generalization and reduced sensitivity to noise in network traffic data [17].

6.3.2 Confusion Matrix Analysis

Analysis of the confusion matrix revealed a low number of false negatives, indicating that the model was effective at identifying attack traffic. False positives were present but within an acceptable range, reflecting a reasonable trade-off between detection sensitivity and alert volume.

In healthcare environments, this balance is essential. While excessive false positives can lead to alert fatigue, false negatives pose a more significant risk due to their potential impact on patient safety [7]. The Random Forest model's confusion matrix suggests a favorable balance aligned with healthcare security priorities.

Figure 6.1 illustrates the Confusion matrix for the Random Forest intrusion detection model. The results show a very low number of false negatives (25 instances), indicating that the model is highly effective in detecting malicious traffic. This is particularly important in healthcare cybersecurity, where undetected attacks may directly impact patient safety and service availability.

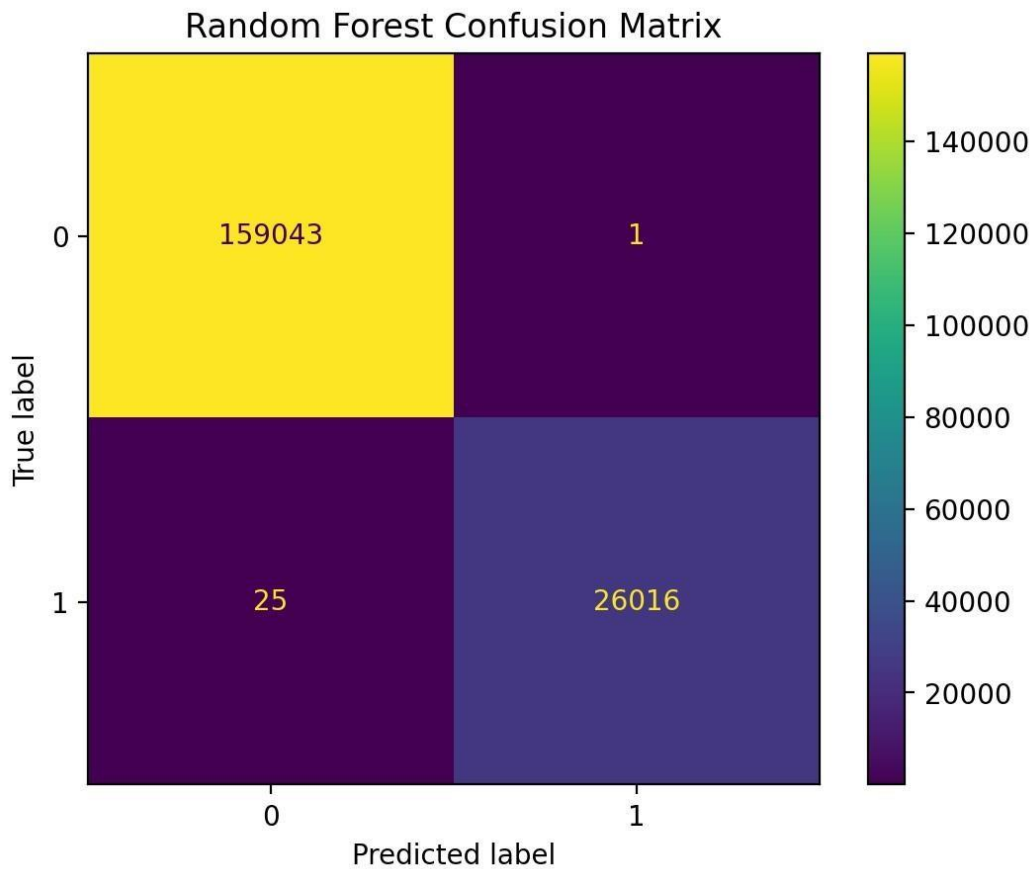


Figure 6.1: Confusion matrix for the Random Forest intrusion detection model.

6.3.3 Discussion of Results

The strong performance of Random Forest aligns with findings in existing intrusion detection research, where ensemble-based models consistently outperform single classifiers in terms of stability and recall [10], [17]. The model's ability to handle high-dimensional feature spaces and class imbalance further contributes to its effectiveness.

Additionally, feature importance measures generated by Random Forest support interpretability, enabling security analysts to understand which network attributes influence detection decisions. This capability enhances trust and auditability in regulated healthcare environments [25].

6.4 Support Vector Machine Model Performance

6.4.1 Overall Performance

The Support Vector Machine model also achieved satisfactory performance but exhibited notable differences compared to Random Forest. While overall accuracy remained high, recall was comparatively lower, indicating that a higher proportion of attack instances were misclassified as benign.

SVM's reliance on optimal decision boundaries can be sensitive to class imbalance and parameter selection, which may explain the observed performance differences [18], [34].

6.4.2 Confusion Matrix Analysis

The Confusion matrix for the SVM model revealed a higher number of false negatives relative to Random Forest. This finding is particularly significant in healthcare cybersecurity, where undetected attacks may have severe operational and safety consequences.

False positives were generally lower than those observed by Random Forest, suggesting that SVM was more conservative in classifying traffic as malicious. While this reduces alert volume, it also increases the risk of missed detections.

Figure 6.2 presents the confusion matrix for the Support Vector Machine Model. Compared to Random Forest, the SVM model exhibits a substantially higher number of false negatives (1,081 instances). This indicates reduced sensitivity to malicious traffic, which is a significant limitation

in healthcare cybersecurity environments where missed detections may have serious operational and safety consequences.

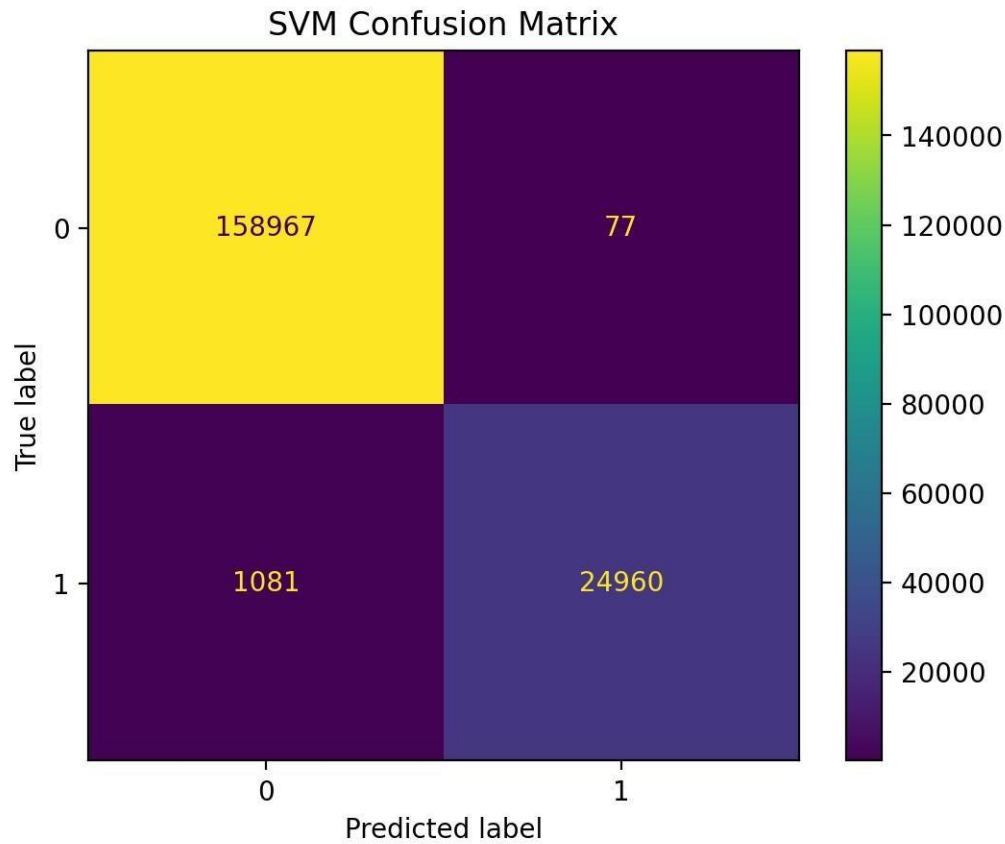


Figure 6.2: Confusion matrix for the Support Vector Machine intrusion detection model.

6.4.3 Discussion of Results

The performance of the SVM model highlights important trade-offs between precision and recall in intrusion detection systems. While SVM demonstrated strong classification boundaries, its sensitivity to imbalanced data and computational complexity limited its effectiveness in this context.

These findings are consistent with prior research, which suggests that SVM may require extensive parameter tuning and feature optimization to achieve optimal performance in largescale intrusion detection tasks [18], [34].

6.5 Comparative Analysis of Models

6.5.1 Accuracy and F1-Score Comparison

Both models achieved high accuracy, indicating that they were capable of distinguishing between benign and malicious traffic. However, accuracy alone does not capture the full performance picture due to class imbalance.

The F1-score comparison revealed that Random Forest achieved a higher balanced performance between precision and recall, making it more suitable for healthcare intrusion detection scenarios.

Figure 6.3 compares the performance of the Random Forest and Support Vector Machine models across key evaluation metrics. While both models achieve high accuracy, Random Forest demonstrates superior recall and F1-score, reinforcing its suitability for healthcare cybersecurity scenarios where minimizing false negatives is critical.

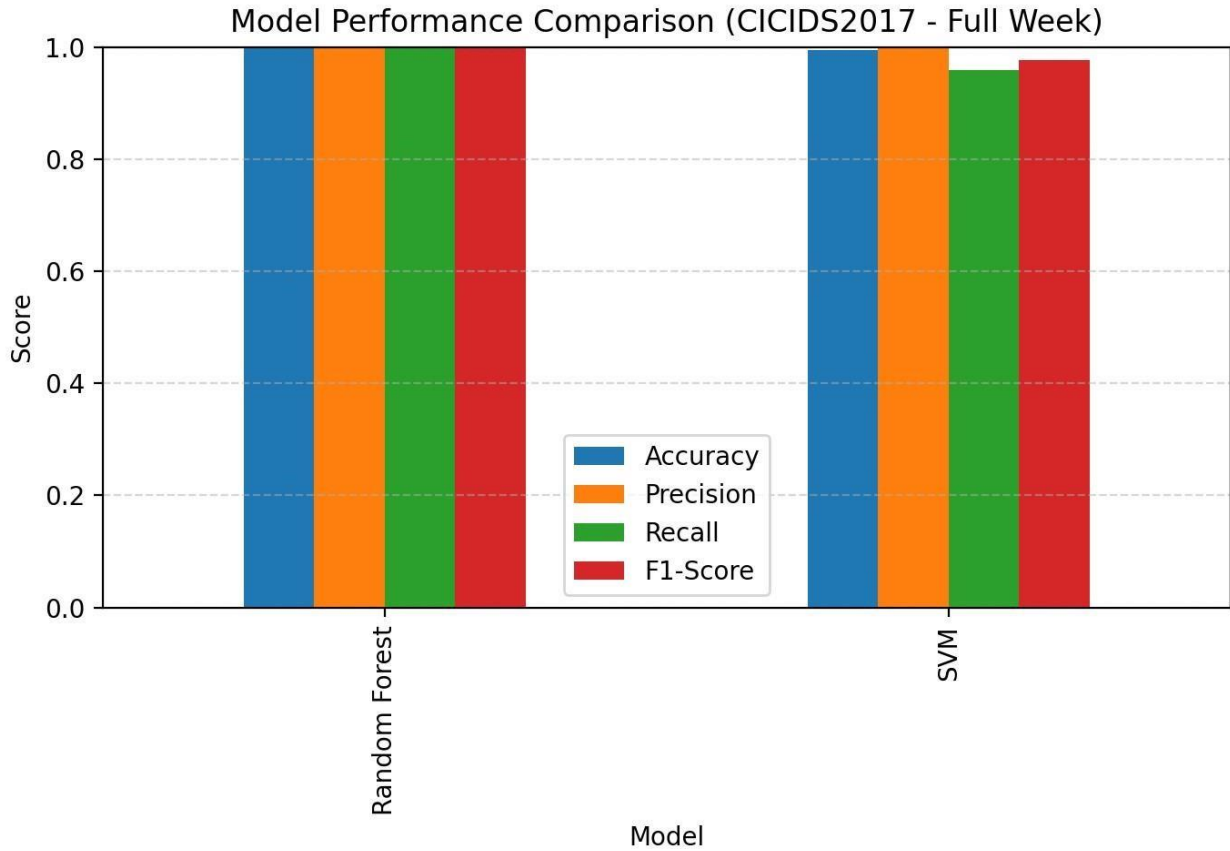


Figure 6.3: Performance comparison of Random Forest and Support Vector Machine models using the CICIDS2017 dataset.

6.5.2 Recall and False Negative Analysis

Recall analysis demonstrated that Random Forest outperformed SVM in detecting malicious traffic. The lower false negative rate observed for Random Forest is particularly important in healthcare contexts, where undetected intrusions may disrupt critical clinical services [7], [12]. This finding reinforces the prioritization of recall over precision in healthcare cybersecurity, as highlighted in the literature review.

6.5.3 Computational Efficiency

From a computational perspective, Random Forest demonstrated faster training and prediction times compared to SVM. The higher computational cost of SVM, particularly when using nonlinear

kernels, raises feasibility concerns for real-world healthcare deployments with limited computational resources [21].

This observation further supports the suitability of Random Forest for practical healthcare cybersecurity applications.

6.6.4 Quantitative Performance Summary

Table 6.1: Quantitative performance metrics for Random Forest and Support Vector Machine models.

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	0.9999	0.99996	0.99904	0.99950
Support Vector Machine	0.9938	0.99693	0.95849	0.97730

The results clearly indicate that Random Forest outperforms Support Vector Machine in terms of recall and F1-score. This confirms its effectiveness in minimizing false negatives, which is a critical requirement for intrusion detection in healthcare environments.

6.6 Alignment with Research Objectives

The evaluation results directly address the research objectives defined in Chapter 1:

- Effective detection of malicious traffic: Achieved by both models, with Random Forest demonstrating superior performance.
- Minimization of false negatives: Random Forest achieved a lower false negative rate, aligning with patient safety priorities.

- Interpretability and feasibility: Random Forest provided greater transparency and lower computational overhead.

These outcomes confirm that the proposed AI-driven intrusion detection framework meets its intended objectives.

6.7 Comparison with Existing Literature

The experimental findings are consistent with existing intrusion detection research. Prior studies have reported strong performance of Random Forest in detecting network intrusions, particularly in imbalanced datasets [10], [17]. The limitations observed for SVM are also well-documented in the literature [18].

Furthermore, the prioritization of recall and false negative reduction aligns with healthcare focused cybersecurity research emphasizing patient safety implications [7], [12].

6.8 Limitations of Evaluation

While the evaluation results are promising, certain limitations must be acknowledged. The experiments were conducted using a simulated dataset rather than live healthcare network traffic. Although CICIDS2017 is widely accepted, it may not capture all nuances of real-world healthcare environments.

Additionally, binary classification does not differentiate between specific attack types, which may limit the system's ability to support detailed incident response. These limitations are addressed through critical reflection in Chapter 7.

CHAPTER 7: CONCLUSIONS AND RECOMMENDATIONS

This chapter includes the conclusions that were arrived at from the research, as well as recommendations for further research and application of the research findings. The purpose of this chapter is to bring together all the conclusions that were presented in the previous chapters, as well as evaluate how successful the research objectives have been met. Contrary to the previous chapters, this chapter does not include new research findings or results from any experiment that was conducted but rather includes critical thoughts about the research findings and how they can be applied to healthcare cybersecurity.

This chapter includes conclusions from the research, evaluation of how the research findings meet the research objectives, the contribution of the research, limitations of the research, as well as recommendations for further research.

7.1 Summary of Research

The primary aim of this research was to design, implement, and evaluate an AI-driven intrusion detection framework suitable for healthcare cybersecurity environments. This aim was motivated by the increasing frequency and severity of cyberattacks targeting healthcare systems and the direct impact such attacks can have on patient safety, service continuity, and data confidentiality [7], [12]. A comprehensive literature review (Chapter 2) identified key challenges in healthcare cybersecurity, including legacy systems, IoMT vulnerabilities, and the limitations of traditional intrusion detection mechanisms. The review also highlighted the growing adoption of machine learning techniques in cybersecurity and identified gaps related to healthcare-specific validation, recall prioritization, and interpretability.

To address these gaps, a quantitative, experimental research methodology was adopted (Chapter 3). The CICIDS2017 dataset was selected as a realistic and ethically appropriate data source, and two supervised machine learning models—Random Forest and Support Vector Machine—were implemented and evaluated using a controlled experimental framework (Chapters 4 and 5).

7.2 Achievement of Research Objectives

The research objectives defined in Chapter 1 have been successfully addressed through the design, implementation, and evaluation of the proposed framework.

Objective 1: Design an AI-driven intrusion detection framework for healthcare systems This objective was achieved through the development of a layered system architecture that integrates data acquisition, preprocessing, machine learning-based detection, and evaluation components. The design prioritized healthcare-specific requirements, including interpretability, scalability, and ethical compliance [25], [12].

Objective 2: Implement and evaluate supervised machine learning models

Both Random Forest and Support Vector Machine models were successfully implemented using a consistent preprocessing and evaluation pipeline. The implementation demonstrated that supervised machine learning techniques can effectively detect malicious network traffic in a healthcare-oriented cybersecurity context [10], [17].

Objective 3: Compare model performance using healthcare-relevant metrics

The evaluation results presented in Chapter 6 demonstrated that Random Forest outperformed SVM across key metrics, particularly recall and false negative rate. This finding is significant given the patient safety implications of undetected cyberattacks in healthcare environments [7].

Objective 4: Assess feasibility and suitability for healthcare environments The results indicated that Random Forest offers a favorable balance between detection performance, interpretability, and computational efficiency. These characteristics support its suitability for deployment in healthcare environments with limited resources and strict regulatory requirements [17], [21].

7.3 Key Findings and Contributions

The research makes several important contributions to the field of healthcare cybersecurity. Firstly, the study provides empirical evidence that ensemble-based machine learning models, such as

Random Forest, are particularly effective for intrusion detection in healthcare-oriented scenarios. The superior recall performance observed for Random Forest aligns with healthcare security priorities and reinforces findings reported in existing literature [10], [17].

Secondly, the research demonstrates that careful preprocessing, class imbalance handling, and metric selection are critical for meaningful evaluation of intrusion detection systems. The prioritization of recall over accuracy reflects a healthcare-centric evaluation perspective that is often underemphasized in generic cybersecurity research [12].

Thirdly, the proposed framework emphasizes interpretability and feasibility, addressing concerns related to trust and regulatory compliance in healthcare environments. By avoiding opaque deep learning models and focusing on transparent algorithms, the study contributes a practical and defensible approach to AI-driven healthcare cybersecurity [25].

7.4 Practical Implications

The findings of this research have several practical implications for healthcare organizations and cybersecurity practitioners.

Healthcare organizations can benefit from machine learning-based intrusion detection systems to improve their capacity to handle and respond to cyber threats. The findings indicate that detection methods based on Random Forest can be effective and computationally feasible.

Additionally, the emphasis on recall and false negative reduction highlights the need for healthcare organizations to prioritize patient safety when evaluating cybersecurity solutions. Security tools should be assessed not only on accuracy but also on their ability to minimize missed detections, which may have severe clinical consequences [7].

The framework proposed in this study may serve as a foundation for developing or enhancing network-based intrusion detection solutions in healthcare environments, particularly where legacy systems and IoMT devices present persistent security challenges.

7.5 Limitations of the Research

Despite its contributions, this research has several limitations that must be acknowledged. Firstly, the experiments were conducted using a simulated dataset rather than live healthcare network traffic. Although CICIDS2017 is widely accepted in intrusion detection research, it may not fully capture the complexity and variability of real-world healthcare networks [27]. Secondly, the use of binary classification limits the system's ability to differentiate between specific attack types. While binary detection is sufficient for identifying malicious activity, more granular classification could support improved incident response and forensic analysis.

Thirdly, the study focused exclusively on supervised learning approaches. While this aligns with the research objectives, it limits the system's ability to detect entirely novel or zero-day attacks that are not represented in the training data [32].

7.6 Recommendations for Future Research

Based on the findings and limitations of this study, several recommendations for future research are proposed.

Future studies could explore the integration of unsupervised or hybrid intrusion detection approaches to enhance detection of previously unseen attacks. Combining supervised and unsupervised techniques may improve adaptability to evolving threat landscapes. Further research could also evaluate the proposed framework using real-world healthcare network traffic, subject to ethical approval and regulatory compliance. Such validation would strengthen confidence in the system's applicability to operational healthcare environments. Additionally, future work could investigate multi-class classification, enabling differentiation between various attack types. This enhancement would support more detailed threat analysis and response strategies.

Finally, the incorporation of explainable AI (XAI) techniques, such as SHAP or LIME, could further improve transparency and trust in intrusion detection decisions, particularly in regulated healthcare settings [25].

7.7 Final Conclusion

This research has demonstrated that AI-driven intrusion detection systems can play a critical role in enhancing cybersecurity within healthcare environments. Through the design, implementation, and evaluation of a supervised machine learning-based framework, the study has shown that Random Forest offers superior performance in terms of recall, interpretability, and feasibility when compared to Support Vector Machine.

By addressing healthcare-specific requirements and prioritizing patient safety, this research contributes a practical and academically rigorous approach to healthcare cybersecurity. The findings provide a strong foundation for future research and offer valuable insights for healthcare organizations seeking to strengthen their cyber defense capabilities.

CHAPTER 8: REFLECTION

This chapter provides a critical reflection on the research journey undertaken during the completion of this MSc project. Reflection is an essential component of postgraduate research, as it enables evaluation of personal learning, skill development, challenges encountered, and professional growth achieved throughout the project lifecycle. Unlike previous chapters, which focused on technical design and empirical evaluation, this chapter adopts a reflective perspective to assess the research process and its contribution to academic and professional development.

The reflection considers methodological decision-making, technical skill acquisition, ethical awareness, time management, and alignment with professional practice in cybersecurity. It also evaluates how the project has enhanced readiness for future research or industry roles within the field of cybersecurity.

8.1 Reflection on Research Planning and Methodology

One of the most significant learning outcomes of this project was the development of structured research planning skills. At the outset, defining a feasible and academically rigorous research scope proved challenging. Healthcare cybersecurity is a broad domain, and narrowing the focus to AI-driven intrusion detection required careful consideration of dataset availability, ethical constraints, and technical feasibility.

The decision to adopt a quantitative, experimental research methodology proved appropriate, as it aligned well with the technical nature of intrusion detection research and allowed objective evaluation of system performance [30]. Through this process, I gained a deeper understanding of how research paradigms influence methodological choices and the importance of aligning research objectives with evaluation strategies.

Selecting suitable datasets and machine learning models requires critical engagement with existing literature. The literature review phase strengthened my ability to assess prior work, identify research gaps, and justify methodological decisions based on evidence rather than assumptions.

This process reinforced the importance of grounding technical implementation in established academic research [10].

8.2 Technical Skills Development

This project significantly enhanced my technical competencies in cybersecurity and machine learning. Implementing Random Forest and Support Vector Machine models using Python and Scikit-learn improved my practical understanding of supervised learning techniques and their application to real-world security problems.

A key technical challenge involved handling large-scale datasets and addressing class imbalance issues inherent in intrusion detection data. Implementing preprocessing techniques such as feature scaling and oversampling improved my understanding of how data quality directly impacts model performance [33]. This experience highlighted that successful machine learning systems depend as much on data preparation as on algorithm selection.

Additionally, working within a cloud-based environment (Google Colab) improved my ability to design reproducible experiments and manage computational resources effectively. These skills are directly transferable to professional cybersecurity and data science roles, where scalable and reproducible analysis is essential.

8.3 Challenges Encountered and Problem-Solving

Several challenges were encountered throughout the project. One of the most significant was managing the complexity of the CICIDS2017 dataset, which required careful preprocessing to avoid errors and ensure consistency. Resolving issues related to missing values, feature compatibility, and memory constraints strengthened my problem-solving abilities and attention to detail.

Another challenge involved balancing detection performance with interpretability and feasibility. While deep learning models could potentially achieve higher accuracy, their lack of transparency and higher computational requirements made them less suitable for healthcare environments. This

challenge reinforced the importance of making context-aware technical decisions rather than optimizing performance metrics in isolation [25].

Time management was also a critical challenge, particularly when balancing literature review, implementation, and evaluation phases. Developing a structured workflow and adhering to milestones improved my ability to manage complex projects under time constraints, a skill essential for both academic research and professional practice.

8.4 Ethical Awareness and Professional Responsibility

Ethical considerations played a central role in shaping this research. The use of publicly available, de-identified datasets ensured compliance with data protection regulations such as GDPR and HIPAA [12], [13]. This reinforced my understanding of ethical responsibility in cybersecurity research, particularly in healthcare contexts where misuse of data can have serious legal and moral implications.

Through this project, I developed a heightened awareness of the ethical dimensions of AI-driven security systems, including issues related to transparency, accountability, and trust. The emphasis on interpretability and avoidance of opaque models reflects a responsible approach to deploying AI in safety-critical environments [25].

This ethical awareness will inform my future professional conduct, particularly when working with sensitive data or developing security solutions that may impact individuals and organizations.

8.5 Reflection on Learning Outcomes and Professional Growth

This MSc project has contributed significantly to my academic and professional development. Academically, it strengthened my ability to conduct independent research, critically evaluate literature, and communicate complex technical concepts in a structured and scholarly manner.

Professionally, the project enhanced my readiness for roles in cybersecurity and data analysis. Skills developed during this research—such as intrusion detection, machine learning

implementation, experimental evaluation, and ethical risk assessment—are directly relevant to roles such as Security Analyst or Cybersecurity Engineer.

Furthermore, the project increased my confidence in tackling complex technical problems and reinforced the importance of continuous learning in the rapidly evolving cybersecurity landscape.

8.6 Reflection on Limitations and Areas for Improvement

Reflecting on the project, several areas for improvement have been identified. While the experimental evaluation produced meaningful results, validation using real-world healthcare network data would strengthen the applicability of the findings. Future work could also explore hybrid or unsupervised learning approaches to enhance detection of novel attacks [32]. Additionally, greater emphasis on automated explainability tools could further improve transparency and trust in intrusion detection decisions. These reflections highlight opportunities for further learning and research beyond the scope of this MSc project.

8.7 Final Reflective Summary

In conclusion, this MSc project has been a valuable and challenging learning experience that has significantly enhanced my technical, analytical, and professional skills. Through the design, implementation, and evaluation of an AI-driven intrusion detection framework for healthcare cybersecurity, I have developed a deeper understanding of both the technical and ethical dimensions of cybersecurity research.

The knowledge and skills gained during this project provide a strong foundation for future academic research and professional practice in cybersecurity. This reflective process has reinforced the importance of critical thinking, ethical responsibility, and continuous development in addressing complex security challenges in healthcare and beyond.

REFERENCES

- [1] J. Adler-Milstein and A. Jha, "HITECH Act drove large gains in hospital electronic health record adoption," *Health Affairs*, vol. 36, no. 8, pp. 1416–1422, 2017
- [2] A. Botta, W. de Donato, V. Persico, and A. Pescapé, "Integration of cloud computing and internet of things: A survey," *Future Generation Computer Systems*, vol. 56, pp. 684–700, 2016, doi: 10.1016/j.future.2015.09.02
- [3] R. McLeod and R. Dolezel, "Cyber-analytics: Modeling factors associated with healthcare data breaches," *Decision Support Systems*, vol. 108, pp. 57–68, 2018
- [4] M. Sittig and H. Singh, "A new sociotechnical model for studying health information technology in complex adaptive healthcare systems," *Qual. Saf. Health Care*, vol. 19, no. Suppl 3, pp. i68–i74, 2010, doi: 10.1136/qshc.2010.042085
- [5] Verizon, "Data Breach Investigations Report," Verizon Enterprise, 2023
- [6] J. Martin et al., "Hospital ransomware attacks and patient outcomes," *JAMA Network Open*, vol. 4, no. 12, e2139126, 2021, doi: 10.1001/jamanetworkopen.2021.39126.
- [7] A. K. Sittig, H. Singh, "Patient safety and cybersecurity," *BMJ Qual. Saf.*, vol. 25, no. 12, pp. 954–957, 2016, doi: 10.1136/bmjqs-2015-004804.
- [8] D. He, S. Zeadally, and N. Kumar, "Security issues and challenges in IoT-based healthcare systems," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3731–3745, 2018
- [9] ENISA, "Cybersecurity challenges in healthcare," European Union Agency for Cybersecurity, 2021, doi: 10.2824/857356.
- [10] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cybersecurity intrusion detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2016, doi: 10.1109/COMST.2015.2494502.
- [11] M. Kim et al., "Deep learning-based intrusion detection," *Computers & Security*, vol. 92, 2020, doi: 10.1016/j.cose.2020.101743.
- [12] HIPAA, "Health Insurance Portability and Accountability Act," U.S. Dept. of Health & Human Services, 2013.
- [13] ISO/IEC 27001, "Information security management systems," ISO, 2017.
- [14] Coveware, "Ransomware attack trends in healthcare," Quarterly Ransomware Report, 2021
- [15] N. Japkowicz and S. Stephen, "The class imbalance problem," *Intelligent Data Analysis*, vol. 6, no. 5, pp. 429–449, 2002, doi: 10.3233/IDA-2002-6504.
- [16] N. Behl and K. Behl, *Cyberwar: The Next Threat to National Security and What to Do About It*, Oxford University Press, 2017
- [17] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001
- [18] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995, doi: 10.1007/BF00994018.
- [19] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset," *ICISSP*, 2018, doi: 10.5220/0006639801080116.

- [20] K. Martin, A. Shacham, and R. Hill, "Cybersecurity failures and patient harm in healthcare," *Health Affairs*, vol. 39, no. 10, pp. 1702–1709, 2020, doi: 10.1377/hlthaff.2020.00698
- [21] A. Armbrust et al., "A view of cloud computing," *Communications of the ACM*, vol. 53, no. 4, pp. 50–58, 2010, doi: 10.1145/1721654.1721672.
- [22] J. A. Kruse et al., "Cybersecurity concerns with telehealth," *JMIR Med. Inform.*, vol. 5, no. 4, e94, 2017, doi: 10.2196/medinform.7825.
- [23] L. Da Xu, W. He, and S. Li, "Internet of Things in industries: A survey," *IEEE Trans. Ind. Informatics*, vol. 10, no. 4, pp. 2233–2243, 2014, doi: 10.1109/TII.2014.2300753
- [24] E. Ullah et al., "A survey on intrusion detection in IoMT," *IEEE Access*, vol. 7, pp. 86659– 86679, 2019, doi: 10.1109/ACCESS.2019.2924828.
- [25] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv:1702.08608, 2017
- [26] F. T. Liu, K. M. Ting, and Z. Zhou, "Isolation forest," *IEEE ICDM*, 2008, doi: 10.1109/ICDM.2008.17.
- [27] A. H. Lashkari et al., "Characterization of CICIDS2017 dataset," *Information Systems Security*, 2017
- [28] J. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, Sage, 2014
- [29] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *NeurIPS*, 2017
- [30] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, 2009.
- [31] A. Géron, *Hands-On Machine Learning with Scikit-Learn*, O'Reilly, 2019.
- [32] M. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, vol. 41, no. 3, 2009, doi: 10.1145/1541880.1541882
- [33] N. Chawla et al., "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.
- [34] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009, doi: 10.1109/TKDE.2008.239.

Appendix

CodeLink:<https://github.com/Mahabub-alam-tuhin/-AI-Driven-Intrusion-Detection-Framework-for-Healthcare-Cybersecurity-.git>