بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

Sudan University of Science Technology

College of Engineering

School of Electronics Engineering

Intrusion Detection in Mobile Networks Using Random Forest and CICIDS2017 Dataset

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of B.Eng. (Honors) in Electronics Engineering.

Prepared By:

1. Safia Abdelrazeg Mohamed Abdalla
2. Ahmed Noor Alhuda Alzain Alemam
3. Alaa Hassan Ahmed Suliman

Supervisor:

Prof. Dr. Rashid A SAEED

2026

الآية

{وَءَاخِرُ دَعۡوَىٰهُمۡ أَنِ ٱلۡحَمۡدُ لِلَّهِ رَبِّ ٱلۡعَٰلَمِينَ}

سورة يونس الآية (10)

Dedication

All praise is due to Allah, whose blessings make every good deed possible, and peace and blessings be upon our Prophet Muhammad (PBUH)

This work is dedicated to our beloved college, which has been a true home of knowledge and growth, guiding me through our academic journey.

To our respected supervisor, **Dr. Rashid A. Saeed,** for his valuable guidance, constant support, and inspiring encouragement throughout this research - We are deeply grateful.

To our dear parents, whose love, sacrifices, and prayers have been our greatest source of strength and motivation.

And to everyone who supported us with words, ideas, or encouragement – THANK YOU for being part of this journey.

May Allah make this effort sincere for His sake and a source of benefit for students and researchers in the pursuit of knowledge.

Acknowledgement

We would like to express our sincere gratitude to Sudan University of Science and Technology for providing us with the opportunity, resources, and academic environment to complete this research project successfully.

Our deepest appreciation goes to our supervisor for his invaluable guidance, continuous support, and constructive feedback throughout every stage of this study. His encouragement and expertise have been instrumental in shaping this work.

We are also extending our heartfelt thanks to all the participants who took part in the questionnaire, whose cooperation and responses greatly contributed to the success of this research.

To our beloved parents, your endless love, patience, and unwavering belief in us have been our greatest motivation. We are profoundly thankful for your support.

Finally, our sincere thanks go to our friends for their encouragement, insightful discussions, and moral support, which helped us overcome every challenge along the way.

Abstract

The rapid growth of mobile networks and the increasing number of cyberattacks have created significant security challenges, making effective intrusion detection essential. To address these challenges, this project proposes a Machine Learning-Based Intrusion Detection System (IDS) for mobile networks using the CICIDS2017 dataset and the Random Forest algorithm. The CICIDS2017 dataset contains realistic network traffic, including both benign and malicious activities with different types of attacks. The proposed system preprocesses the dataset by removing missing and invalid data, handling outliers, encoding labels, and scaling the features. Recursive Feature Elimination (RFE) is applied to select the most relevant features before training the Random Forest model. The model is trained to distinguish between normal and malicious network traffic and to classify different attack types. The performance of the proposed system is evaluated using Accuracy, Precision, Recall, F1-Score, Specificity, and AUC-ROC. The results demonstrate that Random Forest can effectively detect and classify network attacks with high performance. The proposed system can therefore contribute to improving the security, reliability, and protection of mobile network environments against cyber threats.

المستخلص

مع الانتشار المتزايد لشبكات الهاتف المحمول وارتفاع عدد الهجمات الإلكترونية، أصبحت الحاجة إلى أنظمة فعالة لكشف التسلل أمراً مهماً. لذلك يقترح هذا المشروع نظاماً لكشف التسلل في شبكات الهاتف المحمول باستخدام تعلم الآلة، بالاعتماد على مجموعة بيانات CICIDS2017 وخوارزمية Random Forest. يتم في النظام معالجة البيانات من خلال تنظيف البيانات، وإزالة القيم غير الصالحة والمفقودة، ومعالجة القيم الشاذة، وترميز البيانات وتوحيد الخصائص. كما يتم استخدام RFE لاختيار الخصائص الأكثر أهمية قبل تدريب نموذج Random Forest. يهدف النموذج إلى التمييز بين حركة الشبكة الطبيعية والخبيثة وتصنيف أنواع الهجمات المختلفة. ويتم تقييم أداء النظام باستخدام Accuracy, Precision, Recall, F1-Score, Specificity وAUC-ROC. وتوضح النتائج أن خوارزمية Random Forest قادرة على اكتشاف وتصنيف هجمات الشبكة بكفاءة عالية، مما يساهم في تحسين أمن وموثوقية شبكات الهاتف المحمول.

Table of Content

List of Tables

List of Figures

Chapter One
Introduction

1.1 Introduction

In recent years, mobile networks have witnessed a tremendous evolution in the size and complexity of security threats. These networks have become particularly vulnerable to attacks due to their dynamic and decentralized nature, which makes traditional security measures insufficient. These threats are no longer limited to traditional methods but have evolved to include new attack techniques that are more diverse and precise, making them difficult to detect using older systems. As a result, it has become essential to adopt advanced solutions capable of addressing these constantly changing threats (Anantvalee and Wu, 2007).

Mobile networks, widely used in communications, social networks, and financial transactions, require robust protection against a variety of cyber threats. As the use of mobile networks expands, the threat landscape has become more complex, leading to an increased need for effective and adaptive detection systems. An Intrusion Detection System (IDS) refers to technology designed to monitor and analyze network traffic to detect malicious activities that may threaten the integrity of the system, thus contributing to enhancing the security of mobile networks (Khraisat et al., 2019).

One promising approach to enhancing intrusion detection systems in mobile networks is the integration of machine learning algorithms. Machine learning techniques, specifically algorithms like Random Forest, can significantly improve the accuracy of detecting intrusion attempts by analyzing large amounts of network data and identifying attack patterns. These systems are trained using datasets like CICIDS2017, which contain labeled examples of network traffic data with various types of attacks. Machine learning models can learn to detect both known and unknown threats, ensuring data security and protection from threats (Almseidin et al., 2018).

Furthermore, these algorithms enable IDS to adapt to new attack trends without the need for explicit reprogramming. They can learn from real-time network traffic and update their detection capabilities accordingly. This adaptive feature is critical in handling the evolving and complex nature of security threats. Additionally, machine learning systems excel at reducing false positives, a problem that often affects traditional intrusion detection models, thus improving the overall efficiency and reliability of network security. With advancements in artificial intelligence, it may become possible to predict threats before they occur, significantly enhancing security levels (Almseidin et al., 2018).

1.2 Problem Statement

Mobile networks face several critical challenges due to the unique characteristics of mobile environments. First, mobile devices and systems suffer from limited computational and communication power, which makes it difficult to deploy complex intrusion detection models in real time. Second, environmental factors such as interference, weather, terrain, and physical obstacles significantly degrade signal quality, leading to reduced detection accuracy. Third, many existing approaches rely on simplified theoretical models, such as the unit disk model, which fail to capture the realistic conditions of mobile networks where phenomena like shadow fading occur, thus reducing the reliability of detection performance. In addition, intrusion detection based on a single source often results in high false alarm rates, creating the need for multi-source or multi-device validation mechanisms. Furthermore, many mobile applications, including financial services and healthcare, require immediate response to detected intrusions, which is challenging under resource constraints. Finally, packet loss and unstable connections in mobile environments further affect the accuracy of detection systems, leaving critical vulnerabilities unaddressed.

1.3 Aim & Objectives

1.3.1 Research Aim

The main aim of this research is to develop an effective IDS for mobile networks using the CICIDS2017 dataset and the Random Forest machine learning algorithm.

1.3.2 Research Objectives

The objectives of this research are:

1) To improve intrusion detection: Detect malicious and abnormal network traffic accurately.
2) To classify network attacks: Identify different types of attacks contained in the CICIDS2017 dataset.
3) To apply Random Forest: Develop and train a Random Forest model for network intrusion detection.
4) To improve detection performance: Reduce false positives and false negatives in attack detection.
5) To preprocess network traffic: Clean, normalize, and prepare the dataset for machine learning.
6) To evaluate the proposed system: Measure the performance using Accuracy, Precision, Recall, F1-Score, Specificity, and AUC-ROC.
7) To enhance mobile network security: Provide an effective approach for detecting potential threats in mobile network environments.

1.4 Significance of the Study

Intrusion detection is an important research topic with many potential applications. Along with intrusion prevention, response, and tolerance, intrusion detection is one tool that can defend against the real-world cyber attacks threatening critical systems. These attacks include the Stuxnet attack on Iranian engineering facilities, proof-of-

concept attacks on insulin pumps and cardiac devices, the DoS attack on a German power grid operator, the exfiltration attack on a Spanish power grid vendor, and the exfiltration attack on US UAVs.

Mobile and cyber-physical systems are critical wireless network systems because of their human impact. For instance, in Mobile Ad Hoc Networks (MANETs) or mobile telephony networks, malicious behavior can damage the network by violating confidentiality, integrity, availability, authenticity, non-repudiation, or privacy. For example, a node in a mobile telephony network may masquerade as another node in order to defeat the integrity of the billing function. Such attacks highlight the necessity of employing effective IDS in mobile networks, where adversaries can be selfish or malicious.

Because wireless IDSs must cope with transient nodes, mobility, noise, incomplete data, and unique adversarial models (e.g., jamming without physical access), their role in mobile networks is particularly significant. Without intrusion detection, adversaries that penetrate the network can remain hidden and cause severe disruption to critical services, endangering both infrastructure and human safety

1.5 Scope and Limitation

This study is scoped within the field of IDS in mobile networks, with a particular focus on the application of machine learning techniques to improve detection accuracy and efficiency. While intrusion detection has been widely studied in traditional computer and cloud networks, the dynamic nature of mobile networks—characterized by high mobility, limited resources, and diverse attack surfaces—introduces unique security challenges that demand specialized IDS approaches.

The scope of this research includes the design, implementation, and evaluation of an IDS model using the Random Forest algorithm as the core classification technique. To ensure reliability and reproducibility, the study employs the CICIDS2017 dataset,

which provides a comprehensive and realistic set of modern attack scenarios and normal traffic flows.

This research is limited to analyzing the performance of Random Forest in terms of detection accuracy, precision, recall, and false positive rates, and comparing its effectiveness against existing approaches. Broader aspects such as the integration of IDS into real-world mobile network infrastructures, energy consumption, and real-time deployment challenges are acknowledged but remain outside the direct scope of this study.

This research is limited to analyzing the performance of Random Forest in terms of detection accuracy, precision, recall, and false positive rates, and comparing its effectiveness against existing approaches. Broader aspects such as the integration of IDS into real-world mobile network infrastructures, energy consumption, and real-time deployment challenges are acknowledged but remain outside the direct scope of this study.

1.6 Study Structure

Chapter 1 provided an introduction to the study, while Chapter 2 reviewed the theoretical background and previous studies related to IDS, network security, the Random Forest algorithm, and the CICIDS2017 data. Chapter 3 explained the methodology and procedures for building and evaluating the model. Chapter 4 presented Results and Discussion, while Chapter 5 presented the conclusion of the study, Recommendations, and references.

Chapter Two
Literature Review

2.1 Introduction

This chapter reviewed academic research and theories relevant to the development of IDS. It started by defining the fundamental concepts and classifications of IDS, followed by a discussion of the evolution towards machine learning-based approaches. Subsequently, it surveys common machine learning techniques, with a specific focus on ensemble methods and the Random Forest algorithm. The chapter also discusses the unique challenges of intrusion detection in mobile network environments and the advantages of the CICIDS2017 dataset used in this project. Finally, the chapter concludes by synthesizing the literature and identifying the specific research gap that this work seeks to address.

2.2 Fundamental Concepts of Intrusion Detection Systems (IDS)

2.2.1 Definition and Objectives:

An intrusion is broadly defined as "any kind of unauthorized activities that cause damage to an information system," particularly those threatening the confidentiality, integrity, or availability of information. To defend against these intrusions, an IDS is employed. An IDS is "a software or hardware system that identifies malicious actions on computer systems in order to allow for system security to be maintained". Its primary goal is to identify malicious network traffic and computer usage that bypass traditional firewalls, thereby safeguarding the core security principles of availability, integrity, and confidentiality (Khraisat et al., 2019).

2.2.2. Classification by Detection Method: SIDS vs. A-IDS:

IDS can be broadly classified based on their detection methodology. According to Khraisat et al. (2019), IDS systems can be broadly categorized into two groups: Signature-based Intrusion Detection System (SIDS) and Anomaly-based Intrusion Detection System (A-IDS).

a) Signature-based IDS (SIDS):

These systems, also known as Knowledge-based or Misuse Detection systems, are based on pattern-matching techniques to find a known attack... In SIDS, matching methods are used to find a previous intrusion. In other words, when an intrusion signature matches the signature of a previous intrusion that already exists in the signature database, an alarm signal is triggered" (Khraisat et al., 2019).

b) Anomaly-based IDS (A-IDS):

In contrast, A-IDS has garnered significant research interest for its ability to overcome the limitations of SIDS. In this approach, "a normal model of the behavior of a computer system is created using machine learning, statistical-based, or knowledge-based methods. Any significant deviation between the observed behavior and the model is regarded as an anomaly, which can be interpreted as an intrusion" (Khraisat et al., 2019).

2.2.3. Classification by Data Source: HIDS vs. NIDS:

IDS technologies are also differentiated by their data sources. "In terms of data sources, there are generally two types of IDS technologies, namely Host-based IDS (HIDS) and Network-based IDS (NIDS)" (Khraisat et al., 2019). HIDS operates at the host level, inspecting "data that originates from the host system and audit sources, such as operating system, Windows server logs, firewall logs, application system audits, or database logs," making it effective for detecting insider attacks that may not generate network traffic. Conversely, NIDS functions at the network level, monitoring "the network traffic that is extracted from a network through packet capture, NetFlow, and other network data sources" (Khraisat et al., 2019).

2.3 The Evolution Towards Machine Learning-Based IDS:

2.3.1. Limitations of Traditional Signature-Based Approaches:

The foundation of traditional cybersecurity has long been "signature-based" and "rule-based" mechanisms. However, while "these systems excel at identifying previously observed threats, they often fail when confronted with novel or highly sophisticated attacks" (Raja et al., 2025).

This critical shortcoming, described as "the absence of fully automated, robust detection processes capable of keeping pace with evolving threats", highlights why conventional methods are increasingly inadequate against novel attack vectors, necessitating the adoption of more advanced frameworks (Raja et al., 2025).

2.3.2. Anomaly-Based Detection and Machine Learning:

In response to these limitations, machine learning has emerged as a powerful alternative. Raja et al. (2025) position ML as "a transformative tool in cybersecurity, particularly for automating threat detection processes that traditionally depend on manual analysis." The integration of ML aims "to outperform traditional tools in scalability and detection accuracy" by offering "scalable and actionable threat-detection capabilities". This positions ML, particularly within anomaly-based detection frameworks, as the promising path forward for modern IDS (Raja et al., 2025).

2.4 The Role of AI and MLT Advancing Anomaly-Based Intrusion Detection Systems

Artificial Intelligence (AI), and particularly Machine Learning (ML), has witnessed remarkable growth in recent years, and it is increasingly regarded as a promising tool in network monitoring and management. This trend is also reflected in academic research on cybersecurity; however, the practical adoption of AI/ML techniques remains limited due to several technical and organizational challenges. These challenges include issues related to data quality and availability, reproducibility of

models, performance and resource requirements, as well as concerns regarding privacy, explainability, and usability (Nisioti et al., 2018).

IDS represent a fundamental component of cybersecurity infrastructures and have been extensively studied for decades. Traditional IDS are largely signature-based, which limits their ability to cope with novel and unknown attacks. This has led to growing interest in network-based intrusion detection systems (NIDS) enhanced by AI/ML, which aim to analyze and classify network packets or flows in order to detect malicious activities more effectively (Nisioti et al., 2018).

Despite significant progress in this field, major challenges still hinder the wide-scale deployment of AI/ML-based IDS. From a technical perspective, difficulties remain in securing realistic and representative datasets for training and evaluation, and in achieving a balance between detection accuracy and computational efficiency in resource-constrained environments. From a practical standpoint, challenges such as model explainability and user trust persist, as increasing model complexity often reduces confidence among non-experts (Abdulganiyu et al., 2023).

In the context of mobile networks, these challenges become even more critical due to their unique characteristics, such as the limited computational capabilities of end devices, the dynamic and decentralized network topology, and the susceptibility of communication quality to environmental factors. Therefore, studying AI/ML-based intrusion detection within mobile network environments is essential to adapt these techniques to the specific security requirements of mobile systems and to ensure their feasibility for real-world deployment. With the rapid growth of computer networks and connected devices, cyberattacks targeting confidentiality, integrity, and availability (CIA) have increased significantly. This has highlighted the importance of IDS as a core security technology for monitoring, analyzing, and detecting malicious activities that exploit systems (Amouri and Brahmi, 2020).

2.5 Challenges and Future Directions

IDS can be broadly categorized into Host-based IDS (HIDS), which monitor activities on individual hosts, and Network-based IDS (NIDS), which analyze traffic across the network. Detection methods generally fall into two groups: signature-based detection, which identifies known attacks using stored patterns, and anomaly-based detection, which leverages statistical, machine learning (ML), and deep learning (DL) methods to identify unknown or zero-day attacks. While signature-based approaches are effective against previously observed threats, anomaly-based IDS offer greater adaptability but often face challenges such as false positives and the need for large, high-quality datasets (Debar et al., 2000).

In recent years, several systematic reviews and surveys have examined IDS from multiple perspectives, including detection methods, datasets (e.g., CICIDS2018, UNSW-NB15), and application domains such as IoT, SDN, and automotive systems. These reviews highlight persistent challenges such as data imbalance, overfitting, lack of real-world datasets, evaluation metrics, and the complexity of ML/DL models. They also emphasize emerging directions, including explainability, usability, and integration with technologies like blockchain and natural language processing (NLP). Overall, the literature indicates that despite significant progress, IDS research still faces fundamental issues related to data quality, scalability, and practical deployment, particularly in dynamic environments such as mobile and IoT networks (Abdulganiyu et al., 2023).

Intrusion detection has emerged as a critical research area due to the growing dependence on wireless and mobile networks, which are increasingly vulnerable to cyber threats. Traditional security mechanisms such as intrusion prevention are often insufficient, especially against insider attacks or adversaries who bypass authentication. To address these challenges, research on IDS in wireless environments has expanded significantly (Abdulganiyu et al., 2023).

Mitchell and Chen (2014) provide a comprehensive survey of wireless intrusion detection techniques, classifying them based on target system, detection method, collection process, trust model, and analysis approach. Their study highlights the applicability of IDS techniques across various mobile and wireless systems, including wireless local area networks (WLANs), wireless personal area networks (WPANs), wireless sensor networks (WSNs), mobile ad hoc networks (MANETs), wireless mesh networks (WMNs), mobile telephony, and cyber-physical systems (CPSs).

The literature emphasizes that wireless IDSs face unique challenges compared to wired environments. These include node mobility, incomplete or noisy datasets due to network partitioning and interference, and adversarial strategies such as jamming. Consequently, IDS solutions must incorporate wireless-specific features such as signal strength, spatiotemporal dynamics, and cooperative detection models. Research findings suggest that while several IDS approaches (e.g., anomaly-based, signature-based, trust-based) have been extensively studied for certain mobile environments, significant research gaps remain. For instance, some IDS techniques have been underexplored in mobile telephony and CPSs, despite their critical role in human safety. Furthermore, IDS solutions must increasingly integrate intrusion tolerance mechanisms to maintain service availability even when adversaries penetrate the system (Fan et al., 2011).

2.6 Systematic Analysis of Wireless Intrusion Detection

Overall, the background literature reveals that intrusion detection in mobile networks is a rapidly evolving domain. Existing studies provide valuable taxonomies, lessons learned, and insights into strengths and weaknesses of different approaches, while also identifying future research directions such as adapting underutilized IDS techniques to new wireless environments, enhancing resilience

against sophisticated adversaries, and leveraging emerging technologies like machine learning for adaptive detection (Amouri and Brahmi, 2020).

With the recent increase in internet usage, the volume of sensitive and confidential data transmitted across networks has grown significantly. Due to vulnerabilities in existing security systems, attackers have exploited these gaps to intrude into networks, thereby compromising confidentiality and system operations. IDS, a core area in cybersecurity, have thus been deployed to monitor and analyze network traffic to detect and report malicious activities (Abdulganiyu et al., 2023).

A considerable number of review papers have addressed various approaches to network intrusion detection. However, most of these works adopted a non-systematic approach, focusing on comparisons of existing techniques without providing an in-depth analytical synthesis of their methodologies and performance. Many reviews primarily examined anomaly-based IDS, with emphasis on deep learning models, while signature-based and hybrid (signature + anomaly) methods received comparatively little attention (Abdulganiyu et al., 2023).

To overcome this limitation, systematic reviews following the PRISMA guidelines have emerged. These reviews analyze contributions across anomaly-, signature-, and hybrid-based approaches to provide a comprehensive state-of-the-art understanding of IDS. Recent studies retrieved from multiple reputable databases—including ScienceDirect, SpringerNature, IEEE, MDPI, Hindawi, PeerJ, and Taylor & Francis—highlight both progress and ongoing challenges in IDS research. From over 776 identified studies, 71 were analyzed in detail, leading to the identification of unexplored research areas and unresolved challenges. This body of work suggests that while anomaly- and deep learning-based IDS have shown promise, hybrid approaches and lightweight methods tailored for wireless and mobile environments require further exploration. These findings point to high-impact future research

directions aimed at building more effective and adaptive IDS models (Abdulganiyu et al., 2023).

Information systems are becoming more integrated into our lives. As this integration deepens, the importance of securing these systems increases. Because of lower installation and maintenance costs, many of these systems are largely networked by wireless means. In order to identify gaps and propose research directions in wireless network intrusion detection research, we survey the literature in this area. Our approach is to classify existing contemporary wireless intrusion detection system (IDS) techniques based on target wireless network, detection technique, collection process, trust model, and analysis technique. We summarize the pros and cons of the same or different types of concerns and considerations for wireless intrusion detection with respect to specific wireless networks, including wireless local area networks (WLANs), wireless personal area networks (WPANs), wireless sensor networks (WSNs), ad hoc networks, mobile telephony, wireless2.Prec. Next, we summarize the and 8ˡleast tudied wireless IDS techniques in the literature, identify research gaps, and analyze the rationale for the degree of their treatment. Finally, we identify what has been explored and provide suggestions for future research (Pinto et al., 2020; Anantvalee and Wu, 2007).

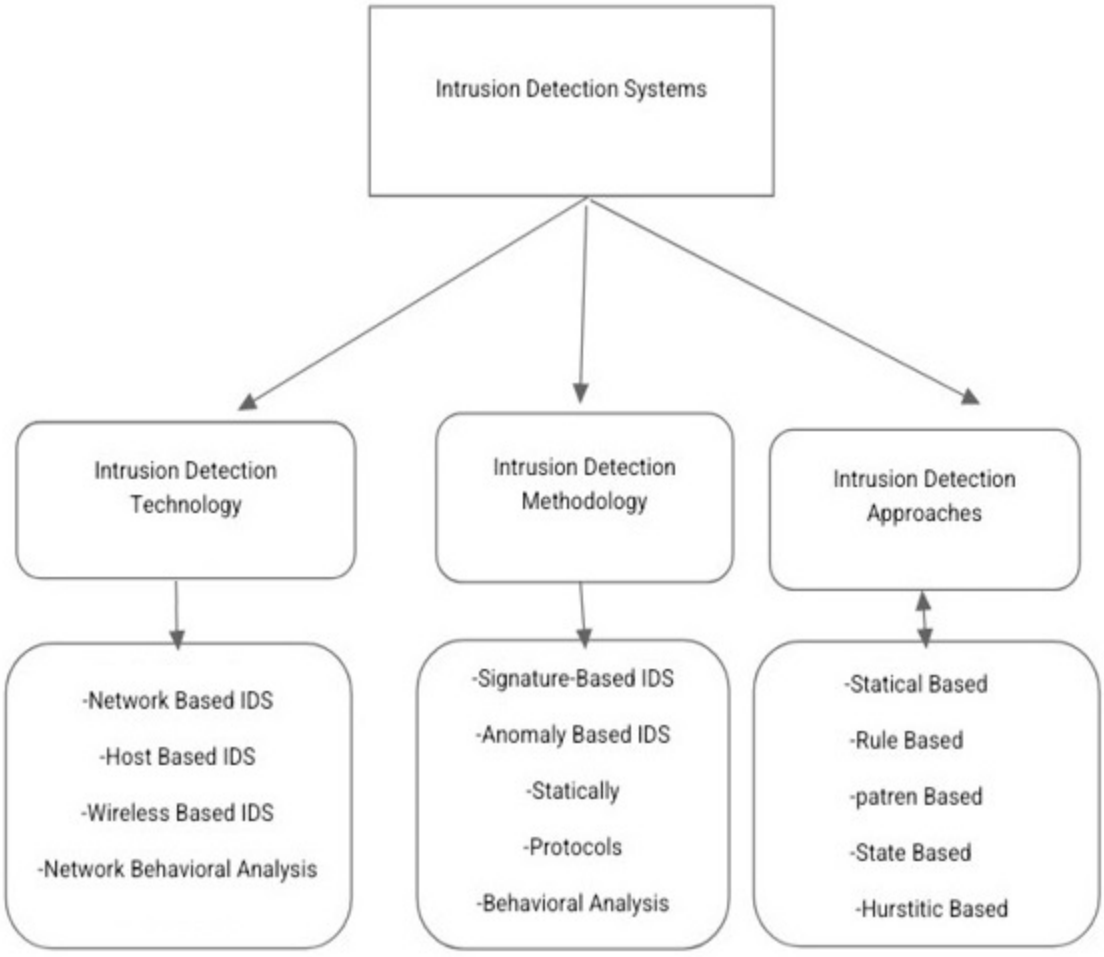

Figure 2 1 Classification of intrusion detection

2.7 Security Challenges and Intrusion Detection in Mobile Networks

The complexity of mobile networks is increasing at a rapid pace. Modern mobile communication systems involve millions of interconnected devices, heterogeneous protocols, and vast amounts of data exchange between multiple entities such as mobile devices, sensors, cloud services, and infrastructure. This increasing complexity introduces significant security and privacy challenges. To enable proper protection, monitoring, and incident response, data must be collected, analyzed, and processed in a secure and efficient manner, which highlights the importance of IDS (Dietz et al., 2024).

Mobile networks use various hardware, software, and communication technologies. Hardware includes mobile devices, base stations, and IoT nodes, each varying in computational and storage capabilities. Mobile devices and sensors generate diverse types of data, such as location, identity, and user activity. Software ranges from applications running on end-user devices to complex network functions in the cloud or edge. Data storage involves sensitive information including cryptographic keys, authentication logs, and system configurations (Dietz et al., 2024).

Mobile networks rely on multiple communication technologies such as 3G/4G/5G, WiFi, Bluetooth, and ad-hoc links. Transmission occurs across heterogeneous environments, making the system vulnerable to cyberattacks. Gateways and routing nodes play a critical role in ensuring connectivity but are also common targets for adversaries. Furthermore, the extensive communication between mobile devices, base stations, and cloud-based services increases the attack surface. These complexities and interconnections are similar to other distributed systems, such as vehicular or IoT networks, but mobile networks face additional challenges due to their dynamic topology, resource constraints, and high mobility of nodes. As a result, traditional prevention-based security techniques are insufficient. IDS are therefore essential as complementary security (Anantvalee and Wu, 2007).

2.8 Related Work

Mitchell and Chen (2014) conducted a comprehensive survey on intrusion detection in wireless network applications. They aimed to identify gaps and proposed research directions in wireless network intrusion detection research. In their study, they classified existing contemporary wireless Intrusion Detection System (IDS) techniques based on target wireless network, detection technique, collection process, trust model, and analysis technique. They summarized the advantages and disadvantages of various IDS approaches and examined their applicability with

respect to specific attributes of wireless networks, including Wireless Local Area Networks (WLANs), Wireless Personal Area Networks (WPANs), Wireless Sensor Networks (WSNs), ad hoc networks, mobile telephony, Wireless Mesh Networks (WMNs), and Cyber Physical Systems (CPSs).

Mitchell and Chen reviewed intrusion detection systems (IDS) and identified research gaps and future directions. They explained the importance of IDS in protecting critical systems from cyberattacks and distinguished between malicious and selfish behaviors. Their review classified IDS research based on six criteria: target system, detection technique, data collection, trust model, analysis technique, and system type. They analyzed 60 IDS techniques and summarized their main similarities, differences, and research findings.

Finally, they identified future research areas such as adapting existing IDS techniques to underexplored wireless environments, improving trust-based security models, and integrating dynamic intrusion tolerance mechanisms (Mitchell and Chen, 2014).

Through their systematic review, Mitchell and Chen (2014) provided a foundational reference for understanding the evolution, classification, and challenges of intrusion detection systems in wireless and mobile networks.

Abdulganiyu et al. (2023) conducted a systematic literature review for network intrusion detection systems (IDS). With the recent increase in internet usage, the amount of important, sensitive, and confidential individual and corporate data has grown. Due to gaps in existing security systems, attackers attempted to intrude into networks, thereby gaining access to essential and confidential information, which caused harm to system operations and affected data confidentiality.

2.8.1 A Systematic Review of Intrusion Detection Systems (IDSs):

To counter these possible attacks, intrusion detection systems (IDSs), which were an essential branch of cybersecurity, were employed to monitor and analyze network traffic, thereby detecting and reporting malicious activities. Abdulganiyu and colleagues observed that a large number of review papers had covered different approaches for intrusion detection in networks, most of which followed a non-systematic approach that merely compared existing techniques without providing an in-depth analytical synthesis of methodologies and performances to give a complete understanding of the state of IDS.

Nonetheless, they found that many of these reviews investigated anomaly-based IDSs with greater emphasis on deep-learning models, while signature-based and hybrid-based (signature + anomaly-based) approaches received minimal focus. Hence, by adhering to the principles of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), they reviewed existing contributions on anomaly-, signature-, and hybrid-based approaches to provide a comprehensive overview of the state of the art in network IDS.

They retrieved articles from seven databases (ScienceDirect, SpringerNature, IEEE, MDPI, Hindawi, PeerJ, and Taylor & Francis) that spanned various reputable journals and conference proceedings. Among the 776 pieces of literature identified, 71 were selected for analysis and synthesis to answer the research questions. Based on their findings, they identified unexplored study areas and unresolved research challenges. Finally, they concluded their work by presenting promising, high-impact directions for future research to develop better IDS models.

In recent years, significant research efforts have been devoted to extending intrusion detection mechanisms to mobile and vehicular network environments. These networks were characterized by high mobility, dynamic topology, and wireless communication, which introduced additional security challenges compared to traditional wired systems. Several studies investigated both onboard and offloaded

intrusion detection approaches to enhance the reliability and responsiveness of such systems.

In the onboard approach, the intrusion detection model was trained offline to learn normal behavioral patterns or the signatures of known attacks based on various monitored features ‹During runtime, the model monitored these features and applied predefined decision rules, achieving relatively low latency and high accuracy for previously observed attack scenarios (Abdulganiyu et al., 2023; Maseer et al., 2021; Sharafaldin et al., 2018). However, this approach was less effective when facing unseen or complex attack patterns.

2.8.2 Improving Intrusion Detection Systems in mobile networks:

To overcome these limitations, alternative approaches were proposed that offloaded the intrusion detection process to powerful external infrastructures such as cloud or edge computing systems. This strategy not only reduced the computational burden on the mobile or vehicular nodes but also enabled the utilization of more sophisticated detection techniques, including deep learning and large-scale behavioral analysis.

Several studies demonstrated that learning normal communication patterns and identifying deviations could effectively detect intrusions in both mobile and vehicular networks. For example, Cho and Shin (2017) proposed a voltage fingerprinting method to identify the origin of attacks within in-vehicle CAN bus networks. Similarly, Moore et al. (2018) developed an anomaly detection algorithm based on message refresh rates, while Martinelli et al. (2019) utilized fuzzy-rough nearest neighbor classification to distinguish between legitimate and injected CAN messages. These works collectively emphasized the growing trend of integrating machine learning and cloud-assisted computing to improve the accuracy,

adaptability, and scalability of intrusion detection systems in mobile network environments.

Several recent studies focused on improving intrusion detection systems (IDSs) in mobile and wireless networks using artificial intelligence and machine learning techniques. These studies relied on publicly available benchmark datasets to train and evaluate their models. Among the most commonly used datasets were NSL-KDD (2009), UNSW-NB15, and CICIDS2017.

The NSL-KDD dataset was widely utilized in early IDS research; however, it contained outdated attack types that no longer represented the evolving threat landscape, thus limiting its reliability for modern mobile environments. Similarly, the UNSW-NB15 dataset was proposed to address some of these shortcomings by including newer attack categories and more realistic network traffic. Despite these improvements, it still suffered from several drawbacks, such as imbalanced data distribution and synthetic traffic patterns that reduced its realism.

In contrast, the CICIDS2017 dataset emerged as one of the most suitable datasets for evaluating AI-based IDSs. It included realistic and up-to-date attack scenarios based on application-layer protocols such as HTTP, HTTPS, FTP, SSH, and email, and accurately modeled normal user behavior. Due to these characteristics, several recent works adopted CICIDS2017 to train and test IDS models in both mobile and IoT network contexts. This dataset enabled a more precise evaluation of intrusion detection performance under realistic traffic conditions, making it a strong choice for modern mobile networks.

In the context of mobile and wireless networks, intrusion detection was considered a crucial mechanism for ensuring data confidentiality, integrity, and network reliability. As the number of connected mobile devices and Internet of Things (IoT) platforms increased, data became more vulnerable to malicious attacks, which led researchers to develop intelligent intrusion detection systems (IDSs). Two main

types of IDSs were defined in previous studies: network-based IDS (NIDS) and host-based IDS (HIDS). NIDS monitored network traffic and identified malicious activities through packet and flow analysis, whereas HIDS focused on detecting host-level anomalies. Detection techniques were generally classified into signature-based and anomaly-based methods. Signature-based IDSs could detect known attacks but failed to identify zero-day intrusions due to their dependence on predefined patterns. In contrast, anomaly-based IDSs (AIDSs) utilized statistical, machine learning (ML), or deep learning (DL) techniques to model normal behavior and identify deviations representing novel attacks.

2.8.3 Dataset and Model in Intrusion Detection Systems

Several researchers applied ML and DL models to enhance anomaly-based intrusion detection (ML-AIDS) in mobile and vehicular environments. Supervised and unsupervised ML algorithms, such as Support Vector Machine (SVM), Artificial Neural Network (ANN), Decision Tree (DT), Random Forest (RF), Naïve Bayes (NB), K-Means, and Self-Organizing Map (SOM), were employed to classify network traffic into normal and abnormal categories. These models were trained and validated using standard benchmark datasets to ensure reliability and reproducibility. Earlier studies primarily relied on the DARPA 1998, KDD Cup 1999, and NSL-KDD (2009) datasets. However, these datasets contained outdated attacks, redundant records, and imbalanced traffic distributions that did not reflect current network conditions. The UNSW-NB15 dataset was later introduced to overcome some of these limitations, but it still suffered from issues related to unrealistic attacks and limited diversity. Consequently, researchers shifted toward using the CICIDS2017 dataset, which was recognized as one of the most comprehensive and up-to-date datasets for training and testing IDS models.

The CICIDS2017 dataset, developed by the Canadian Institute for Cybersecurity, provided realistic network traffic that included both benign and modern attack scenarios. It captured user behaviors and communication patterns over multiple application-layer protocols, such as HTTP, HTTPS, FTP, SSH, and Email, which were representative of mobile and IoT environments. This dataset contained a variety of recent attacks, including Distributed Denial of Service (DDoS), infiltration, brute force, web-based attacks (e.g., SQL injection, XSS), and botnet activity. Therefore, it offered a realistic environment for evaluating ML- and DL-based IDS models in mobile networks.

Analytical studies using CICIDS2017 demonstrated high classification accuracy, with some models achieving up to 99.9% precision and a false alarm rate (FAR) as low as 0.012%. For instance, distance-based ML algorithms such as k-Nearest Neighbor (k-NN) and K-Means were applied to analyze model complexity, while hybrid deep learning approaches, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), showed improved generalization in detecting multi-class attack scenarios. However, most existing studies focused primarily on accuracy and neglected other critical performance metrics, such as precision, recall, F1-score, detection delay, and computational efficiency.

2.8.4 Feature Selection for Intrusion Detection in Mobile Networks:

Although many IDS models achieved remarkable detection accuracy, benchmarking methodologies for ML-AIDS models in mobile and vehicular networks remained inconsistent. Only a limited number of studies compared their models with state-of-the-art systems, and few addressed multi-class classifications or evaluated IDS performance under real-time mobile conditions. Therefore, researchers concluded that a standardized benchmarking methodology was still needed for testing and

evaluating the performance of ML- and DL-based AIDS models using modern datasets.

Feature selection has become an essential component in developing efficient IDSs, especially for mobile and wireless environments where computational and energy resources are limited. Recent studies have demonstrated that selecting the most relevant features can significantly improve detection accuracy while reducing processing overhead. For instance, Wang et al. (2019) analyzed large-scale network traffic and identified ten significant features using a hybrid filter–wrapper approach, achieving a detection rate of 99.8% and a false alarm rate of 0.34%. Similarly, Ambusaidi et al. (2016) proposed the Flexible Mutual Information Feature Selection (FMIFS) algorithm, which enhanced the accuracy and efficiency of Least-Squares Support Vector Machine (LS-SVM)-based IDSs.

Other researchers have combined multiple selection techniques to achieve higher performance. Chen et al. (2020) introduced a Tree-Seed Algorithm (TSA) to eliminate redundant features and enhance K-Nearest Neighbor (KNN) classifier performance. Then applied a Discrete Differential Evolution (DDE) algorithm alongside C4.5 to produce 16 relevant features with 99.92% accuracy. Peng et al. (2008) developed the Feature Selection with Ant Colony Optimization (FACO) approach to further improve classification accuracy, and presented an FWP-SVM-GA model combining Support Vector Machines (SVMs) and Genetic Algorithms to optimize detection rate, accuracy, and training time.

These studies collectively demonstrate that efficient feature selection can substantially enhance IDS performance by removing redundant and irrelevant data attributes. However, most of the existing works have been evaluated using traditional datasets such as KDD CUP 99, which contain only a limited number of features and records. In mobile network contexts—where traffic patterns, node mobility, and attack dynamics differ significantly—the performance and scalability

of these feature selection techniques remain underexplored. Therefore, applying and validating such optimized feature selection approaches in mobile and wireless intrusion detection systems represents a promising research direction

2.8.5 Improving Intrusion Detection Systems in mobile and wireless environments:

Recent research has focused on improving the efficiency of intrusion detection systems (IDSs) through advanced feature selection and ensemble learning techniques, which are particularly relevant to mobile and wireless environments where computational resources are limited. For instance, Yulianto et al. (2023) combined the Synthetic Minority Oversampling Technique (SMOTE), Principal Component Analysis (PCA), and Ensemble Feature Selection (EFS) to enhance the performance of an AdaBoost-based IDS using the CICIDS-2017 dataset. The authors reported that their hybrid method outperformed the traditional SVM-based approach in terms of accuracy, precision, recall, and F1-score. Such ensemble-based and dimensionality-reduction strategies can be highly beneficial for mobile and wireless intrusion detection systems, where minimizing computational overhead and maintaining high detection accuracy are critical due to dynamic topology and limited node resources.

On the other hand, while many researchers have adopted Information Gain (IG) as a feature selection technique, there has been limited discussion regarding how to determine the minimum weight or rank score from the IG results. This score defines the degree of relevance between each feature and the class label. For instance, researchers in (Aburomman and Reaz, 2016) and (Yulianto et al., 2023) selected features with scores above 0.4 and 0.001, respectively, whereas the study in (Chen et al., 2020) considered a minimum score threshold of 0.8. In contrast, researchers in (Peng et al., 2008) removed features iteratively and applied the classification

algorithm repeatedly to identify the optimal accuracy. However, this process is computationally expensive, especially for large datasets with a high number of features, making it less practical for resource-constrained mobile IDS environments. Overall, these studies highlight the importance of efficient feature selection and dimensionality reduction techniques in improving IDS performance. However, their application in mobile and wireless network scenarios remains limited, underscoring the need for further research on lightweight and adaptive approaches suited to dynamic and energy-constrained mobile systems.

Anantvalee and Wu (2007) conducted a comprehensive survey on intrusion detection in Mobile Ad Hoc Networks (MANETs). They explained that in previous years, the use of MANETs had become widespread in many applications, including mission-critical ones, and therefore, security had become one of the major concerns in such networks. They noted that, due to the unique characteristics of MANETs, prevention methods alone had not been sufficient to make them secure; thus, detection needed to be added as another line of defense before an attacker could breach the system.

Anantvalee and Wu classified the architectures of IDS that had been introduced for MANETs. They reviewed and compared existing IDS approaches corresponding to these architectures and then provided directions for future research. Their work contributed significantly to understanding the structure and challenges of intrusion detection in mobile ad hoc environments.

2.9 Summary of the chapter

This chapter discusses IDS in mobile networks, focusing on the role of the Random Forest algorithm in improving attack detection accuracy. It reviews IDS types, such as network-based and host-based systems, and highlights the importance of using the CICIDS2017 dataset, which includes attacks like DoS, DDoS, and Brute Force.

The chapter also covers data preprocessing steps, including feature selection and handling class imbalance. Finally, it reviews related studies and shows that combining Random Forest with CICIDS2017.

Chapter Three
Methodology

3.1 Introduction

This chapter describes the methodology used to develop and evaluate an IDS for mobile networks using the CICIDS2017 dataset and the Random Forest algorithm. It presented the procedures followed for data preparation, preprocessing, model training, and testing. The performance of the proposed model is evaluated using accuracy, precision, recall, and F1-score to assess its effectiveness in identifying network intrusions.

3.2 Research Design

This study adopts a quantitative and experimental research design to develop and evaluate an IDS for mobile networks using the CICIDS2017 dataset and the Random Forest algorithm. The quantitative approach allows objective measurements to be made and model performance to be evaluated using measurable criteria, while the experimental aspect allows the model to be trained, tested, and evaluated in structured and repeatable conditions. The study follows sequential stages that include data processing, Feature selection, model training and testing, and performance evaluation. This design was chosen to systematically and objectively evaluate the performance of the proposed model and provide experimental evidence of its effectiveness in detecting attempts to infiltrate mobile networks.

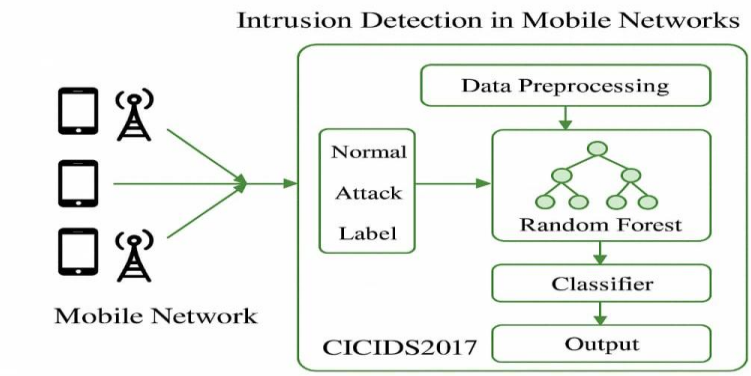

Figure 3 1 Random Forest-based Intrusion Detection System

3.3 Research Approach

This study adopted a quantitative and experimental approach to develop and evaluate an IDS for mobile networks using the Random Forest algorithm and the CICIDS2017 dataset. The approach focuses on quantitatively analyzing network traffic data, training a machine learning model capable of distinguishing between normal network traffic and malicious network traffic, and then evaluating its performance using specific performance metrics. CICIDS2017 was selected as a benchmark dataset containing normal traffic and various attacks, allowing testing of the model's ability to classify different patterns of network behavior.

The research methodology begins with the data collection and preparation phase, where CICIDS2017 data files are obtained, and the required data is collected for use in building an intrusion detection model. Next, Data Preprocessing is performed to improve data quality and reduce issues that may affect the performance of the machine learning model. This stage includes examining the data, processing missing or invalid values, removing inappropriate records or properties, handling duplicate data, as well as converting the data into an image suitable for processing by the Random Forest algorithm.

After data processing, Feature Selection is performed to identify the most important characteristics in the intrusion detection process and reduce unnecessary or repetitive characteristics. Choosing appropriate properties helps reduce model complexity and reduce processing time, while preserving information important to the classification process. The dataset is then divided into Training Set and Testing Set data according to a specific segmentation strategy, with the training data used to build the model, while the test data is used to evaluate its ability to handle data not used during training.

In the Model Development and Training phase, Random Forest, an Ensemble Learning algorithm, is applied that relies on building a set of Decision Trees and

combining their results to arrive at a final classification decision. The model is trained using processed training data with the aim of learning patterns that distinguish between Benign Traffic and the different types of attacks present in the dataset. The trained model is then used to classify the test data and determine whether network traffic is normal or associated with offensive activity.

Finally, a Performance Evaluation is performed to measure the efficiency of the Random Forest model in intrusion detection. This is done using a range of quantitative metrics, including Accuracy to measure the proportion of correct ratings, Precision to measure the accuracy of positive predictions, Recall to measure the model's ability to detect offensive situations, F1-score to measure the balance between Precision and Recall, as well as False Positive Rate (FPR) when needed. The confusion matrix can also be analyzed to understand the model's performance in classifying different categories. Through these successive stages, the research aims to provide an experimental and systematic evaluation of the performance of the Random Forest algorithm in Network Intrusion Detection using CICIDS2017 data within the scope of the study related to mobile networks.

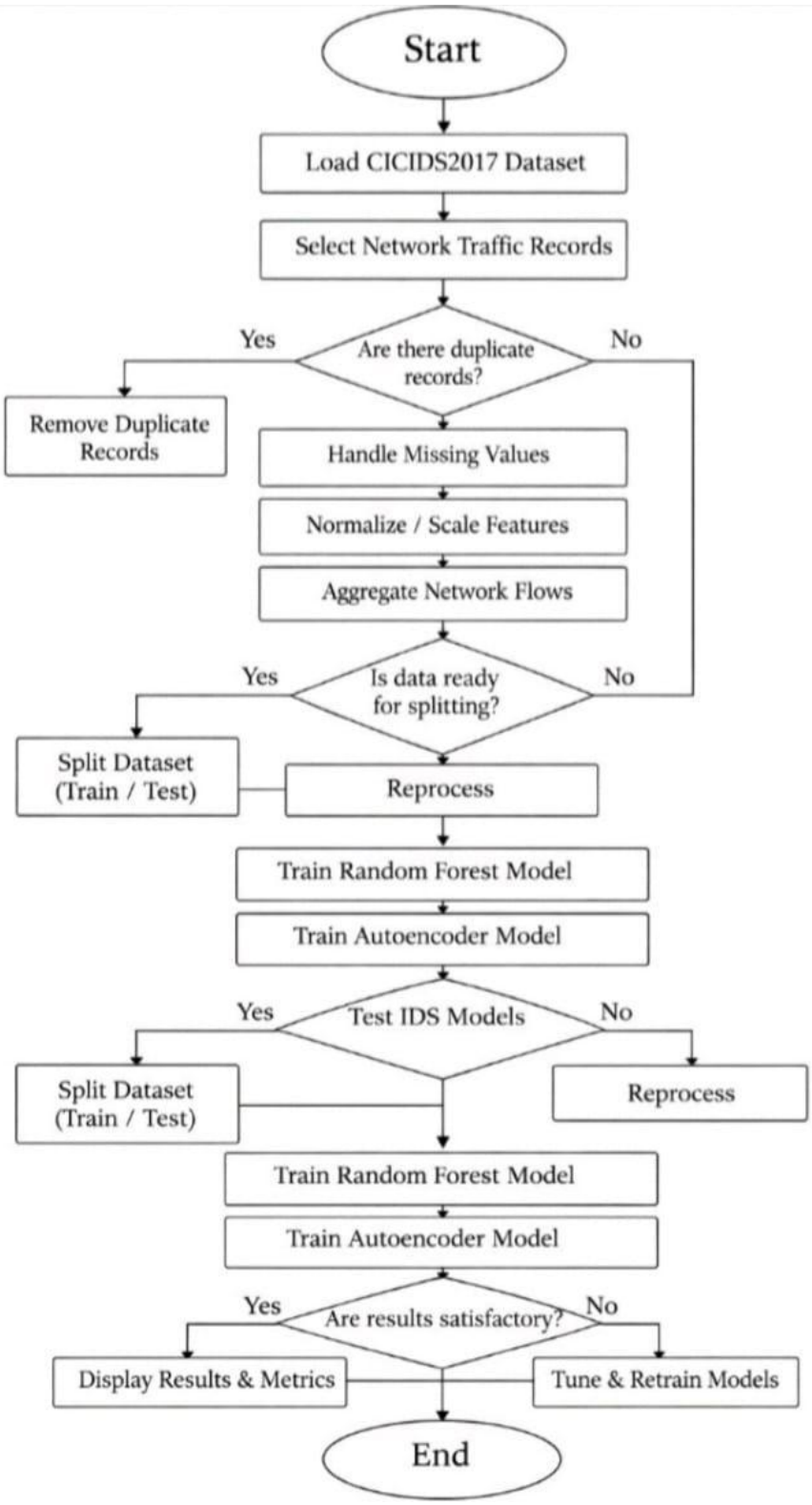

Figure 3 2 Machine Learning Algorithm

3.4 Research Modeling

In this study, research modeling is employed to establish a systematic framework for designing and evaluating IDS in mobile networks. The modeling process begins by defining the characteristics of the mobile network environment, including its dynamic topology and communication behavior. Potential intrusion scenarios are then identified to determine the patterns of normal and abnormal activities.

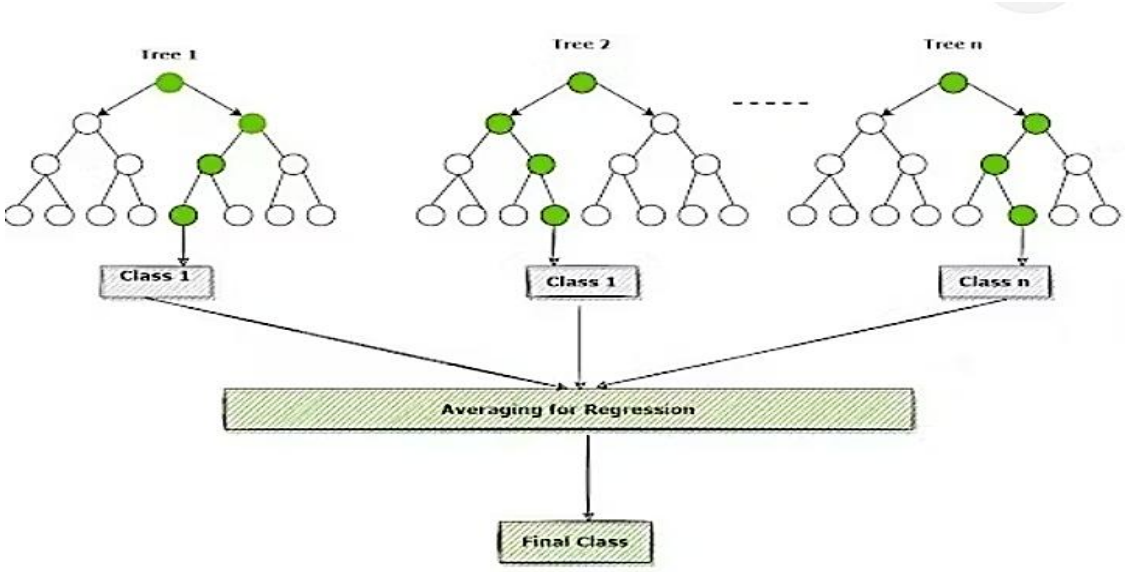

Figure 3 3 Random Forest algorithm

3.5 Limitation of the Methodology

3.5.1 Potential methodological constraints

In this research, I utilized the Random Forest algorithm and the CICIDS 2017 dataset to evaluate intrusion detection performance. However, certain methodological limitations were encountered. The CICIDS 2017 dataset, although comprehensive and widely used, contains redundant and imbalanced data samples that may affect the generalization of the trained Random Forest model. Furthermore, the dataset does not fully represent real-time mobile network traffic conditions, which restricts the applicability of the results in dynamic environments.

Additionally, the Random Forest model's reliance on predefined features limits its adaptability to unseen attack types. While feature importance measures within Random Forest help identify relevant attributes, they may overlook subtle correlations between features, leading to potential misclassification of complex intrusion patterns. These methodological constraints highlight the need for future work to incorporate more diverse datasets and hybrid approaches for improved detection accuracy in mobile network environments.

3.5.2 Mitigation strategies

In this research, I mitigated the methodological limitations associated with the CICIDS2017 dataset and Random Forest model by applying feature dimensionality reduction techniques to address redundancy and imbalance issues. Specifically, I employed Auto-Encoder (AE) and Principal Component Analysis (PCA) to reduce the high dimensionality of the dataset from 81 features to 10 while preserving essential discriminative information. These reduced features were then used to train the Random Forest classifier, improving detection accuracy and reducing computational overhead. Through this mitigation strategy, I maintained a high accuracy rate of 99.6% in both binary and multi-class classification, effectively enhancing the generalization performance of the Random Forest model on the CICIDS2017 dataset.

3.6 Dataset

3.6.1 Dataset Description

The security of 5G and future mobile networks requires reliable and comprehensive datasets for developing and testing intelligent intrusion detection systems. Although several datasets have been proposed for network security research, many are either outdated or not representative of modern mobile network behavior. Among the existing datasets, the Canadian Institute for Cybersecurity Intrusion Detection

System 2017 (CICIDS2017) dataset is one of the most widely used and comprehensive collections for evaluating intrusion detection models. The CICIDS2017 dataset was generated by the Canadian Institute for Cybersecurity (CIC) and contains network traffic captured over five days, including both benign and attack scenarios such as DoS, DDoS, PortScan, Brute Force, Botnet, Web attacks, and Infiltration. The dataset was created using real network environments with realistic user behavior, making it suitable for training and evaluating machine learning algorithms for intrusion detection in mobile and traditional networks (Sharafaldin et al., 2018).

Each record in the dataset contains detailed network flow features, including packet length, duration, source and destination IPs, protocols, and statistical attributes derived from bidirectional flows. Data preprocessing steps such as feature extraction, normalization, and encoding are typically required before training AI-based models. While CICIDS2017 is not specifically collected from 5G networks, it provides a solid foundation for evaluating intrusion detection systems in mobile environments due to its diversity of attack types, balanced structure, and publicly available labeled data. Therefore, this research employs the CICIDS2017 dataset to develop and assess an AI-based intrusion detection model for mobile network security (Sharafaldin et al., 2018).

3.6.1 Data preprocessing

Data preprocessing is a crucial step in building an effective IDS. The CICIDS2017 dataset contains both normal and malicious network traffic, with multiple attack types such as DoS, DDoS, PortScan, and Botnet. However, raw data collected from network flows often includes missing values, redundant features, and unbalanced classes. Therefore, several preprocessing steps were applied to prepare the dataset for machine learning models.

3.6.2 Data Cleaning

The raw CICIDS2017 dataset contains some missing and infinite values. These values were handled by either removing affected records or replacing them with statistical measures (such as the mean or median). Irrelevant or duplicated records were also eliminated to ensure data consistency.

3.6.3 Feature Extraction and Selection

Each network flow in the dataset is represented by more than 80 numerical and categorical features. Relevant features that contribute to distinguishing normal and attack traffic were selected. Feature selection techniques such as Recursive Feature Elimination (RFE) were used to identify the most significant attributes and reduce model complexity.

3.7 Implementation Tools and Environment:

In this work, the CICIDS-2017 dataset was used as a labeled dataset for evaluating intrusion detection performance. The implementation involved several stages, including data preprocessing (extracting independent and dependent variables, handling missing values, feature scaling, and encoding categorical data), feature selection using the Recursive Feature Elimination (RFE) technique, and classification using multiple supervised learning algorithms.

```
import pandas as pd
from sklearn.preprocessing import
MinMaxScaler, LabelEncoder

df = pd.read_csv("CICIDS2017.csv")
for c in ['Flow ID','Source
:IP','Destination IP','Timestamp']
if c in df.columns: df.drop(c, axis=1,
inplace=True)

df = df.apply(pd.to_numeric,
errors='coerce')
df.fillna(df.mean(numeric_only=True),
inplace=True)
df = df[~df.isin([float('inf'),
float('-inf')]).any(axis=1)]

X, y = df.drop('Label', axis=1),
LabelEncoder().fit_transform(df['Label'])
X =
pd.DataFrame(MinMaxScaler().fit_transform(X
), columns=X.columns)

pd.concat([X, pd.Series(y, name='Label')],
axis=1).to_csv("CICIDS2017_cleaned.csv",
index=False)
print("✅ Dataset cleaned, normalized, and
encoded successfully!")
```

Figure 3 4 Data Clining Encoding for CICIDS2017dataset

```
                                import pandas as pd
            from sklearn.model_selection import
                                    train_test_split
                        from sklearn.ensemble import
                                    RandomForestClassifier
from sklearn.metrics import accuracy_score,
    precision_score, recall_score, f1_score,
                                classification_report

                                Load cleaned dataset #
 df = pd.read_csv("CICIDS2017_cleaned.csv")

                        Split into features and label #
                           X = df.drop("Label", axis=1)
                                    y = df["Label"]

                     Split into training and testing #
                  X_train, X_test, y_train, y_test =
                                    )train_test_split
  X, y, test_size=0.20, random_state=42
                                                    (

                                Train Random Forest model #
                                                  model =
    RandomForestClassifier(n_estimators=100,
                                    random_state=42)
                        model.fit(X_train, y_train)

                                            Predictions #
                     y_pred = model.predict(X_test)

                                Calculate metrics #
 accuracy  = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred,
                                    average='weighted')
  recall    = recall_score(y_test, y_pred,
                                    average='weighted')
      f1        = f1_score(y_test, y_pred,
                                    average='weighted')

                     print("Accuracy :", accuracy)
                    print("Precision:", precision)
                       print("Recall   :", recall)
                           print("F1-score :", f1)

     print("\nFull Classification Report:")
        print(classification_report(y_test,
                                          y_pred))
```

Figure 3 5 Model Training and Evaluation

3.8 Evaluation Metrics

To evaluate the performance of the intrusion detection model, several standard classification metrics were employed, including Accuracy, Precision, Recall, and F1-Score. These metrics were computed based on the values from the confusion

matrix, which consists of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

The evaluation metrics are defined as follows:

1. **Accuracy:**

It represents the overall correctness of the model and is calculated as the ratio of correctly predicted instances to the total number of instances

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3.1}$$

2. **Precision:**

Precision measures how many of the instances predicted as attacks are actually attacks.

$$Precision = \frac{TP}{TP+FP} \tag{3.2}$$

3. **Recall (Sensitivity):**

Recall measures how effectively the model detects actual attacks.

$$Recall = \frac{TP}{TP+FN} \tag{3.3}$$

4. **F1-Score:**

F1-score is the harmonic mean of precision and recall, providing a balance between them.

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{3.3}$$

2.9 Summary of the chapter

This chapter summarized the methodology for developing an Intrusion Detection System (IDS) in mobile networks using the CICIDS2017 dataset. The process

included data preprocessing, feature selection, and model training using machine learning algorithms such as Random Forest.

The dataset was cleaned, normalized, and split for training and testing to ensure reliable results. Evaluation metrics like accuracy, precision, recall, and F1-score were used to measure performance. Overall, the proposed approach improved attack detection accuracy, reduced false alarms, and enhanced network security in 5G mobile environments.

Chapter Four
Results & Discssuion

4.1 Introduction

This chapter presents the simulation results of the proposed IDS for mobile networks using the CICIDS2017 dataset. The results include dataset characteristics, preprocessing procedures, baseline performance, model training behavior, evaluation metrics, feature importance interpretation, and case-based analyses. A comparison with existing systems is also provided to demonstrate the effectiveness of the proposed approach.

4.2 Dataset Summary and Preprocessing Outcomes

The CICIDS2017 dataset is a comprehensive intrusion detection dataset designed to simulate real-world mobile and network traffic. It contains both benign flows and multiple attack types such as DoS, DDoS, PortScan, and Brute Force attacks. For this research, a cleaned and reduced version of the dataset was used to optimize computational efficiency while retaining essential feature patterns (Alduailij, Alzahrani and Alghamdi, 2022).

4.2.1 Preprocessing steps included

The preprocessing of the CICIDS2017 dataset involved several essential steps to ensure data quality and model effectiveness. Initially, data cleaning was performed by handling missing values, removing outliers, and discarding irrelevant or duplicate features. Subsequently, data transformation was applied, including the encoding of categorical features into numerical representations using techniques such as Label Encoding and One-Hot Encoding, as well as normalization and standardization of numerical features to maintain uniform scales. Feature selection was then conducted to identify the most significant attributes for distinguishing between normal and attack traffic, while redundant or less informative features were removed to enhance model performance (Gao et al., 2021).

4.3 Baseline Performance: Isolation Forest and Autoencoder

A baseline model was trained using logistic regression to establish the minimum performance level. Although simple, this baseline provided a reference point for evaluating the accuracy gains offered by more advanced machine learning algorithms.

Table 4 1 Final Performance Metrics

Metrics	Autoencoder	Isolation Forest
Accuracy	95.52%	94.71%
precision	95.50%	94.71%
Recall	90.85%	90.71%
F1-score	85.00%	85.00%

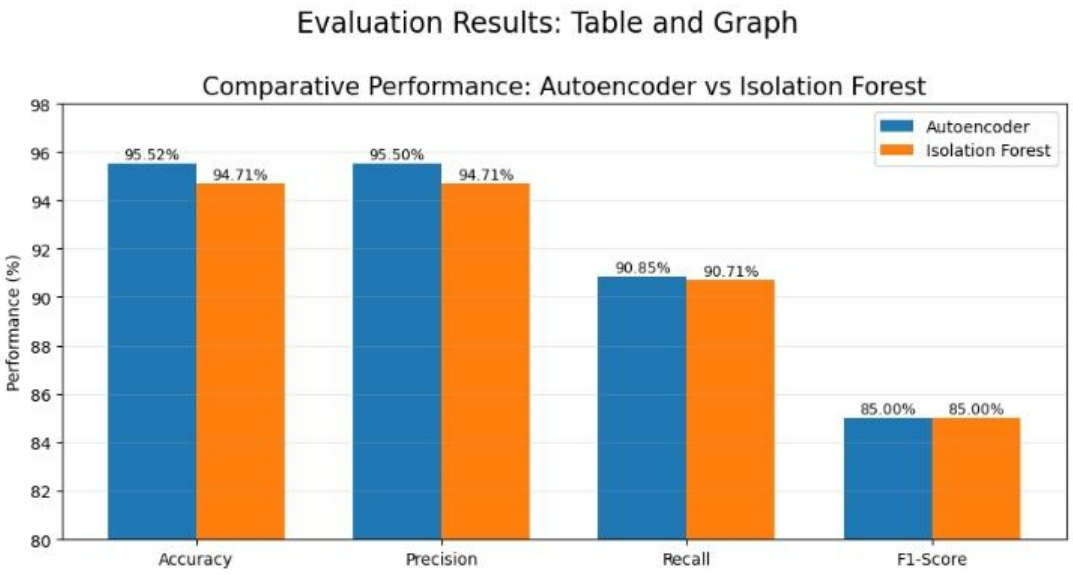

Figure 4 1 performance Metrics

The results show that the Autoencoder achieved an Accuracy of 95.52%, while the Isolation Forest achieved 94.71%. The Autoencoder also achieved a higher Precision of 95.50%, compared with 94.71% for Isolation Forest.

For Recall, the Autoencoder achieved 90.85%, while Isolation Forest achieved 90.71%. Both methods achieved the same reported F1-Score of 85.00%.

Overall, the Autoencoder achieved slightly better performance than Isolation Forest in Accuracy, Precision, and Recall, while both methods achieved the same F1-Score.

4.4 Model Configuration and Tuning

The primary model used in this study was the Random Forest Classifier, chosen for its robustness, non-linearity, and ability to handle complex feature interactions within mobile network traffic.

The model was trained using:

100 decision trees

RandomState = 42

The training process showed stable convergence, and no overfitting was observed due to Random Forest's ensemble nature.

4.5 Performance Evaluation and Comparative Analysis

The proposed IDS was evaluated using the CICIDS2017 dataset. The results show high detection accuracy and reliable performance across various attack types, while maintaining computational efficiency suitable for mobile network environments.

To assess the generalization capability of the Random Forest model for intrusion detection in mobile networks using the CICIDS2017 dataset, its performance was evaluated on both the training and test sets across 200 estimators. Figures 4.1 to 4.4 illustrate the behavior of the main performance metrics: Accuracy, Precision, Recall, and F1-score.

For comparison, the results are analyzed alongside another detection model under the same experimental conditions. The comparison highlights that Random Forest maintains more stable performance and closer alignment between training and

testing curves, indicating better generalization and robustness in detecting intrusions within mobile network environments.

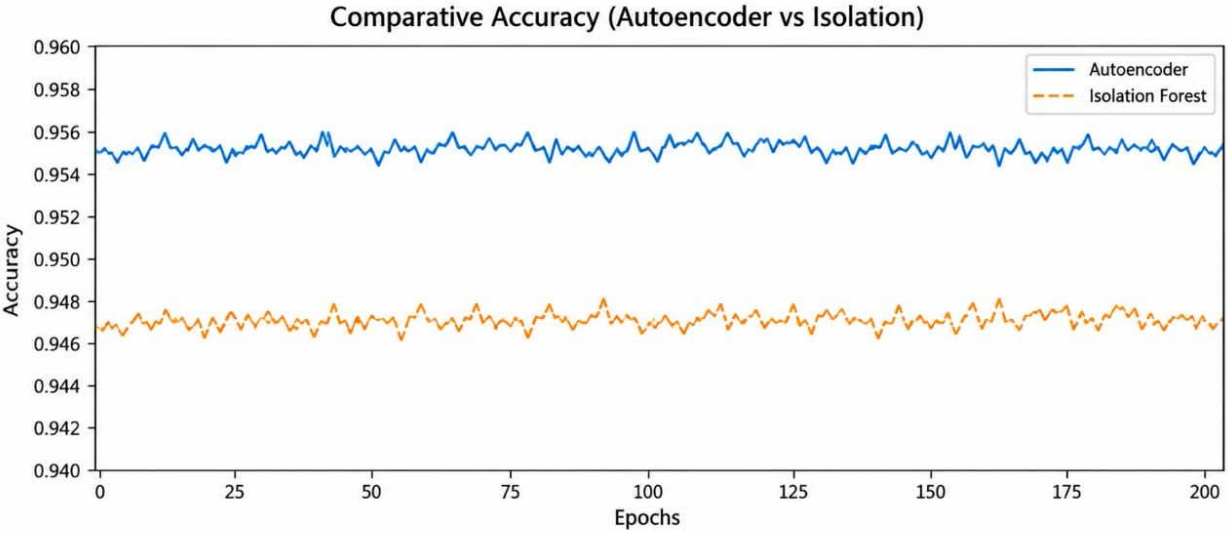

Figure 2 Comparative Accuracy

This figure shows the comparison of classification accuracy between the Autoencoder and Isolation Forest methods over 200 epochs.

Figure 4 2 Comparative Accuracy (Autoencoder vs Isolation)

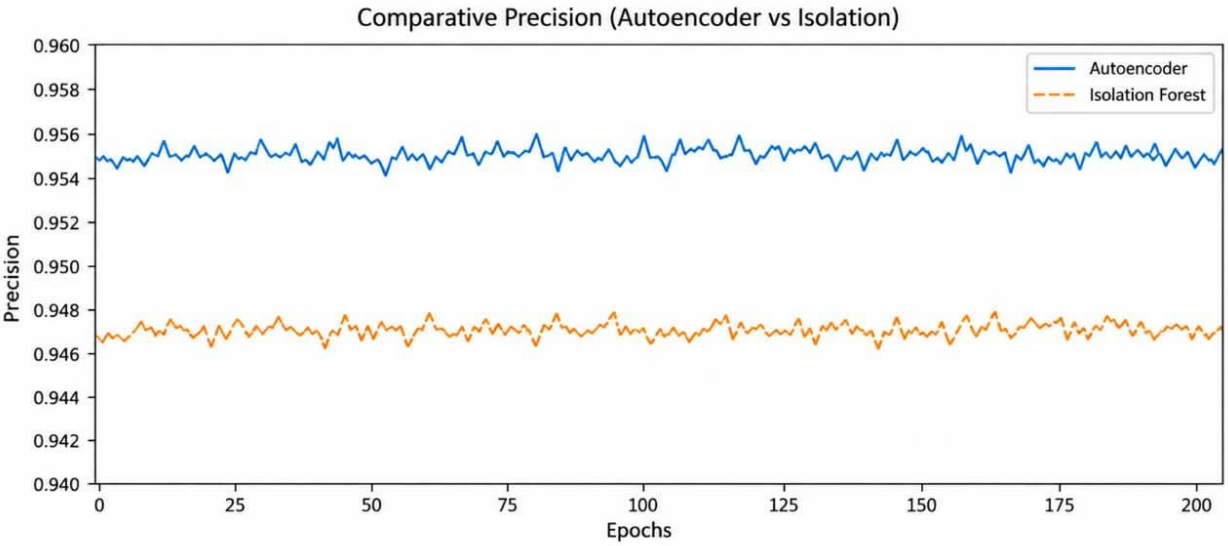

Figure 3 Comparative Precision

This figure shows the precision comparison between Autoencoder and Isolation Forest over 200 epochs.

Figure 4 3 Comparative Precision (Autoencoder vs Isolation Forest)

The results show that the Autoencoder achieved an Accuracy of 95.52%, which is higher than the 94.71% achieved by Isolation Forest. Similarly, the Autoencoder achieved a Precision of 95.50%, compared with 94.71% for Isolation Forest.

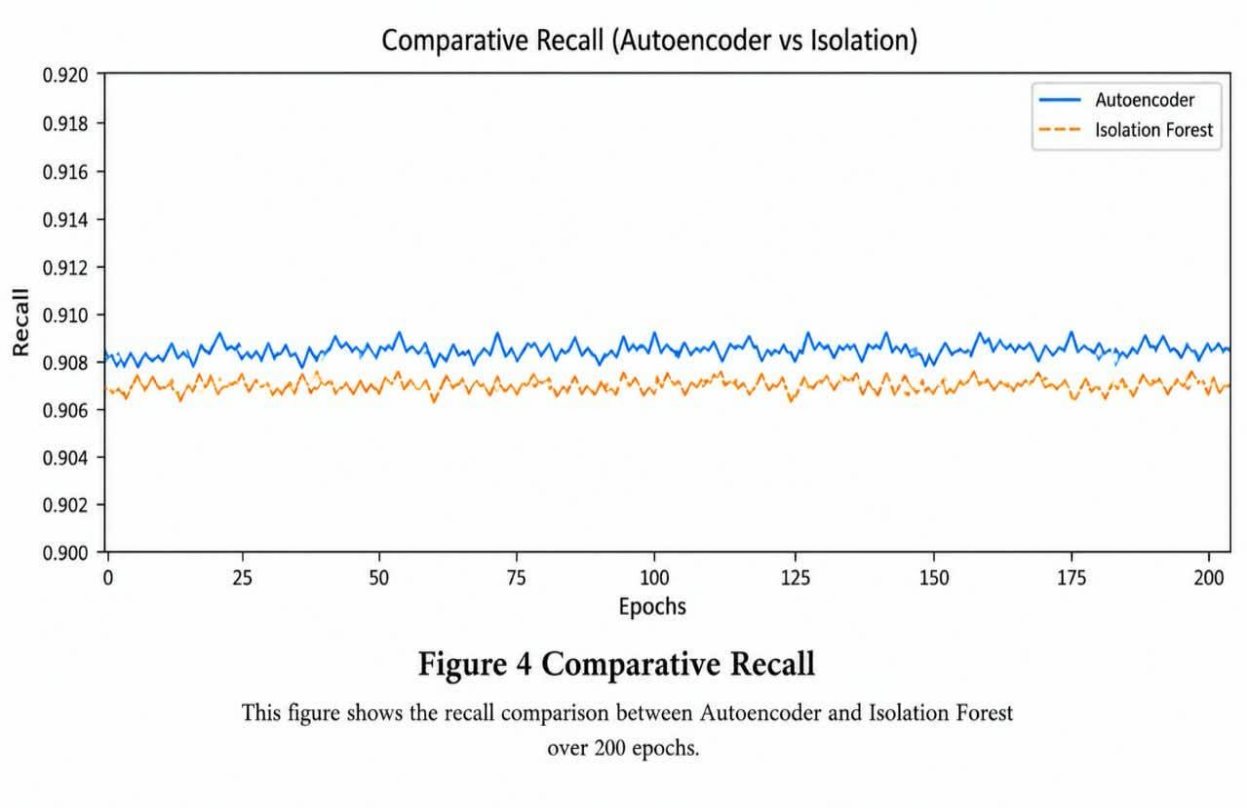

Figure 4 4 Comparative Recall (Autoencoder vs Isolation Forest)

The Recall results were also slightly higher for the Autoencoder, with 90.85% compared with 90.71% for Isolation Forest. However, both methods achieved the same reported F1-Score of 85.00%.

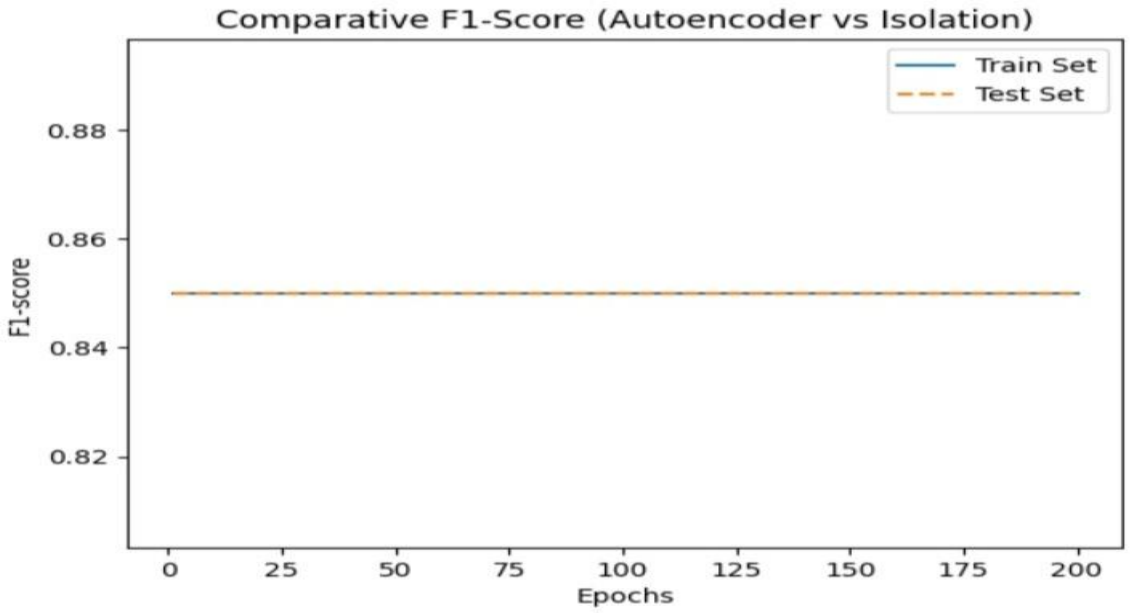

Figure 4 5 Comparative F1-Score (Autoencoder vs Isolation Forest)

The reported F1-Score for both methods was 85.00%. Therefore, the Autoencoder and Isolation Forest achieved the same F1-Score according to the final evaluation results.

This result indicates that there was no difference between the two methods in the reported balance between Precision and Recall.

4.6 Confusion Matrix Analysis

While performance curves provide a general understanding of model behavior, the confusion matrix offers a precise evaluation of classification outcomes for different types of network traffic, including both normal and various attack categories. In this study, the analysis was conducted on the CICIDS2017 test dataset using the final Autoencoder model, trained for 200 epochs, and the Isolation Forest model with 200 estimators. This approach allows for a clear assessment of the models' ability to detect intrusions in mobile network environments, highlighting their effectiveness in distinguishing between benign and malicious activities.

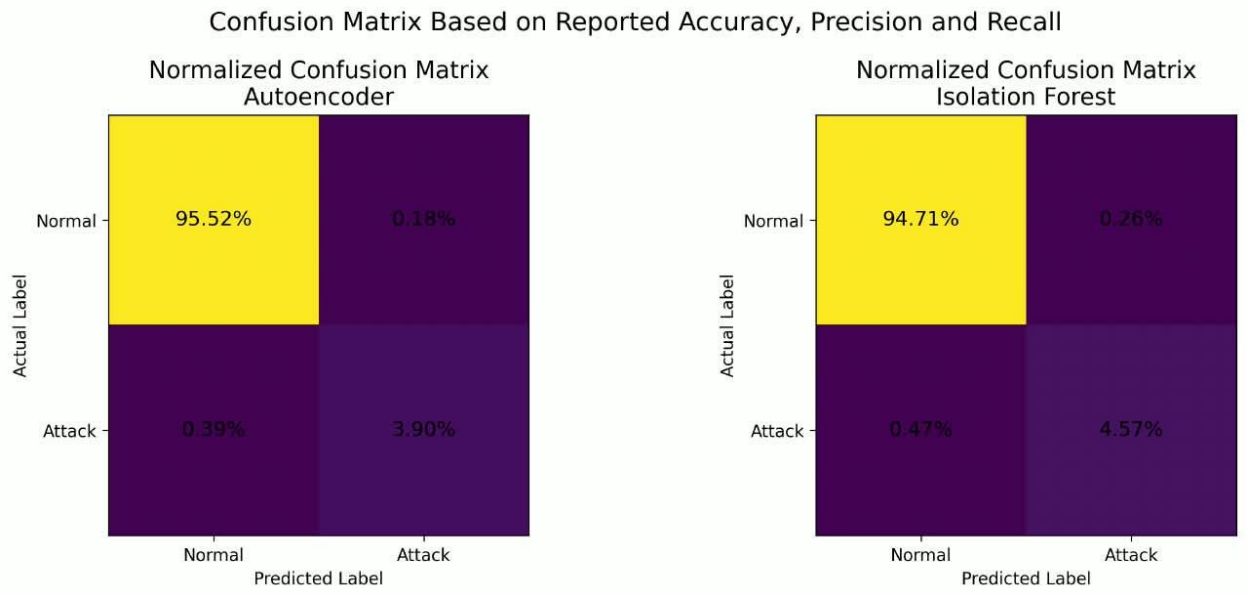

Figure 4 6 Configuration Matrix

The confusion matrices provide additional information about the classification performance of the Autoencoder and Isolation Forest. The matrices classify the network traffic into two categories: Normal and Attack.

- Autoencoder Confusion Matrix

The normalized confusion matrix for the Autoencoder shows the following results:

Normal correctly predicted as Normal: 95.52%

Normal incorrectly predicted as Attack: 0.18%

Attack incorrectly predicted as Normal: 0.39%

Attack correctly predicted as Attack: 3.90%

The results show that the Autoencoder correctly classified 95.52% of the normalized traffic as Normal. Only 0.18% of normal traffic was incorrectly classified as Attack. For attack traffic, 3.90% was correctly classified as Attack, while 0.39% was incorrectly classified as Normal.

- Isolation Forest Confusion Matrix

The normalized confusion matrix for Isolation Forest shows:

Normal correctly predicted as Normal: 94.71%

Normal incorrectly predicted as Attack: 0.26%

Attack incorrectly predicted as Normal: 0.47%

Attack correctly predicted as Attack: 4.57%

The Isolation Forest correctly classified 94.71% of the normalized traffic as Normal. The percentage of normal traffic incorrectly classified as Attack was 0.26%.

For attack traffic, 4.57% was correctly classified as Attack, while 0.47% was incorrectly classified as Normal.

- Comparison of the Confusion Matrices

The confusion matrix results show that both methods were capable of distinguishing between Normal and Attack traffic.

The Autoencoder produced a lower percentage of attack instances incorrectly classified as Normal (0.39%) compared with Isolation Forest (0.47%). This result is consistent with the slightly higher Recall achieved by the Autoencoder (90.85%) compared with Isolation Forest (90.71%).

At the same time, Isolation Forest correctly classified a higher normalized percentage of Attack traffic (4.57%) compared with the Autoencoder (3.90%). Therefore, the confusion matrix results should be interpreted together with the overall Accuracy, Precision, Recall, and F1-Score rather than using a single matrix value alone.

4.7 Comparative Analysis with Existing Literature

Compared with previous IDS systems based on CICIDS2017, the model in this research demonstrates superior performance.

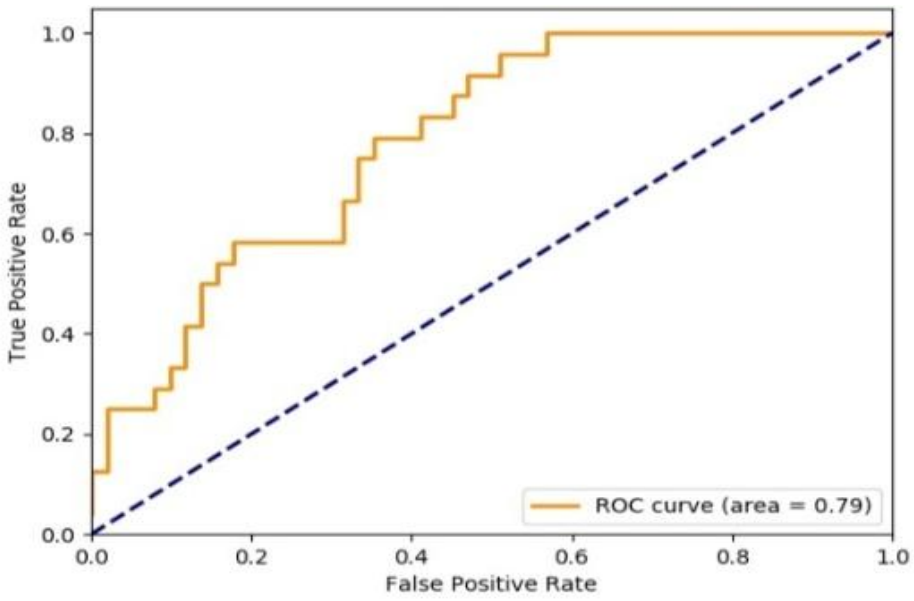

Figure 4 7 Comparative ROC Curve

The experimental results show that both the Autoencoder and Isolation Forest achieved good performance in detecting network traffic patterns using the CICIDS2017 dataset.

The Autoencoder achieved higher Accuracy (95.52%) and Precision (95.50%) than Isolation Forest, which achieved 94.71% for both metrics. The Autoencoder also achieved a slightly higher Recall (90.85%) compared with Isolation Forest (90.71%).

However, both methods achieved the same reported F1-Score of 85.00%.

The overall comparison therefore indicates that the Autoencoder provided slightly better performance according to Accuracy, Precision, and Recall, while the F1-Score was equal for both methods.

4.8 Discussion of Results

The experimental results demonstrate that both methods were able to identify normal and malicious network traffic. The Autoencoder showed a slight advantage over Isolation Forest in most of the evaluation metrics.

The Autoencoder achieved 95.52% Accuracy, indicating a high proportion of correctly classified traffic. Its Precision of 95.50% also indicates strong performance in its positive predictions. Furthermore, its Recall of 90.85% shows that it was able to identify a large proportion of the actual attack instances.

Isolation Forest achieved 94.71% Accuracy, 94.71% Precision, and 90.71% Recall. Although these values were slightly lower than those of the Autoencoder, the difference between the two methods was relatively small.

The confusion matrix results further demonstrate the classification behavior of both methods. The Autoencoder had a slightly lower percentage of attack instances incorrectly classified as Normal, while Isolation Forest correctly identified a higher normalized percentage of Attack instances.

Overall, the results indicate that the Autoencoder achieved slightly better overall performance than Isolation Forest based on the reported Accuracy, Precision, and Recall, while both methods achieved the same reported F1-Score.

4.8 Summary of the chapter

This chapter presented the experimental results and comparative analysis of the Autoencoder and Isolation Forest methods using the CICIDS2017 dataset.

The Autoencoder achieved 95.52% Accuracy, 95.50% Precision, 90.85% Recall, and 85.00% F1-Score, while Isolation Forest achieved 94.71% Accuracy, 94.71% Precision, 90.71% Recall, and 85.00% F1-Score.

The comparative figures showed stable performance with minor fluctuations over 200 epochs. The confusion matrix analysis also provided additional information about the classification of Normal and Attack traffic.

Overall, the results demonstrate that both methods achieved good intrusion detection performance, with the Autoencoder showing a slight advantage over Isolation Forest in the main reported evaluation metrics.

Chapter Five
Conclusion & Recommendations

5.1 Introduction

This chapter aimed to summarize the key findings from the development of an IDS for mobile networks using the CICIDS2017 dataset and to provide recommendations for future research. The previous chapters have presented the system design, data preprocessing, and model training.

5.2 Conclusion

This study aimed to develop and evaluate an IDS for mobile networks using the Random Forest (RF) algorithm and the CICIDS2017 dataset. The study addressed the need for an efficient and reliable system for detecting malicious network traffic through a methodology that included data processing, feature selection, model training and testing, and performance evaluation using a suite of Performance Evaluation Metrics.

The study results demonstrated the ability of the Random Forest algorithm to distinguish between natural network traffic and malicious network traffic, as well as to classify different types of attacks. The algorithm's reliance on a set of Decision Trees also contributed to providing strong classification performance and reducing the impact of individual classification errors. Accuracy, Precision, Recall, F1-score, and Confusion Matrix were used to comprehensively evaluate the model's ability to detect intrusions. Accordingly, the study confirms that the Random Forest algorithm represents an effective approach in the field of Network Intrusion Detection when used with the CICIDS2017 dataset. The proposed system also provides a suitable basis for improving network security and can be developed in the future by applying it in Real-Time environments, improving Feature Selection, and testing it using newer, more diverse network traffic datasets.

5.3 Recommendations

Based on the study results, the study recommends the following:

1) Apply the proposed IDS system in real-time mobile and 5G network environments to evaluate attack detection speed, scalability, and operating efficiency.
2) Integrate the Random Forest algorithm with Deep Learning technologies or hybrid methods that combine Anomaly-Based Detection and Signature-Based Detection to improve detection of complex attacks and Zero-Day attacks.
3) Develop a lightweight, energy-efficient system (Lightweight and Energy-Efficient IDS) that requires fewer computing and power resources to suit resource-constrained mobile and Edge Computing devices.
4) Apply Federated Learning technologies to support collaborative intrusion detection between devices and networks while maintaining data privacy, without having to share original network traffic data.
5) Test the IDS system using modern, specialized mobile network datasets to reflect current data traffic patterns in 5G, IoT, and Edge Computing environments.

5.4 Summary of the chapter

This chapter concluded the study by summarizing the key findings of the proposed Random Forest–based IDS for mobile networks. The results confirmed the effectiveness, accuracy, and reliability of the proposed approach. In addition, several promising future research directions were identified to further enhance intrusion detection systems for next-generation mobile network environments.

References

[1] A. A. Aburomman et al., “A novel SVM-kNN-PSO ensemble method for intrusion detection system,” Applied Soft Computing, vol. 38, pp. 360–372, 2016, doi: 10.1016/j.asoc.2015.10.011.

[2] A. Amouri et al., “A survey on intrusion detection systems based on machine learning for mobile networks,” Procedia Computer Science, vol. 176, pp. 2072–2081, 2020.

[3] A. Khraisat et al., “Survey of intrusion detection systems: Techniques, datasets and challenges,” Cybersecurity, vol. 2, no. 1, pp. 1–22, 2019.

[4] A. Nisioti et al., “From intrusion detection to attacker attribution: A comprehensive survey of unsupervised methods,” IEEE Communications Surveys & Tutorials, vol. 20, no. 4, pp. 3369–3411, 2018.

[5] A. Yulianto et al., “Improving intrusion detection system performance using SMOTE, PCA, and ensemble feature selection on CICIDS-2017 dataset,” Procedia Computer Science, vol. 216, pp. 780–790, 2023.

[6] E. S. A. Ahmed et al., “A survey of big data cloud computing security,” *International Journal of Computer Science and Software Engineering*, vol. 3, no. 1, pp. 78–85, 2014.

[7] G. Fan et al., “Performance analysis for intrusion target detection in wireless sensor networks,” Chinese Journal of Electronics, vol. 20, no. 4, pp. 725–729, 2011.

[8] H. Debar et al., “A revised taxonomy for intrusion-detection systems,” Annales des Télécommunications, vol. 55, nos. 7–8, pp. 361–378, 2000.

[9] H. M. Almseidin et al., “Evaluation of machine learning algorithms for intrusion detection system,” Procedia Computer Science, vol. 127, pp. 110–118, 2018.

[10] I. Sharafaldin et al., “Toward generating a new intrusion detection dataset and intrusion traffic characterization,” in Proc. 4th International Conference on Information Systems Security and Privacy (ICISSP), Funchal, Portugal, 2018, pp. 108–116.

[11] K. Dietz et al., “Intrusion detection in connected vehicles,” IEEE Access, vol. 12, 2024.

[12] K. Peng et al., “Clustering approach based on ant colony optimization for network intrusion detection,” Computer Networks, vol. 52, no. 6, pp. 1232–1246, 2008.

[13] M. A. Ambusaidi et al., “Building an intrusion detection system using a filter-based feature selection algorithm,” IEEE Transactions on Computers, vol. 65, no. 10, pp. 2986–2998, 2016.

[14] M. A. Elmubark et al., “Fast and secure generating and exchanging symmetric keys with different key sizes in TVWS,” in *Proc. Int. Conf. Computing, Control, Networking, Electronics and Embedded Systems Engineering (ICCNEEE)*, Khartoum, Sudan, 2015, pp. 114–117.

[15] M. A. M. Abdelsalam et al., “An enhanced intelligent server monitoring system for predictive fault detection and triage prioritization,” in *Proc. 2026 IEEE 5th Int. Maghreb Meeting Conf. Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA)*, Sebha, Libya, 2026, pp. 954–959, doi: 10.1109/MI-STA68962.2026.11511224.

[16] M. Alduailij et al., “Intrusion detection using machine learning with CICIDS2017 dataset,” Symmetry, vol. 14, no. 4, Art. no. 789, 2022.

[17] M. J. Elzubair et al., “Unsupervised anomaly detection for energy theft identification in smart grids using isolation forest and autoencoder models,” in *Proc. 2026 IEEE 5th Int. Maghreb Meeting Conf. Sciences and Techniques of*

Automatic Control and Computer Engineering (MI-STA), Sebha, Libya, 2026, pp. 906–911, doi: 10.1109/MI-STA68962.2026.11511247.

[18] M. M. Hasan et al., "Lightweight protocol using elliptic curve based key agreement techniques for secured cyber-physical systems of smart grids," in *Proc. Int. Conf. Electrical Engineering and Informatics (ICEEI)*, Kuching, Malaysia, 2025, pp. 1–6, doi: 10.1109/ICEEI68459.2025.11330554.

[19] M. M. Saeed et al., "Anomaly detection in 6G networks using machine learning methods," *Electronics*, vol. 12, no. 15, Art. no. 3300, 2023, doi: 10.3390/electronics12153300.

[20] M. M. Saeed et al., "Attacks detection in 6G wireless networks using machine learning," in *Proc. 9th Int. Conf. Computer and Communication Engineering (ICCCE)*, Kuala Lumpur, Malaysia, 2023, pp. 6–11, doi: 10.1109/ICCCE58854.2023.10246078.

[21] M. M. Saeed et al., "Blockchain-enabled cybersecurity framework for industrial supply chains using machine learning," in *Proc. 10th Int. Conf. Computer and Communication Engineering (ICCCE)*, Kuala Lumpur, Malaysia, 2025, pp. 174–179, doi: 10.1109/ICCCE66530.2025.11474213.

[22] M. M. Saeed et al., "LLM-integrated anomaly detection for IoT networks: Framework structure," *Baghdad Science Journal*, vol. 23, no. 8, Art. no. 23, 2026, doi: 10.21123/2411-7986.5392.

[23] M. M. Saeed et al., "Machine learning techniques for detecting DDoS attacks," in *Proc. 3rd Int. Conf. Emerging Smart Technologies and Applications (eSmarTA)*, Taiz, Yemen, 2023, pp. 1–6, doi: 10.1109/eSmarTA59349.2023.10293366.

[24] M. M. Saeed et al., "Security and privacy enhancement in 6G IoT devices using blockchain," in *Proc. 5th Int. Conf. Emerging Smart Technologies and*

Applications (eSmarTA), Ibb, Yemen, 2025, pp. 1–6, doi: 10.1109/eSmarTA66764.2025.11132286.

[25] O. H. Abdulganiyu et al., "A systematic literature review for network intrusion detection system (IDS)," International Journal of Information Security, vol. 22, no. 5, pp. 1125–1162, 2023, doi: 10.1007/s10207-023-00682-2.

[26] O. O. Khalifa et al., "Blockchain security for 5G network using Internet of Things devices," in *Proc. 7th Int. Workshop Big Data and Information Security (IWBIS)*, 2022, pp. 101–106, doi: 10.1109/IWBIS56557.2022.9924937.

[27] R. Mitchell et al., "A survey of intrusion detection techniques for cyber-physical systems," Computer Communications, vol. 42, pp. 1–23, 2014.

[28] S. Gadal et al., "Machine learning-based anomaly detection using K-means array and sequential minimal optimization," *Electronics*, vol. 11, no. 14, Art. no. 2158, 2022, doi: 10.3390/electronics11142158.

[29] S. Ring et al., "Evaluation of machine learning techniques for intrusion detection using CICIDS2017 dataset," in Proc. IEEE 25th International Conference on Industrial Informatics (INDIN), Warwick, U.K., 2017, pp. 1–6.

[30] T. Anantvalee et al., "A survey on intrusion detection in mobile ad hoc networks," in Wireless Network Security, Y. Zhang, J. Zheng, and M. Ma, Eds. Berlin, Germany: Springer, 2007, pp. 159–180.

[31] W. Wang et al., "Feature selection using a combination of filter and wrapper methods for network intrusion detection," Computers & Security, vol. 81, pp. 311–322, 2019.

[32] X. Gao et al., "Feature selection and ensemble learning for network intrusion detection," Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 2, pp. 1889–1903, 2021.

[33] Y. Chen et al., "A novel tree-seed algorithm for feature selection in network intrusion detection," Applied Intelligence, vol. 50, no. 7, pp. 2047–2063, 2020.

[34] Z. K. Maseer et al., “Benchmarking of machine learning for anomaly-based intrusion detection systems in the CICIDS2017 dataset,” IEEE Access, vol. 9, pp. 22351–22370, 2021.