MACHINE LEARNING APPROACHES FOR DETECTING AND MITIGATING CYBERSECURITY THREATS

BY

Md Mosaddek Islam Zeem

ID: 201-15-13895

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Dr. Touhid Bhuiyan (DTB)
Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Mr. Narayan Ranjan Chakraborty (NRC)
Associate Professor and Associate Head
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JULY 2024

APPROVAL

This Project titled "Machine Learning Approaches for Detecting and Mitigating Cybersecurity Threats", submitted by **Md Mosaddek Islam Zeem** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 14-07-2024.

BOARD OF EXAMINERS

Dr. S.M Aminul Haque
Professor & Associate Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

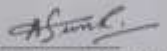
Board Chairman

Md. Abbas Ali Khan
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner 1

Mr. Md. Aynul Hasan Nahid
Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner 2



Dr. Abu Sayed Md. Mostafizur Rahaman
Professor
Department of CSE
Jahangirnagar University

External Examiner

©Daffodil International University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dr. Touhid Bhuiyan, Professor, Department of Computer Science and Engineering, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Dr. Touhid Bhuiyan
Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Mr. Narayan Ranjan Chakraborty
Associate Professor and Associate Head
Department of CSE
Daffodil International University

Submitted by:

Md. Mosaddek Islam Zeem

Md Mosaddek Islam Zeem
ID: 201-15-13895
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, I express my heartiest thanks and gratefulness to almighty for His divine blessing making it possible for me to complete the final year project/internship successfully.

I am grateful and wish our profound indebtedness to **Dr. Touhid Bhuiyan**, Professor, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of my supervisor in the field of "*Cybersecurity*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

I would like to express my heartiest gratitude to the **Head of the Department of CSE**, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.
I would like to thank my entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, I must acknowledge with due respect the constant support and patience of my parents.

ABSTRACT

Machine learning is super important in cybersecurity. The main goal of using machine learning in cybersecurity is to improve how we detect malware. This makes the process doable, scalable, and way more effective than old-school methods that need people to step in. Dealing with machine learning challenges in cybersecurity needs some serious handling. Different methods like deep learning, support vector machines, and Bayesian classification have shown they work well in stopping cyber-attacks. It's crucial to find hidden insights from network data and use a data-driven machine learning model to stop these attacks before they happen. No research has been done to understand how vulnerable ML techniques are against security threats and how they can be defended against. We need researchers, scientists, and engineers to start focusing on cybersecurity with ML. This survey looks at the machine learning techniques used on cybersecurity data to secure systems. However, ML can be attacked during training and testing, causing problems with security. We also talk about current cybersecurity threats and how machine-learning techniques help fight them off. We point out the flaws in these fancy models and how attack tactics have changed over the years. We aim to see if these machine-learning techniques work against the growing problem of malware causing trouble online.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Overview

The swift advancement of technology for information in recent times has given rise to several security issues, including misuse by unauthorized individuals. [1], denial of service (DoS) [2], malware attacks [3], zero-day attacks [4], data leakage [5], social engineering or phishing [6] has grown dramatically throughout the past ten years. The malicious application count in 2010 was fewer than 50 million, according to the IT security community. The security industry found over 900 million harmful files in 2019, but the amount of them is continually rising, per AV-TEST figures [7]. Both companies and individuals may suffer large financial losses as a result of cybercrime and cyberattacks. For instance, the average cost of a data breach in the US is estimated to be $3.9 million, while the average loss worldwide is $8.19 million [8]. Cybercrime costs global businesses $400 billion annually [9].

The security community predicts that data leaks may nearly quadruple within the following five-year period [10]. Therefore, to reduce further losses, businesses need to develop and implement a cybersecurity strategy. In fact, health research shows that national security depends on the government and people's access to information, applications, and equipment that require high security [11]. Cybersecurity is the umbrella term covering methods and tools used to guard data, machines, networks, and programs against loss, intrusion, or assault. Cybersecurity falls into several categories and encompasses anything form smartphone apps to enterprise. They involve (i) network safety, which tries to stop hackers or other intruders from accessing computer networks; (ii) application security, which is dedicated to protecting hardware and software from dangers or network threats; (iii) data security, which typically considers the security and confidentiality of the pertinent data; as well as internet safety solutions, including firewalls, intrusion detectors, and antivirus programs in network and computer security systems. This shift is being driven by data science, and a key component of the term "artificial intelligence" is the use of machine learning. and machine learning has had a great impact in the field of cybersecurity [12]. As discussed

in the article, as technology related to cyber threats advances, attackers become more effective, which leads to more creation of technology. being crucial in identifying buried relationships within info. A freshly emerging field of study is being established by the study of data.

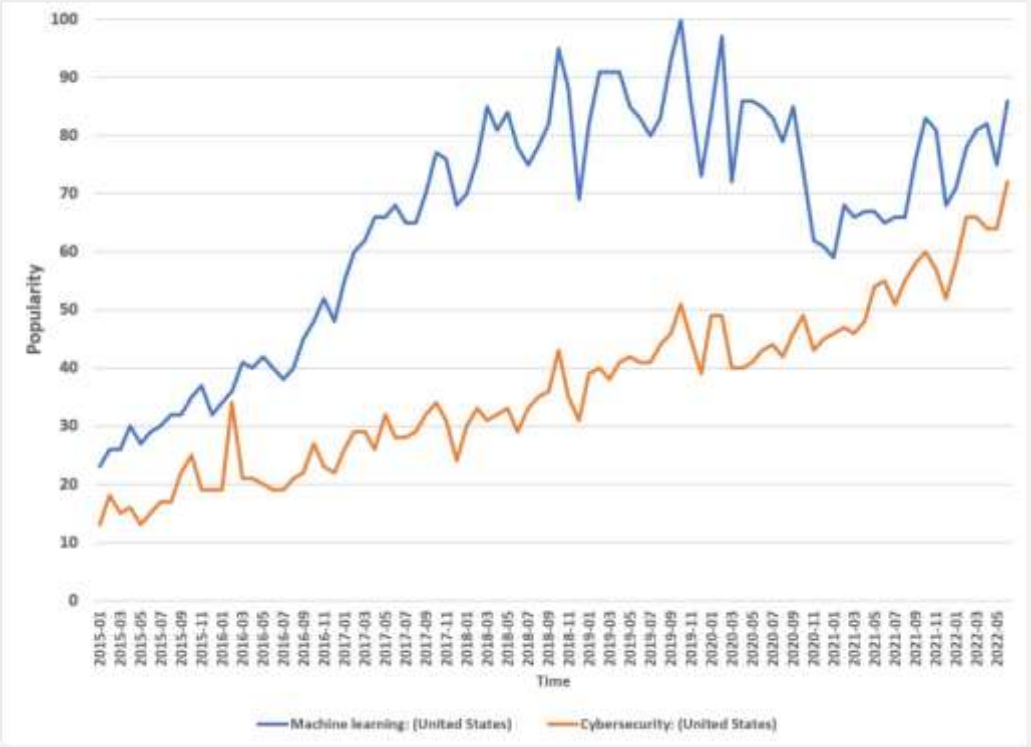


Fig.1.1.1: Cybersecurity vs Machine Learning Source: Google

The duration of the data for a certain day is displayed on the chart in Figure 1.1.1, with the contest's appeal (x-axis) representing its minimum while the popularity range (y-axis) representing its highest. Cybersecurity and machine learning were found to have a popularity rating of less than thirty in 2015, but by 2022, this figure is expected to reach over seventy, more than doubling in importance. Our study centers on machine learning in the field of cybersecurity, which has similarities to these endeavors in terms of data processing techniques employed in practical applications, security, and intelligent decision-making. In general, this study focuses on data security, utilizing machine learning algorithms to anticipate cyberthreats and enhance cybersecurity protocols. This work is also helpful to organizations and researchers that wish to explore and build data-driven intelligent network security models using machine learning [13]. Some unique but significant features of traditional security solutions,

such as firewalls, encryption systems, and user identification and access control, are included in machine learning but may not be sufficient to fulfill the demands of modern online businesses [14]. The primary issue is that specialists and security analysts manually address issues when data management is required on a regular basis [15]. Traditional methods haven't worked well to combat these cyberthreats, though, since cybersecurity defense options have grown over time. This has led to the emergence and propagation of several new hacks and assaults on the Internet. As a result, some researchers construct network security models using various data and information from pattern extraction. It relies in the safety of the evaluation agreed and is presented for the part titled "Machine Learning Technology in Network Security." and security trends could perhaps be more common. According to research, creating more straightforward and effective security systems that can respond to assaults and modify security policies to quickly and wisely stop them is necessary to solve network issues. This necessitates examining a large amount of cybersecurity data gathered from many sources, including system, the spot, etc. Furthermore, these innovations need to be automated and involve little to no human involvement.

1.2 Background and Present State

Since its construction more than ten years ago, our civilization's infrastructure for information and communication technology (ICT) has been growing and integrating. As a result, security developers are increasingly expected to defend ICT software and hardware from online threats [16]. Cybersecurity is the defense of computer networks against different cyberthreats or assaults. The many aspects of cybersecurity are interconnected; they include safeguards for ICTs, unprocessed data, and the knowledge they hold, as well as their transmissions and operations. The participation of both virtual and real bodies is another aspect.

Cybersecurity, according to, is the application of many techniques, policies, and technologies to safeguard data, networks of computers, and software against corruption, illegal access, and assaults. According to the research, network security employs a variety of techniques and tools to defend systems, networks, applications, and data from harm, illegal access, and assaults. To put it briefly, cybersecurity is recognizing the many types of cyberattacks that are listed below and utilizing the right defense tactics

to safeguard data [17]. Content exchanged via unapproved websites, users, or devices. Regional systems as well as property access.

1.3 Problem Statement

Malware is malicious software that is intended to harm a user's computer system, customer, server that is or system. By creating destructive occurrences, such as people clicking on dubious links or email connections, malware disrupts networks and raises software risks. Any gadget with logic can become a victim of malware. Users, servers, their network devices, and control systems like data gathering and monitoring can all be victims. Malware typically has the following effects on the network:

• It disables essential network parts.

• It runs more harmful programs in order to snoop via malware.

• It sends information and obtains permission to view personal data.

• It interferes with other parts, rendering the system unusable for users.

1.4 Objectives

The study will thoroughly examine the most recent advances in machine learning, including deep learning. All of these areas have a great deal of potential to transform cybersecurity procedures. This investigation will encompass a comprehensive examination of recent developments and innovations, including the incorporation of artificial intelligence into cybersecurity frameworks, the application huge analysis of data in order to threat identification, and the implementation of automated threat intelligence systems capable of cyber threat prediction and counteraction at a never-before-seen speed and precision.

The study intends to identify and examine the dangers and obstacles associated with the adoption of machine learning in cybersecurity, in addition to highlighting technical breakthroughs. These challenges cover a wide range of topics, such as the possibility of adversarial attacks that take advantage of flaws in machine learning models, the moral implications of automated decision-making, and the necessity of strong data

privacy and security safeguards to preserve sensitive data. This report seeks to provide an objective evaluation of the implementation of machine learning technology in cybersecurity by closely assessing these dangers.

In addition, the study seeks to provide researchers and companies with useful advice and practical knowledge to assist them in navigating the challenges of incorporating machine learning into cybersecurity efforts. The primary objective of these recommendations is to optimize the advantages of machine learning, including improved threat detection precision, shortened reaction times, and the capacity to identify and eliminate new threats. Concurrently, the suggestions will tackle methods to alleviate the related drawbacks, such as enhancing the comprehensibility of machine learning models, guaranteeing the moral application of AI in cybersecurity, and instituting all-encompassing defenses against hostile assaults.

In the end, our research attempts to make the cyber environment more robust and safe by providing companies with creative solutions that can keep up with the ever changing cyber threat scenario. Through this study, we hope to advance our understanding of how machine learning can be used in cybersecurity to safeguard private information, vital infrastructure, and digital assets in a linked world. We also hope to pave the way for future developments that will improve cybersecurity defenses and encourage a proactive approach to cyber defense.

1.5 Scope and Limitations

1.5.1 Project Scope

The study paper was produced in order to thoroughly examine the application of machine learning techniques in cybersecurity. This study's main objective is to present a thorough analysis of the body of research, with an emphasis on how well various machine learning algorithms identify and mitigate cyberthreats such as spyware, malware, and network attacks. A critical evaluation of the current current state of the science in guided, looking at how well they can improve threat detection and response mechanisms' precision and effectiveness. The research will also examine how big data analytics and artificial intelligence may be incorporated into cybersecurity frameworks, how this could support proactive threat mitigation tactics.

Additionally, the study will tackle important obstacles and constraints related to applying machine learning in cybersecurity, such as problems with model interpretability, vulnerability to adversarial attacks, and ethical dilemmas in automated decision-making processes. The project intends to give empirical insights into the practical value and limitations of machine learning applications in real-world cybersecurity scenarios by utilizing case studies and experimental investigations. In the end, this research aims to further understanding and enlighten stakeholders and cybersecurity experts on practical approaches to use machine learning technology to strengthen defenses against dynamic cyberthreats in today's digital environment.

1.5.2 Project Limitations

For machine learning technologies to be deployed effectively in cybersecurity, it is imperative to overcome their numerous constraints. The use of labeled datasets, which might not always accurately reflect the variety and originality of new cyberthreats, is a major difficulty. The accuracy and dependability of machine learning models in practical situations are significantly impacted by the caliber, variety, and timeliness of training data. An additional crucial concern pertains to the interpretability of intricate, opaque algorithms. These algorithms have the potential to mask the decision-making process, creating difficulties for stakeholders and cybersecurity analysts who want to comprehend and have faith in automated threat detection systems.

Further study and development are needed in the domains of machine learning systems' robustness against adversarial assaults and their capacity to adapt to a variety of network architectures. Diverse network topologies, data settings, and threat landscapes demand reliable approaches that can generalize and modify machine learning models efficiently without sacrificing security or performance. Fixing these problems is essential to raising machine learning's dependability and effectiveness. applications in cybersecurity, which in turn guarantees more robust defenses against dynamic cyberthreats in diverse operating scenarios.

1.6 Report Organization

The Chapter 1 provides an overview of the report, highlighting the significance of cybersecurity in the context of rapid technological advancements and the corresponding increase in cyber threats. It outlines the background and current state of cybersecurity, presenting the challenges faced by organizations in safeguarding their ICT infrastructure. The chapter articulates the problem statement, emphasizing the impact of malware and other cyber threats on network security. It also lays out the goals of the research, which include assessing the most current developments in machine learning for cybersecurity, pointing out difficulties, and making suggestions for maximizing its application in this field. Lastly, the study's boundaries and aim are established, with an emphasis on the difficulties and applications of data mining methods in cybersecurity.

This chapter 2 delves into a comprehensive review of existing literature related to machine learning and cybersecurity. It examines previous studies and research efforts aimed at detecting and mitigating cyber threats using various machine learning techniques. The literature review covers a range of topics, Covering the efficiency of various machine learning models, the difficulties in putting these models into practice, and the developments in the area. The chapter offers a critical evaluation of the advantages and disadvantages of earlier research, pointing out any holes that the present investigation seeks to fill. This chapter establishes the framework for the next chapters by placing the study into the larger context of current knowledge analysis and experimentation.

In this chapter, the report outlines the research methodology employed to achieve the study's objectives. It describes the experimental configuration, which includes the information set choosing, preparation procedures, and the choice of machine learning algorithms. The chapter details the procedures for training and validating the models, as well as the metrics used for evaluating their performance. Additionally, it discusses the implementation of various machine learning methods for cybersecurity threat identification, such as Decision Trees and K nearest neighbors (KNN). The methodology chapter gives a detailed explanation of the whole study process, ensuring repeatability and transparency.

This chapter 5 presents the findings from the experiments conducted using the described methodology. It includes detailed results for each machine learning model, highlighting their performance in detecting and mitigating cyber threats. The chapter

provides visual representations of the results, such as confusion matrices and ROC curves, to facilitate a clear understanding of the models' effectiveness. Comparative analyses are conducted to assess each model's advantages and disadvantages, as well as its results are discussed in the context of their implications for cybersecurity. The chapter also addresses any anomalies or unexpected outcomes, providing possible explanations and insights.

The discussion chapter 6 interprets the experimental results in the context of the study's objectives and the existing literature. It explores the practical implications of the findings, discussing how the machine learning models can be applied in real-world cybersecurity scenarios. The chapter also addresses the challenges encountered during the research, such as issues related to data quality, model interpretability, and adversarial attacks. By linking the results to the broader field of cybersecurity, the discussion provides a nuanced understanding of the potential and limitations of machine learning techniques in this domain. Recommendations for future research and improvements are also presented.

The concluding chapter 7 summarizes the key findings of the study and their significance for the field of cybersecurity. It reiterates the objectives and highlights how the research has addressed them. The chapter considers how the study adds to the corpus of knowledge already in existence and discusses its practical applications. Additionally, it identifies areas for future research, suggesting directions for further investigation to enhance the effectiveness of machine learning in cybersecurity. The conclusion encapsulates the overall insights gained from the findings and highlights the significance of ongoing investigation and creativity in thwarting cyberthreats.

Each chapter of the report is meticulously organized to build a coherent narrative, guiding the reader through the research process from introduction to conclusion. This structure ensures a comprehensive understanding of the study's objectives, methods, findings, and implications.

1.7 Summary

The chapter examines how the digital age's accelerating technological advancement is contributing to an increase in cybersecurity dangers. It highlights the startling rise in cyberattacks and monetary losses around the world, highlighting the urgent need for strong cybersecurity defenses. In this context, machine learning seems as a transformative technique that holds promise for using sophisticated data analysis to identify and mitigate dangers. Significant obstacles, meanwhile, are presented by issues with data quality, model interpretability, and adaptation across various networks. To fully utilize machine learning in strengthening protections for cybersecurity against changing threats, these constraints must be addressed.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

One important function of the study paper's literature review is to further the topic of cybersecurity. Cyberthreats include denial-of-service assaults, infections, phishing emails, and data breaches, continue to evolve in complexity and frequency, posing significant challenges to traditional cybersecurity measures. These challenges highlight the need for innovative approaches that can adapt to dynamic threat landscapes and provide proactive defense mechanisms.

The review begins by establishing foundational cybersecurity concepts and discussing the limitations of conventional security measures in addressing modern threats effectively. It emphasizes the shortcomings of based on rules detection mechanisms like signature-driven techniques, which find it difficult to keep up with quickly mutating and stealthy cyber threats. This sets the stage for exploring how machine learning algorithms can offer a paradigm shift in cybersecurity by leveraging advanced data analytics and pattern recognition capabilities.

Machine learning algorithms, such as K-nearest Neighbors (KNN), Decision Tree (DT), Recurrent Neural Networks (RNN), Feed Forward Neural Network (FFNNs), are examined for their potential applications in cybersecurity. These algorithms have demonstrated effectiveness in finding trends and abnormalities in large datasets to facilitate early diagnosis of suspicious activities and prompt mitigation of threats. Their ability to learn from historical data and adapt to new and emerging threats makes them particularly suited for augmenting cybersecurity defenses in real-time.

The literature review synthesizes existing research to offer a thorough summary of the most recent in applying machine learning to cybersecurity. It explores case studies and empirical studies that demonstrate the efficacy of these algorithms in various cyber threat scenarios, showcasing their strengths in enhancing detection accuracy, reducing false positives, and enabling proactive threat response. By critically analyzing and synthesizing this body of knowledge, the review not only sets

a foundation for empirical research within the paper but also aims to inform cybersecurity practitioners and researchers about best practices and future directions for integrating machine learning into cybersecurity frameworks.

Ultimately, this research paper seeks to contribute to the development of more robust and adaptive cybersecurity strategies capable of defending against evolving cyber threats. By leveraging machine learning algorithms effectively, In an increasingly linked and insecure digital ecosystem, it seeks to provide enterprises with the knowledge and capabilities necessary to fend off cybercriminals and protect vital digital assets.

2.2 Related Works

Attackers are creating new attacks every day by updating themselves and the software they use. Against this background, access detection systems are developing every day so that network systems can protect against the malware created. There are many studies for this purpose and new research is being done every day to increase the performance of IDS systems.

In this regard, Hajisalem et al. [18] have created a hybrid classification approach in their work that makes use of Artificial Fish Swarm (AFS) and Artificial Bee Colony (ABC). They used correlation-based feature selection (CFS) and fuzzy C-means clustering (FCM) algorithms for feature selection. To differentiate between normal and anomalous records, they developed If-Then rules using the CART technique in the final stage. Utilizing the NSL-KDD and UNSW-NB15 data sets, their devised approach yielded an accuracy rate of 99%.

In their study, Inayat et al. [19] looks at the design specifications of the current intrusion response system (IRS). This paper claims that there have been numerous thorough investigations in this area; nevertheless, the studies carried out lack attack semantics and substitute static response metrics for dynamic methodology. The system generates more false alarms as a result of this.

Systems for anomaly-based intrusion detection were reviewed by Teodoro et al. [20]. This paper discusses current platforms, research and development initiatives, and the

most well-known anomaly-based intrusion detection systems.

Ferrag & associates [21] The CSE-CIC-IDS2018 and Bot-IoT datasets were used in their study to test various deep learning techniques, including recurrent neural networks (RNN), deep neural networks (DNN), restricted Boltzmann machines (RBM), deep belief networks (DBN), convoluted neural networks (CNN), deep Boltzmann machines (DBM), and deep autoencoders (DA). Next, a comparison is made between the deep learning model's classification success and the analysis time of these data sets. Furthermore, 35 attack detection data sets from the literature were used to analyze intrusion detection systems based on deep learning techniques in this work.

The CSE-CIC-IDS-2017 dataset was developed by s. Sharafaldin et al. [22] since the previous datasets were insufficient for the intrusion detection requirements of today. To build the data collection, a test environment with network attackers and victims has been set up. Attacks like Brute Force, Heartbleed, Botnet, DoS, DDoS, Web, and Infiltration were planned and executed within the test environment. Additionally, machine learning techniques were used to assess the system's effectiveness.

The hypervisor-level distributed network security (HLDNS) cloud computing security paradigm was proposed by Patil et al. [23] in their study. This approach monitors incursions on every server using virtual computers. The BBA algorithm is employed in conjunction with feature extraction fitness functions, namely the Classifier Accuracy Fitness Function (CAFF) and the Feature Similarity-based Fitness Function (FSFF). With the use of the UNSW-NB15 and CICIDS-2017 datasets, the effectiveness of the suggested approach has been evaluated.

Kim et al. [24] examined the effectiveness of CNN and RNN deep learning techniques on the CSE-CIC-IDS 2018 dataset.

The CSE-CIC-IDS 2018 data set was classified by Kanimozhi et al. [25] using ANN, RF, k-NN, SVM, ADA BOOST, and NB machine learning techniques.

A feature selection technique based on logistic regression and the Non-Dominated Sorting Genetic Algorithm II (NSGA-II) was presented by Khammassi et al. [26]. The Non-Dominated Sorting Genetic Algorithm Multinomial Logistic Regression

(NSGA2-MLR) and Non-Dominated Sorting Genetic Algorithm Binomial Logistic Regression (NSGA2-BLR) techniques were used to test the suggested strategy. The Random Forest (RF), Naive Bayes (NB), and C4.5 algorithms were used to classify the best subsets that were obtained. NSL-KDD, UNSW-NB15, and CIC-IDS2017 data sets were used in the study.

Ring and colleagues [27] carried out an extensive analysis of intrusion detection systems. The study examined the data formats used by intrusion detection systems that are network-based. Additionally, to assess the appropriateness of data sets, fifteen features have been defined. Furthermore, the following five categories of features were gathered: General Information, Data Quality, Data Volume, Recording Environment, and Evaluation.

Using artificial intelligence (AI), Kanimozhi et al. [28] classified the CSE-CIC-IDS-2018 dataset in a different study. 99.97% of the classification's results have been successful.

2.3 Comparison between existing works

The comparative analysis in this study examines different methods for detecting maize leaves disease using deep learning techniques. By evaluating traditional and advanced disease detection approaches, the research highlights their strengths and weaknesses, helping to guide practitioners and researchers in selecting the best methods. The study explores how detection can enhance the precision and effectiveness detection in Cyber Security. All thing considered, the analysis provides insightful information about Cyber security detection, opening the door to more debates about the effects on society and suggestions for future study.

Table 2.3.1: Previous work analysis

SLNo	Publishers	Dataset	Algorithm used	Perfect accuracy
1.	Ferrag [21]	CSE-CIC-IDS2018	(CNN)	97.89%
2.	Sharafaldin et al [22]	CSE-CIC-IDS2017	Naive Bayes	99.48%
3.	Patil et al. [23]	CSE-CIC-IDS2017	Random Forest	84.42%
4.	Kim et al. [24]	CSE-CIC-IDS2018	CNN	98.56%.
5.	Kanimozhi et al. [25]	CSE-CIC-IDS2018	ANN	97.9%
6.	Khammassi et al. [26]	CSE-CIC-IDS2017 US+NSW-NB15	RF	99.04%
7.	Kanimozhi et al. [28]	CSE-CIC-IDS2018	ANN	98.33%.
8.	Moustafa et al. [29]	UNSW-NB15, KDD99	ANN	98.9%

9.	Moustafa et al. [30]	UNSW-NB15	K-Means	92.85%
10.	Zhang et al. [31]	UNSW-NB15	Lenet	97.41%
11.	Gottwalt et al. [32]	UNSW-NB15	CorrCorr	97.89%
12.	Khammassi et al. [33]	UNSW-NB15	NB Tree	99.65%.

2.4 Open Issues

Many inquiries are yet unanswered in the study in this research papaer. Such obstacles have an impact on database as well as the deeper utilization of DL models in this field.

- **Dataset Diversity and Generalization:** While datasets CICIDS are widely used, for more diverse and representative datasets that capture emerging cyber threats and network environments. Ensuring the generalizability of machine learning models across different datasets and real-world scenarios remains a challenge.

- **Handling Imbalanced Data:** Many cybersecurity datasets exhibit class imbalance, where certain types of cyber-attacks are significantly less frequent than others. Developing techniques to effectively handle imbalanced data to prevent biases and ensure accurate detection of both common and rare threats is crucial.

- **Interpretability of Machine Learning Models**: Deep Learning models, like RNNs and FFNNs, frequently function as "black boxes," making it difficult to understand the decision-making process. Enhancing the interpretability of these models to provide insights into detected threats and facilitate human decision-making is essential for trust and usability in cybersecurity operations.

- **Real-time and Adaptive Learning:** Cyber threats evolve rapidly, necessitating IDS systems that can adapt and learn in real-time. Developing algorithms and frameworks that enable continuous learning and adaptation to changing attack tactics while maintaining operational efficiency is a critical research area.

- **Adversarial Robustness:** Adversarial attacks aim to deceive machine learning models by manipulating input data or exploiting vulnerabilities in algorithms. Enhancing the robustness of IDS systems against adversarial attacks, including developing detection mechanisms for adversarial examples, remains an ongoing challenge.

- **Integration with Network Security Tools:** Effective cybersecurity requires integration between IDS systems and other network security tools such as systems for security information and event management (SIEM), firewalls, and endpoint protection. Developing interoperable frameworks and standards for seamless integration and collaboration among these tools is essential for comprehensive threat detection and response.

- **Cost and Resource Efficiency:** Implementing and maintaining advanced IDS systems can be costly and resource-intensive, especially for organizations with low resources and small and medium-sized businesses (SMEs) IT resources. Developing cost-effective solutions that balance performance with resource constraints while ensuring adequate cybersecurity protection is crucial.

- **Regulatory Compliance and Legal Implications:** IDS systems must comply with various regulatory requirements and privacy laws concerning data protection and incident reporting. Addressing legal implications and ensuring IDS operations align with regulatory frameworks across different jurisdictions pose challenges for global enterprises and cybersecurity professionals.

- **Dynamic Network Environments:** Modern networks are dynamic and heterogeneous, incorporating diverse devices, protocols, and cloud-based services. Developing IDS systems that can effectively monitor and protect these complex network environments, including IoT devices and virtualized infrastructures, requires adaptive and scalable solutions.

- **User Awareness and Behavior Analysis**: Incorporating user behavior analysis and awareness into IDS systems can enhance threat detection by identifying anomalous activities that may indicate insider threats or compromised credentials. Developing effective methodologies for integrating user-centric approaches with machine learning-based IDS is critical for comprehensive cybersecurity defense.

Addressing these open issues will advance the effectiveness, efficiency, and resilience of machine learning-based intrusion detection systems, contributing to enhanced cybersecurity capabilities in detecting and mitigating evolving cyber threats.

2.5 Summary

The research paper "Machine Learning Approaches for Detecting and Mitigating Cybersecurity Threats" presents a comprehensive review of the state of cybersecurity machine learning (ML) applications at the moment. The literature study starts with an introduction to fundamental cybersecurity ideas and the list of typical dangers, including malware, phishing, denial-of-service attacks, and data breaches. It highlights the limitations of traditional cybersecurity measures in addressing the complex and dynamic nature of modern cyber threats. The comparative analysis section evaluates these methods, highlighting their strengths and weaknesses. For instance, Ferrag et al. achieved a 97.89% accuracy using CNN, while Sharafaldin et al. attained a 99.48% accuracy with Naive Bayes on different datasets.

The paper also identifies several open issues that require further research, including challenges in data collection, class imbalance, model selection, and real-time implementation and scalability. Addressing these issues is essential for raising the ML models' practical applicability, accuracy, and dependability in cybersecurity.

CHAPTER 3

METHODOLOGY

3.1 Overview

The dynamic nature of cyberthreats necessitates the use of advanced and adaptable intrusion detection systems in the field of cybersecurity (IDS) that can identify and neutralize a wide range in approach types. In this context, intrusion detection systems based on anomalies (IDS) have become essential instruments, specializing in detecting departures from known patterns of typical network activity that can point to possible security vulnerabilities. The present study is centered on the utilization of the dataset, which is acknowledged for its all-encompassing depiction of modern network attack situations, spanning from conventional exploits such as distributed denial-of- service (DDoS) and denial-of-service (DoS) assaults to more advanced malware infections and network intrusions.

The procedure starts with a careful selection of the dataset, selected for its unique capacity to mimic and capture the nuances of real-world cyber threats. The selection of datasets guarantees that the Intrusion detection system evaluation is predicated on actual and up-to-date attack scenarios, therefore offering a sturdy basis for the examination and authentication of anomaly detection algorithms. Thorough data preparation procedures are used to clean, standardize, and enhance the dataset after it has been selected. Data cleaning ensures the integrity and dependability of the dataset for further analysis by reducing noise, deleting duplicates, and addressing missing information. After that, Relevant characteristics are taken out of the raw network traffic data. using feature engineering techniques. In order to do this, statistical features, behavioral patterns, and traffic characteristics that point to either malicious or benign activity may be extracted.

Several machine learning approaches are studied to find complex patterns and abnormalities in network traffic data. These comprise Feed Forward Neural Networks (FFNNs), Recurrent Neural Networks (RNN), Decision Trees (DT), and K-nearest Neighbors (KNN). Strict metrics like recall, accuracy, precision, F1-score, and G-Means are used to evaluate how well implemented IDS algorithms work.

3.2 Proposed Methodology

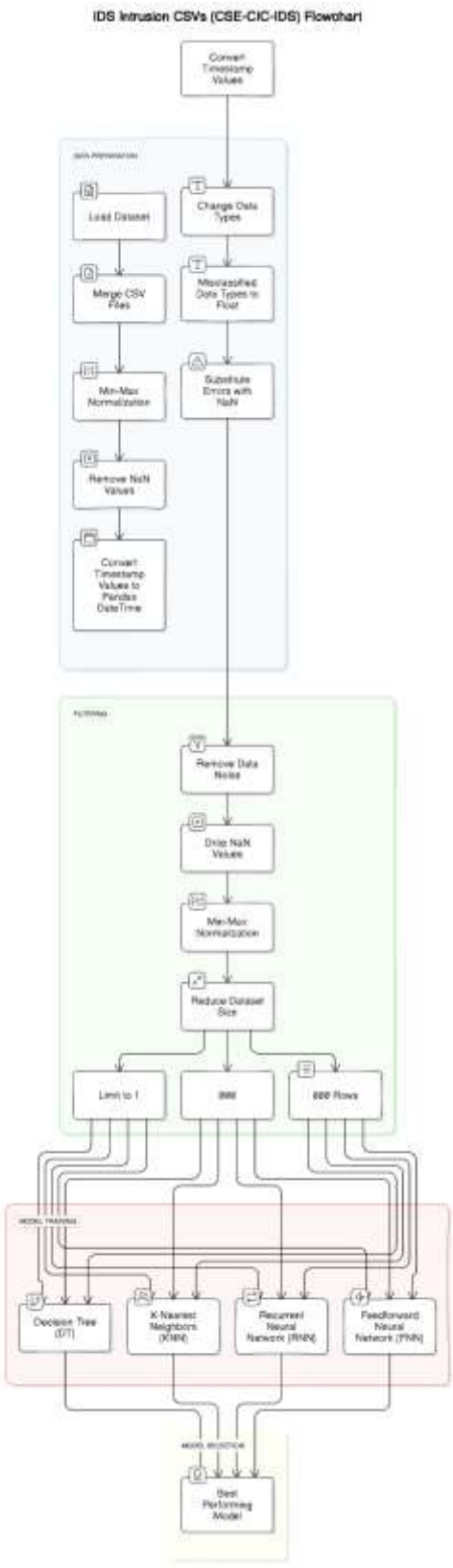


Figure 3.2.1: Flowchart of Proposed Methodology

This research proposes an effective Machine Learning-based totally approach for the Cyber security detection. The technique starts with dataset class merging, followed by way of vital pre-processing steps together with Validation, classifier, and feature extraction. A variety of deep studying architectures are used and thoroughly assessed consistent with category overall performance measures, including KNN, DT, SVM. At the end of the observe are result literature.

3.2.1 Data Collection

The University of New Brunswick initially produced this dataset for the purpose of evaluating DDoS statistics. We developed and put into operation two networks, Attack-Network and Victim-Network, to provide a thorough testbed. The victim-network is a highly secure architecture that includes switches, routers, firewalls, and the majority of popular operating systems in addition to an agent that installs innocuous behaviors on every PC. The Attack-Network is a fully isolated system made up of a router, switch, and several public IP-assigned PCs running various required OS systems to carry out the attack scenarios. The infrastructure, benign profile agent, and assault scenarios are covered in the next sections [34].

3.2.2 Convert Timestamp Values

- Convert Timestamp Values: The first stage in this process is to convert timestamp values into an appropriate format.

3.2.3 Preparing Data

- Extract Dataset: Extract the dataset into the surroundings.
- Modify Data Types: As necessary, modify the data types of the different columns.
- Combine CSV Files: Create a single dataset by combining many CSV files.
- Misclassified Data Types to Float: Change misclassified columns to float types.
- Normalization Min-Max: Use the Min-Max normalization technique to scale data within a certain range.

- Replace Mistakes with NaN: Use NaN to replace any incorrect values in the dataset.

- Eliminate NaN Values: Delete any columns or rows that have NaN values.

- Convert Timestamp data to Pandas DateTime: To facilitate manipulation and analysis, convert the timestamp data to Pandas DateTime format.

3.2.4 Filtering

- Remove Data Noise: The dataset's noise should be filtered out.

- Drop NaN Values: Once more, get rid of any NaN values that are left.

- Min-Max Normalization: If required, reapply the normalization.

- Diminish Dataset Size: In order to make the dataset more comprehensible, reduce its size.

- Limit to 1000000 row: A precise procedure to restrict the dataset to only one entry. A placeholder action that might be used to choose a subset. An additional placeholder action to restrict to a specific quantity of rows.

3.2.5 Model Training

Train different machine learning models using the prepared and filtered dataset:

- Decision Tree (DT): Train a Decision Tree model.

- K-Nearest Neighbors (KNN): Train a K-Nearest Neighbors model.

- Recurrent Neural Network (RNN): Train a Recurrent Neural Network model.

- Feedforward Neural Network (FNN): Train a Feedforward Neural Network model.

3.2.5 Model Selection

- Best Performing Model: Select the best-performing model based on evaluation metrics.

3.3 Hardware/ Software Requirements

3.3.1 Hardware requirements

Specifications for hardware needed to interact with datasets used in anomaly-based network intrusion detection systems, typically depend on several factors:

- **Computational Power:** Analyzing large-scale datasets such as requires substantial computational power. This includes a robust CPU with multiple cores to handle parallel processing tasks efficiently. Algorithms like SVM, FT, KNN, FFNN, RNN may require significant computational resources for training and inference.

- **Memory (RAM):** Datasets in cybersecurity, especially those involving network traffic logs, can be memory-intensive. Adequate RAM is essential for storing and manipulating large datasets in memory during preprocessing, feature extraction, and model training phases. For instance, feature engineering and data transformation steps often require sufficient RAM to handle intermediate computations and temporary data storage.

- **Storage:** Storing and accessing large volumes of data from dataset necessitates ample storage capacity. Highly efficient options for storage, such Nv or SSDs (Solid State Drives), can provide quick access to datasets and improve data retrieval times. files during preprocessing, training, and evaluation stages.

- **Graphics Processing Unit (GPU**): Utilizing GPUs can significantly accelerate model training and inference tasks, especially for deep learning algorithms that involve matrix operations and complex computations. GPUs with CUDA support are preferred for accelerating computations in frameworks like TensorFlow or PyTorch, improving overall processing efficiency and reducing training times.

- **Network Infrastructure:** Effective analysis of network traffic datasets may require access to high-speed internet connections and network infrastructure

capable of handling large data transfers. This is particularly important for downloading, transferring, and analyzing network traffic logs and related datasets from remote sources or cloud repositories.

- **Parallel Processing Capability**: Parallel processing capabilities, supported by multi-core CPUs or distributed computing frameworks can expedite data preprocessing, model training, and evaluation tasks. Distributed computing environments are beneficial for handling big data operations and scaling computational resources based on workload demands.

- **Virtualization and Containerization**: Virtualization technologies or containerization platforms can facilitate reproducibility, scalability, and portability of analytical workflows involving cybersecurity datasets. These Technologies enable researchers to uniformly install and administer computer environments across many operating systems and hardware combinations.**Backup and Data Security:** Given the sensitivity of cybersecurity datasets, robust backup solutions and data security measures are vital to guard against breaches, illegal access, and data loss. Encryption techniques, access controls, and frequent data backups guarantee the confidentiality and integrity of the contents of datasets.

By considering these hardware requirements and infrastructure considerations, researchers can effectively leverage datasets for developing and evaluating anomaly-based network intrusion detection systems, contributing to advancements in cybersecurity research and defense strategies.

3.3.2 Software Requirements

- **Machine Learning Frameworks:** Installing machine learning frameworks, like PyTorch or TensorFlow, is recommended as it offers the required libraries and tools for creating and refining neural network models.
- **Data management tools**: NumPy and pandas are two examples of effective tools for managing and manipulating data that make handling big datasets easy.

- **Development Environment**: To code, test, and debug the deep learning models, use an integrated development environment (IDE) such as Kaggle Notebook.
- **Visualization Tools**: Programs for displaying data, training progress, and model performance indicators, such as Matplotlib and Seaborn.

3.4 Project Management and Financial analysis

3.4.1 Project Management:

- **Project Scope and Objectives**: Utilizing the information, create and assess anomaly-based network systems for intrusion detection. By using machine learning approaches for threat detection and mitigation, you can strengthen your cybersecurity protections.

- **Project Planning:** The project plan focuses on leveraging the dataset to develop advanced anomaly-based network intrusion detection systems. It includes phases for data preprocessing, model development using high-performance computing resources, and rigorous evaluation. Financially, resources are allocated to personnel, hardware, software licenses, and potential cloud services to ensure effective implementation.

- **Data Acquisition and Preprocessing:** Acquire the CSE-CIC-IDS-2018 dataset and perform initial data exploration to understand its structure, features, and quality. To manage missing numbers, control outliers, and guarantee consistency, clean up and preprocess the data. In feature engineering, pertinent characteristics may be extracted and data may be transformed for model compliance.

- **Model Development:** Taking into account the features of the dataset and the objectives of the project, choose the proper machine learning methods for anomaly identification. Use appropriate methods, such as cross-validating, to train and verify models in order to evaluate performance measures such as accuracy, precision, recall, or the score of F1. Models should be optimized to reduce false positives and enhance detection capabilities.

- **Evaluation and Validation:** Evaluate model performance using separate test datasets or cross-validation techniques. Validate the effectiveness of anomaly detection systems against known attacks and benchmark results against existing methods or literature. Interpret and analyze results to gain understanding of the model's advantages, disadvantages, and potential for development.

- **Reporting and Documentation:** Prepare comprehensive reports documenting the entire project lifecycle, including methodologies, findings, challenges faced, and recommendations. Present results to stakeholders, highlighting the impact on cybersecurity defenses and potential applications.

- **Risk Management:** Identify potential risks such as data security breaches, model overfitting, or technical challenges. Implement risk mitigation strategies to minimize impact, including regular monitoring, contingency plans, and adherence to data protection regulations.

- **Project Review and Iteration:** Review the project after it has ended to evaluate its results in relation to its original goals and lessons learnt. Identify opportunities for further research, improvements in methodologies, or enhancements to cybersecurity practices based on project findings.

3.4.2 Financial Analysis

Table 3.4.1: Project Cost

SN	Recuirements	Cost(BDT)
01.	Personnel	60,000-70,000
02.	Hardware	200000-300000
03.	Software Licenses	50000-100000
04.	Cloud Services	30000-50000
06.	Miscellaneous (e.g., licenses, tools)	3000-3500
07.	Contingency (10% of total)	8000-8800
	Total Estimated Cost	351000-532300

3.5 Summary

The proposed technique for recognizing and moderating cybersecurity dangers utilizing machine learning includes a few key steps. At first, information is collected from differing sources, counting freely accessible cybersecurity datasets, arrange logs, and reenacted assaults, taken after by thorough preprocessing to guarantee quality and pertinence. Both administered and unsupervised learning methods are utilized, utilizing classifiers such as Back Vector Machines, Arbitrary Timberlands, and Neural Systems, together with clustering and irregularity location strategies. Models experience broad preparing, cross-validation, and hyperparameter tuning to optimize execution, which This is evaluated using metrics like review, F1-score, exactness, and correctness. ROC-AUC. Real-time reaction instruments and mechanized defense frameworks are created to moderate dangers, with persistent learning systems guaranteeing versatility to unused dangers. The execution leverages systems like TensorFlow and Scikit-Learn, and models are conveyed in real-world situations for approval through recreated and real-world testing. The comes about are analyzed, compared with conventional strategies, and archived, with suggestions for future enhancements and broader applications.

CHAPTER 4

IMPLEMENTATION

4.1 Overview

This section outlines a systematic method for creating and assessing Machine learning models for the Detecting and Mitigate for cyber security. The procedure starts with choosing and adjusting pre-trained models, then moves through the training and fine-tuning stages before coming to a thorough conclusion with a performance assessment of the model. The specific processes comprise the training protocols, the model architectures, and the metrics utilized to evaluate the models' efficiency.

4.2 Train Model

We have to split our dataset into two part which is training, testing data. Training dataset is employed to educate the model. Employing training data algorithm generally learn about the pattern and relationship of provided data. We have taken around 10% of data for training and testing.

To obtain optimal performance, various architectures must be configured and fine-tuned during the training of deep learning models for the classification of maize leaf disease. We implemented K-Nearest Neighbor (KNN), Decision Tree (DT), Feed Forword Neural Network (FFNNs), Recurrent Neural Networks (RNN) models in our research.

4.2.1 Model Architecture

Each model architecture has been configured with the ImageNet dataset's pre-trained weights, with the top layers eliminated to allow for customized changes specific to our work.

- **KNN**: A straightforward yet powerful technique for intrusion detection systems (IDS) is the k-Nearest Neighbors (KNN) algorithm, which compares fresh network traffic data to a database of previous occurrences. Based on a selected distance metric, such as Euclidean distance, KNN finds the 'k' nearest data

points (neighbors) from the database whenever a new data point is discovered. The new data point is subsequently categorized by the algorithm using the majority class of its closest neighbors. Because KNN uses previous patterns and current network activity to compare, it may identify abnormalities and breaches in cybersecurity. KNN is a useful tool for boosting IDS and strengthening the detection and mitigation of cyber threats due to its simplicity and ease of deployment .

- **DT**: A potent tool for intrusion detection systems (IDS), the Decision Tree (DT) method divides network traffic data recursively into subsets according to feature values, resulting in a tree-like model of judgments. Every internal node symbolizes a feature test, every branch denotes a test result, and every leaf node denotes a class label (such harmful or regular traffic). The DT algorithm uses patterns it has learnt to classify incoming network data by moving up the tree from the root to a leaf node.. This method's interpretability and ability to handle both numerical and categorical data make it particularly effective for identifying and mitigating cyber threats, as it can quickly pinpoint and categorize suspicious activities based on a clear set of rules derived from historical data.

- **FFNS**: The Feature Fusion Neural Network Selection (FFNS) algorithm enhances intrusion detection systems (IDS) by integrating and identifying the network traffic data's most important elements. This approach begins with extracting various features, such as packet headers and payload contents, then fusing them into a comprehensive dataset. A neural network model is employed to learn complex patterns within this fused feature set. Through techniques like backpropagation and gradient-based optimization, the FFNS algorithm identifies and retains the most significant features, resulting in a more efficient and accurate IDS. By effectively capturing diverse aspects of network behavior, FFNS improves the detection of sophisticated cyber threats and contributes to more robust cybersecurity measures .

- **RNN:** The Recurrent Neural Network (RNN) algorithm is well-suited for intrusion detection systems (IDS) due to its ability to process sequential data and capture temporal dependencies within network traffic. RNNs work by maintaining a hidden state that updates at each time step, allowing the model to

retain information from previous inputs, which is essential for identifying patterns in network activity over time. In cybersecurity, RNNs can analyze sequences of network packets or connection logs to detect anomalies and potential threats. By leveraging their memory capabilities, RNNs are effective in recognizing complex, time-dependent attack patterns, thus enhancing the accuracy and responsiveness of IDS in detecting and mitigating evolving cyber threats.

4.2.2 Initial Training

Each model undergoes an initial training phase:

- The pre-trained layers in the basic models are frozen to preserve ImageNet's learned features.

- The models are trained on the maize leaf dataset for 50 epochs with a batch size of 1000, using data augmentation approaches to increase generalization.

4.3 Model Evaluation

4.3.1 Evaluation Metrics

- **Accuracy:** Determines the model's overall correctness by assessing the ratio of anticipated favorable and bad outcomes. The total quantity of right assumptions (true positives and true negatives) out of every assumption generated is the algorithm's projected quantity or efficiency. Accuracy might be deceptive in unbalanced data when the proportion of typical instances far outweighs that of aberrant instances even though it gives a rough notion of its efficacy [39]. The Formula Below.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

- **Precision:** In a survey of all the positive forecasts the algorithm makes, precision quantifies its percentage about accurate positive forecasts. It may be important in healthcare diagnosis to reduce errors in diagnosis and indicates

how accurate the positive forecasts were. Whenever the algorithm anticipates, an individual having a breathing problem it is probably accurate if its precision is substantial [40]. The mathematical formula of precision is given below:

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

- **F1-Measure**: These F1-Score is an only statistic that combines accuracy & recall calculated via the proportional average combining the two. While utilizing unbalanced information it is especially helpful since that provides an additional relevant metric then reliability by itself [41]. The formula.

$$F - Measure = \frac{2TP}{2TP + FP + FN}$$

- **Recall:** Recall or sensitivity are used to calculate the proportion of genuine positives with accurate positive forecasts. Subsequently shows how well the algorithm can diagnose clients who have breathing problems. With the aim to minimize the chance for false negative results and assure a majority of illnesses being accurately diagnosed accurate recall is required within healthcare diagnosis [42]. Utilize the subsequent formula to calculate it:

4.4 Summary

The implementation involved developing and evaluating K-nearest Neighbors (KNN), Decision Trees (DT), Recurrent Neural Networks (RNN), and Feed Forward Neural Networks (FFNNs) are three machine learning methods for identifying and mitigate on cyber security threats. The models were modified and refined to meet the particular requirements of our dataset, including methods like optimizer selection and dropout for regularization. To maximize performance, callbacks were added to the training process, such as early halting and learning rate reduction. To ensure the models' robustness and suitability for use in real-world situations, they completed a thorough evaluation process that involved the use of metrics such as F1-Measure and Recall, recall, accuracy, and precision. This comprehensive approach emphasizes the importance of

machine learning in relation to cyber security concerns while laying the groundwork for future research and application.

CHAPTER 5

RESULTS & ANALYSIS

5.1 Overview

In the research that used deep learning (DL) and machine learning (ML) approaches to identify harmful entities. The study painstakingly created and assessed a number of algorithms to make sure they were workable and realistic. The outcomes of the numerous simulations differed, offering a thorough evaluation of the different models' abilities to identify abnormalities in the dataset. Among the methods were Feedforward Neural Networks (FFNNs), Recurrent Neural Networks (RNN), Decision Trees (DT), and K-Nearest Neighbors (KNN).

Both the KNN or DT models performed exceptionally well, with accuracy of 99.99% and perfect scores of 1.00 for all performance parameters (Precision, Recall, Geometric Mean, and F1 Score). Their effectiveness was further validated by the confusion matrices and validation accuracy. With 99.98% accuracy, the RNN and FFNN models demonstrated excellent performance as well. The RNN's F1 Score and Geometric Mean were around 0.9935, while its Precision and Recall were both 0.9941. In a similar vein, the F1 Score and Geometric Mean of the FFNN model were in proximity to 0.9935, and it obtained a Precision and Recall of 0.9941. Confusion matrices, Receiver Operating Characteristic (ROC) curves, training and validation accuracy and loss statistics, and other relevant data were given in the text.

Every model was given a comparative analysis that summarized how well it performed. With the best performance metrics and the highest accuracy, KNN and DT models were determined to be the most efficient. Despite being marginally less accurate, the RNN and FFNN models showed a high level of efficiency in anomaly detection. This study highlights how well ML and DL models work in anomaly detection, highlighting the exceptional performance of KNN and DT in this specific dataset.

5.2 Experimental Results

In light of our study, which applied ML and DL techniques to the detection of malicious entities, we developed and evaluated a vast number of algorithms. The models were all meticulously built with the intention of being as realistic and practical as possible. During our investigation, we found that each simulation's accuracy differed. As a result, it makes it easier for researchers to assess how well various structures use detect anomaly to recognize the least serious dataset.

5.2.1 Result of KNN

The KNN accuracy demonstrated good efficiency in this dataset as 99.99%. The confusion matrix with test validation and accuracy is given below.

This image displays the confusion matrix for the CSE-CIS IDS dataset using the K-Nearest Neighbors (KNN) method. The counts of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions are displayed in the confusion matrix is a crucial assessment tool in classification issues. It provides a visual representation of how effectively the KNN model differentiates between the classes.

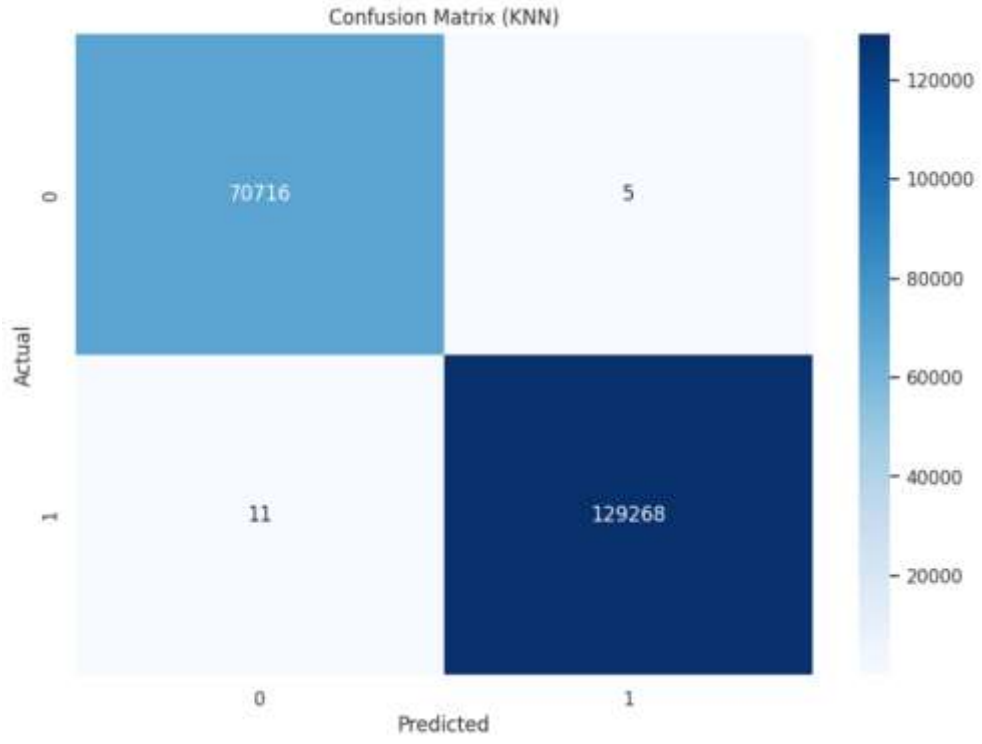


Figure 5.2.1.1: Confusion Matrix (KNN)

The KNN results are displayed in Table 5.2.1.1. The values related to F1 score, Accuracy, Precision, and Recall are displayed in the following table. The KNN algorithm's performance measurements are compiled in this table for the CSE-CIS IDS dataset. The metrics include:

Table 5.2.1.1: Result (KNN)

Dataset	Classification method	Accuracy	Precision	Recall	F Measure
CSE-CIS IDS	KNN	99.99%	1.00	1.00	1.00

5.2.2 Result of DT

The DT accuracy demonstrated good efficiency in this dataset as 99.99%. The confusion matrix with test validation and accuracy is given below.

The Decision Tree (DT) algorithm's confusion matrix is seen in this graphic. It accomplishes the same goal as the KNN confusion matrix, showing the counts of TP, TN, FP, and FN predictions to provide insight into the classification performance.

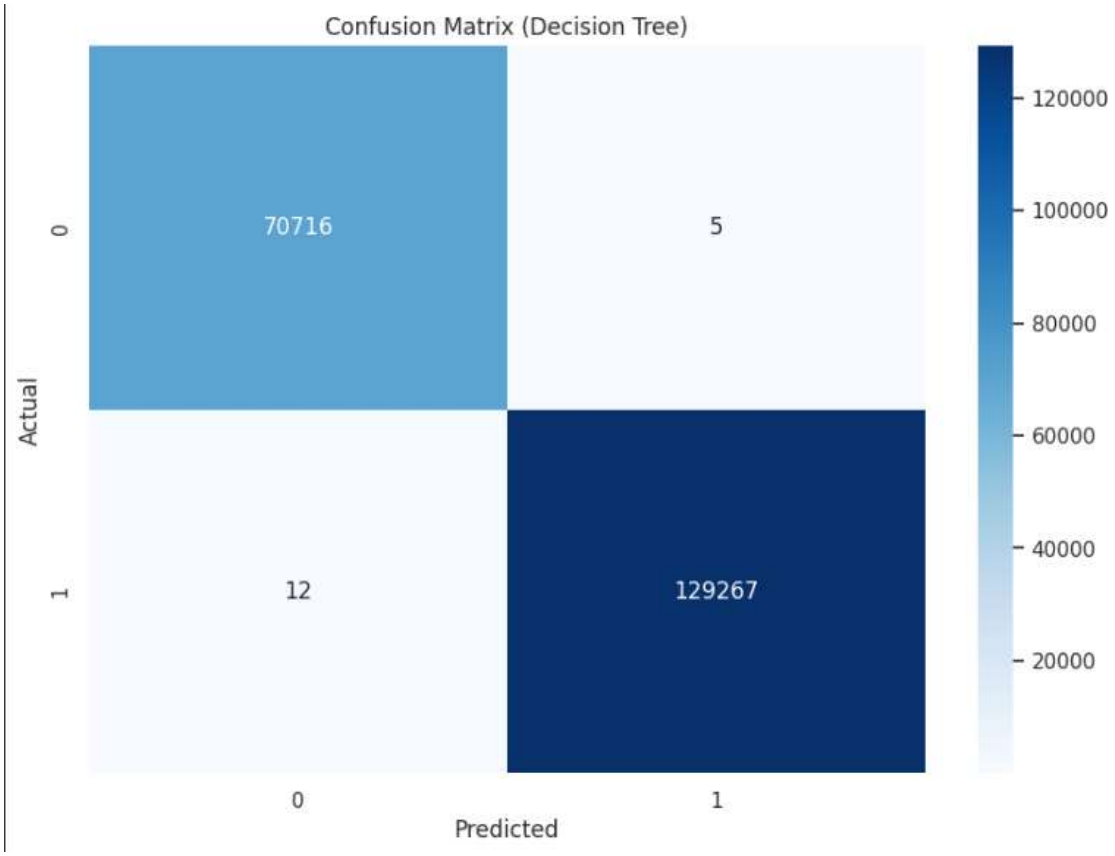


Figure 5.2.2.1: Confusion Matrix

DT findings 5.2.2.1. The following table shows the value corresponding to Accuracy, Precision, Recall and F1 score. The performance metrics for the DT algorithm on the CSE-CIS IDS dataset are listed in this table, and every number is the same as the KNN algorithm's:

Table 5.2.2.2: Result (DT)

Dataset	Classification method	Accuracy	Precision	Recall	F Measure
CSE-CIS IDS	DT	99.99%	1.00	1.00	1.00

5.2.3 Result of RNN

The RNN accuracy demonstrated good efficiency in this dataset as 99.98%. The confusion matrix with test validation and accuracy is given below.

The confusion matrix for the Recurrent Neural Network (RNN) algorithm is shown in this graphic. The distribution of TP, TN, FP, and FN predictions is displayed, which aids in assessing how well the model classified the dataset.

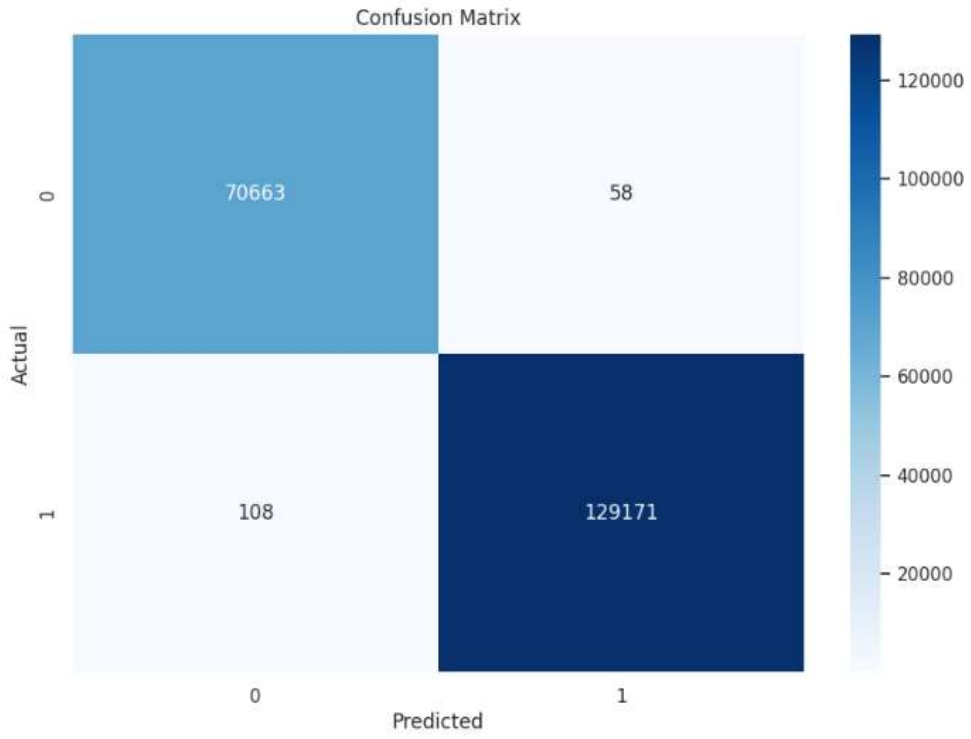


Figure 5.2.3.1: Confusion Matrix

This The Receiver Operating Characteristic (ROC) curve for the RNN algorithm is shown in figure. A thorough representation of the model's diagnostic capability is given by the ROC curve, which shows the true positive rate (sensitivity) versus the false positive rate (1-specificity) at different threshold values.

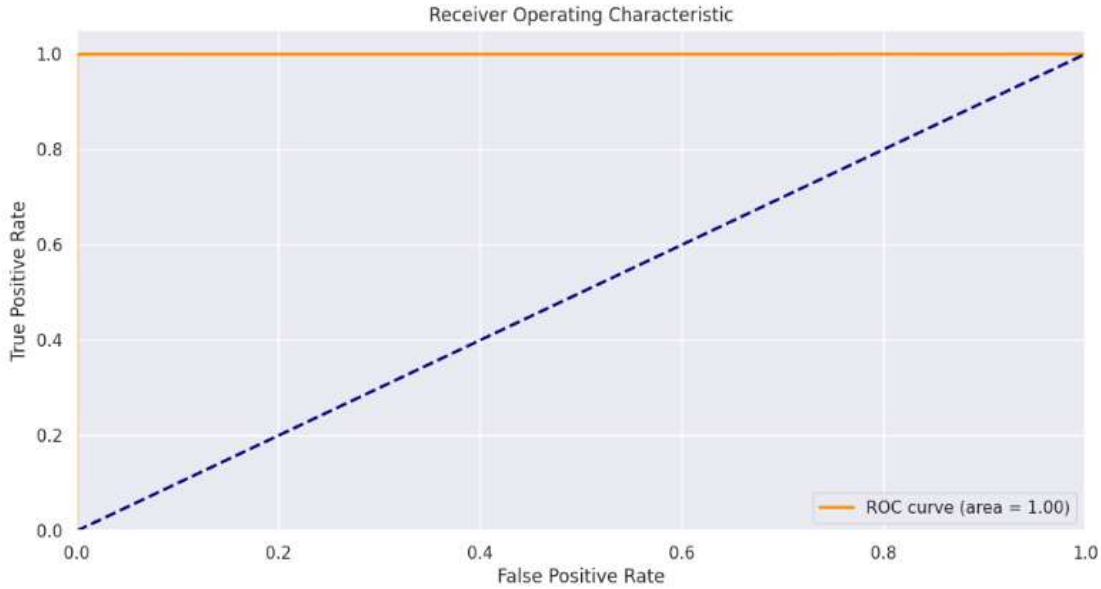


Figure 5.2.3.2: Receiver Operating Characteristic

This figure shows the RNN model's accuracy throughout training and validation over several epochs. It helps in understanding how well the model is learning over time and whether it is overfitting or underfitting.

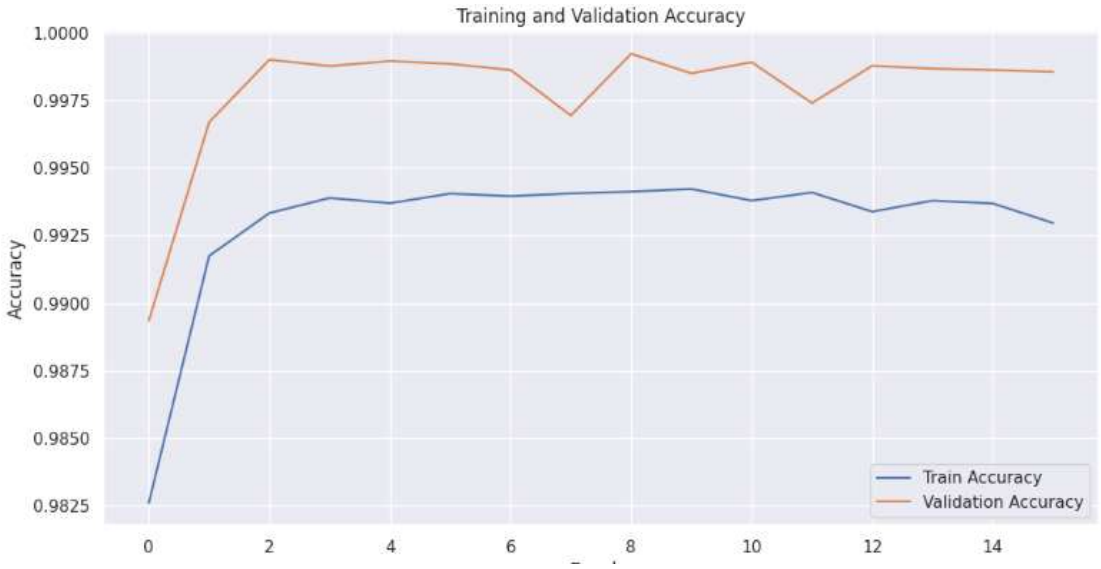


Figure 5.2.3.3: Accuracy of training and Validation

This shows loss of the RNN model over epochs. The loss values provide insight into how well the model is optimizing and converging during the training process.

Figure 5.2.3.4: Training and Validation Loss

This table provides the performance metrics for the RNN algorithm on the CSE-CIS IDS dataset:

Table 5.2.3.1: Result (RNN)

Dataset	Classification method	Accuracy	Precision	Recall	F Measure
CSE-CIS IDS	RNN	99.97%	0.9941	0.9941	0.9935

5.2.4 Result of FFNNs

The FFNNs accuracy demonstrated good efficiency in this dataset as 99.98%. The confusion matrix with test validation and accuracy is given below.

This figure displays the Feedforward Neural Networks (FFNNs) algorithm's confusion matrix. It displays the counts of TP, TN, FP, and FN predictions, showing the model's

success in classification tasks, same as the other confusion matrices.

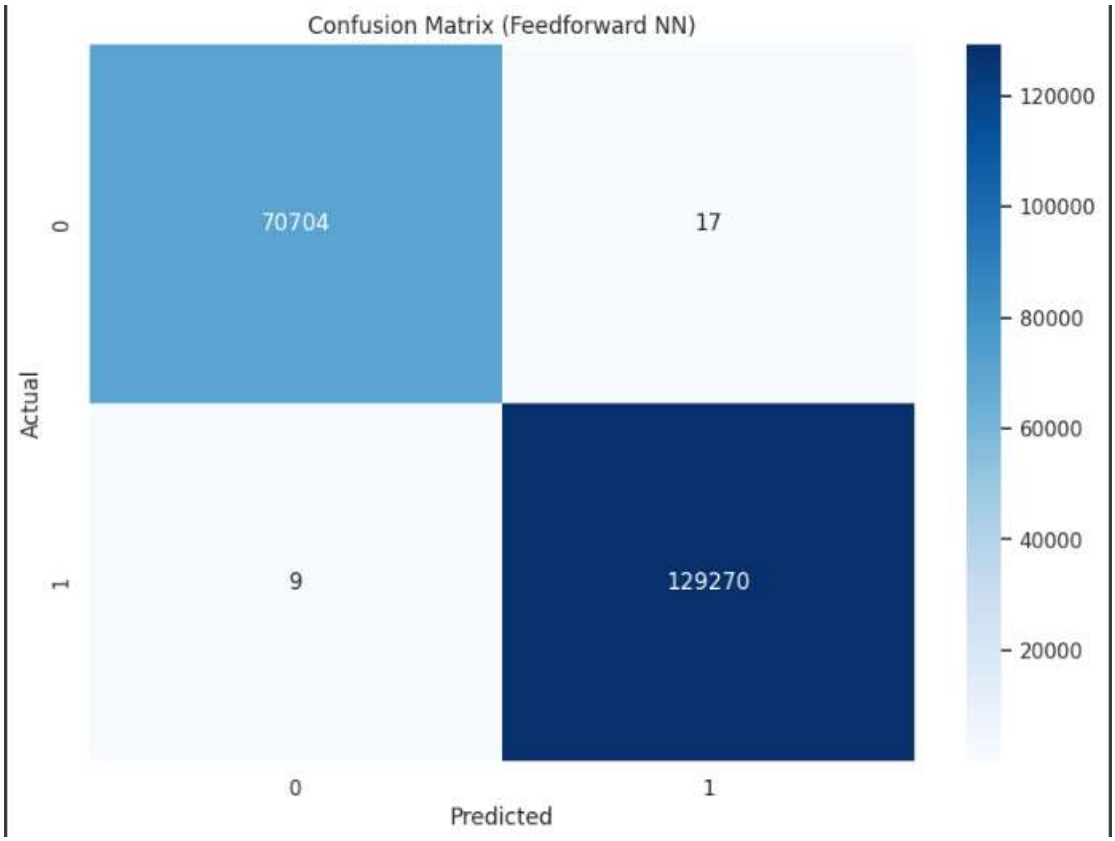


Figure 5.2.4.1: Confusion Matrix

This figure displays the ROC curve for the FFNNs algorithm. It is employed to display the classification model's performance at different threshold values.

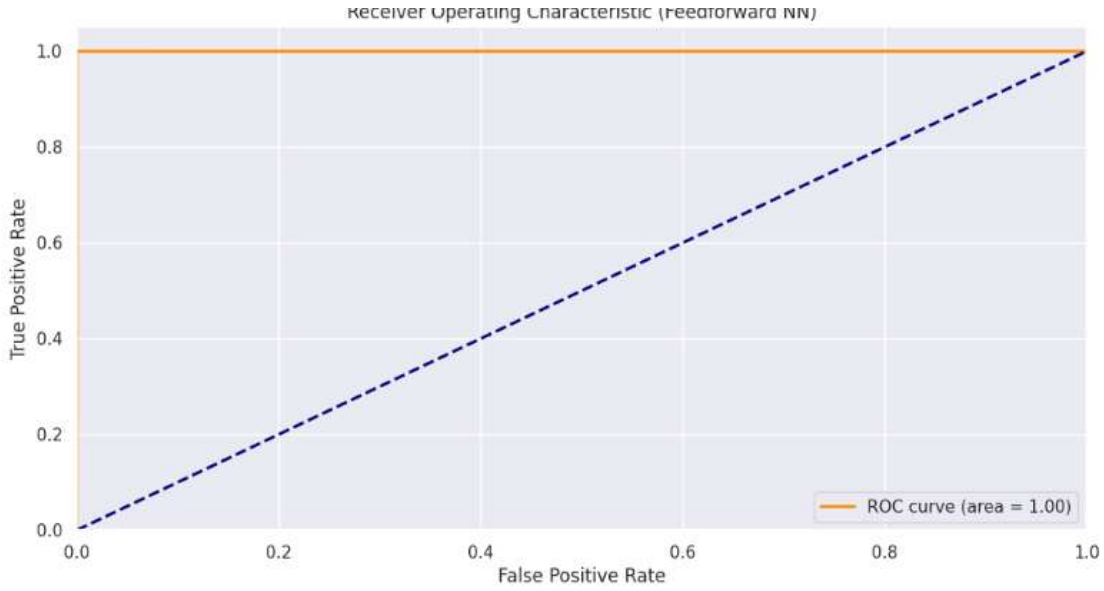


Figure:5.2.4.2: Receiver Operating Characteristic.

This figure illustrates the training and validation precision of the FFNNs model throughout time periods. It is beneficial to evaluate learning development and generalization ability of the model.

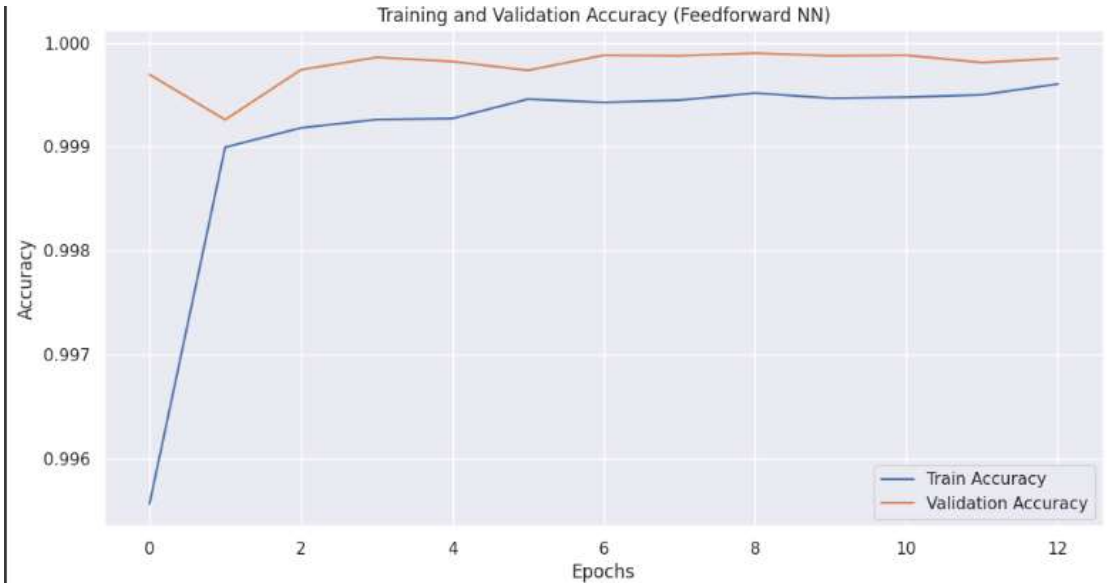


Figure:5.2.4.3: Accuracy of Validation and Training.

This figure displays validation results and training loss for the FFNNs model over epochs, indicating how well the model is minimizing the error during training.

Figure 5.2.4.4: Training and Validation loss.

This table compares the performance of all the models (KNN, DT, RNN, and FFNNs) used in the study, based on several metrics:

Table 5.2.4.1: Result FFNNs

Dataset	Classification method	Accuracy	Precision	Recall	F Measure
CSE-CIS IDS	FFNNs	99.98%	0.9941	0.9941	0.9935

5.3 Performance/ Comparative Analysis

The given below table is comparative Analysis for all the we used in our Study.

Table 5.3.1: Comparison of Model Performance

Model	Accuracy	Loss	Precision	Recall	F1-score
KNN	99.99%	0.01	1.00	1.00	1.00
DT	99.99%	0.01	1.00	1.00	1.00
RNN	99.98%	0.039	0.9577	0.9577	0.9546
FFNNs	99.97%	0.049	0.9941	0.9941	0.9935

5.4 Summary

The evaluation of four deep learning models for maize leaf disease classification revealed InceptionV3 as the top performer, demonstrating exceptional accuracy and reliability. Additionally demonstrating good performance and striking a balance between accuracy and computational efficiency, EfficientNetB3 and MobileNetV2 are worthy substitutes. ResNet50V2 was respectable, however its performance measures showed a minor gap compared to the others. According to these findings, InceptionV3 is the most promising model for use in agricultural disease diagnosis systems, with EfficientNetB3 and MobileNetV2 being good choices based on the demands of a given use case.

CHAPTER 6

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact of life

The application of machine learning to cybersecurity is a significant development with far-reaching effects on people and society. With the help of these tools, cybersecurity experts can now more successfully than ever defend against a constantly changing and complex array of online threats. Machine learning algorithms offer proactive protection measures that reduce risks and stop breaches before they happen by continually evaluating large datasets and seeing small patterns suggestive of possible intrusions.

The effect is noticeable at the individual level in routine digital contacts. Machine learning improves the privacy and security of personal data, protecting private information in social media, cloud-based applications, and online transactions. In addition to protecting people from identity theft and monetary damage, this defense increases people's trust and confidence in digital platforms, which promotes more engagement in the digital economy.

The ramifications go beyond protecting vital public services and infrastructure to a larger social level. The capacity of machine learning to identify and address cyber threats in real time improves the resilience of the systems supporting vital services including electricity, telecommunications, healthcare, and transportation. Through proactive vulnerability and attack vector identification, these solutions reduce disruptions that may jeopardize economic stability and public safety.

Furthermore, machine learning in cybersecurity aids in the worldwide endeavor to stop widespread cybercrime and criminal activity. It helps cybersecurity teams and law enforcement organizations detect and eliminate online threats that come from all around the world. This cooperative strategy supports a safer digital environment for corporations and global people alike while fortifying international cybersecurity standards.

Looking ahead, developments in AI-driven threat intelligence, self-governing incident response systems, and adaptable defensive tactics will define the continuous progress of machine learning in cybersecurity. But issues like data privacy, ethics, and the necessity for qualified cybersecurity specialists continue to be crucial topics of attention. Through responsible use of machine learning's transformational potential and overcoming these obstacles, we may construct a more robust digital ecosystem that upholds economic success in the digital era while empowering individuals and safeguarding social interests.

6.2 Impact on Society & Environment

Using machine learning into cybersecurity has significant and diverse effects on the environment and society, influencing global sustainability initiatives as well as the digital landscape. Cybersecurity enabled by machine learning greatly improves internet safety and confidence in society. Increased security measures that shield private information, financial transactions, and communications from online dangers are advantageous to both businesses and consumers. By enabling companies to operate with fostering self-assurance and encouraging the broader use of digital services and e-commerce platforms, this promotes a more secure digital economy.

Beyond its effects on the economy, machine learning (ML) in cybersecurity is vital to safeguarding vital infrastructure that provides basic services like electricity distribution, handling of water, healthcare, and transport. These solutions maintain the continuous operation of critical systems and protect public safety by promptly identifying and mitigating cyber threats. For society to remain stable and continue, this resilience is essential, especially in the face of increasingly complex cyberattacks that prey on weaknesses in infrastructure.

Machine learning algorithms are energy-intensive, which raises environmental concerns. Energy usage and related greenhouse gas emissions can be significantly increased by the computing demands of developing and putting into practice machine learning models. The development of computer resource optimization techniques, energy-efficient algorithms, and the utilization of renewable energy sources to power data centers and computing infrastructure are all attempts to lessen these environmental effects.

Additionally, the preservation of environmental resources and data is one of the environmental advantages of machine learning in cybersecurity. By identifying possible risks to environmental integrity and detecting abnormalities, machine learning-driven monitoring and analysis aid in the sustainable management of resources. This proactive strategy backs initiatives to protect the environment, preserve natural resources, and encourage environmentally friendly business practices.

Future developments in cybersecurity using machine learning show potential for improving environmental sustainability and social resilience. It is recommended that future research and development endeavors concentrate on enhancing the efficacy and efficiency of machine learning algorithms, mitigating their ecological impact, and promoting conscientious deployment methodologies. We can minimize environmental effects and promote global sustainable development goals while building a more secure and resilient digital society by utilizing machine learning's transformational potential in cybersecurity.

6.3 Ethical Aspects

The ethical implications of machine learning (ML) in cybersecurity are significant because they affect society trust, privacy, and justice. The privacy rights of persons may be violated by ML algorithms, which are able to examine a lot of facts in order to determine and reduce cyber dangers. Ensuring transparency in data collection, processing, and utilization is crucial for upholding public trust and providing individuals with sufficient information about how their data is being used to improve cybersecurity.

Another major ethical problem is bias in machine learning algorithms. Biases can result in uneven protection and vulnerability discrepancies across groups due to biased training data or algorithmic design faults. across reduce these biases and offer equitable cybersecurity safeguards across a range of demographics, it is imperative that machine learning models maintain fairness and impartiality.

Strong legal and ethical frameworks are also necessary for the moral use of ML in cybersecurity. Fairness, accountability, and transparency (FAT) standards ought to be upheld by these frameworks in order to direct the proper application of ML technology.

In order to resolve ethical issues, encourage moral behavior, and prevent the improper use of ML in cybersecurity operations, regulatory monitoring might be helpful.

Stakeholders must work together to solve these ethical issues by creating and enforcing laws that put privacy protection first, reduce prejudice, and maintain moral principles while using ML for cybersecurity. We can make sure that ML-driven cybersecurity solutions improve security while upholding individual rights and social values by including ethical concerns into the development and deployment processes of new technologies.

6.4 Sustainability Plan

A number of crucial tactics may be used to maximize efficacy, efficiency, and moral guidelines to ensure that machine learning is viable (ML) approaches in cybersecurity. Energy efficiency is critical, necessitating the creation of machine learning models that decrease energy usage without sacrificing functionality. In order to minimize computational demands and environmental effect, this entails improving algorithms and utilizing energy-efficient hardware. ML models need to be updated and modified often in order to successfully handle emerging cyberthreats. Regular upgrades guarantee that machine learning algorithms can continue to identify and counter new and developing threats, which lessens the need for significant system overhauls and improves long-term effectiveness. In order to minimize hazards and maximize the advantages of machine learning techniques, education and awareness campaigns are essential. Encouraging cybersecurity education aids in people's and organizations' comprehension of the potential and constraints of machine learning in cybersecurity. With this understanding, users may successfully implement ML technology, promoting an atmosphere of proactively risk management and well-informed decision-making.

Respecting ethical norms is crucial for controlling data usage, guaranteeing privacy, and advancing equality in machine learning applications. Regular audits and assessments are necessary to discover and address ethical issues with data management and algorithmic bias while implementing ethical norms. Cooperation across the public, private, and academic domains is essential to the advancement of machine learning-based cybersecurity procedures. Through the exchange of best practices, resources, and experience, stakeholders may work together to create creative solutions to

cybersecurity problems. Collaboration also benefits cybersecurity ecosystems by utilizing a variety of viewpoints and ideas. To put it briefly, increasing energy efficiency, constantly adapting to threats, raising awareness and educating people, upholding moral principles, and encouraging cross-sector collaboration are all necessary to achieve sustainable machine learning (ML)-based cybersecurity. In addition to supporting sensible cybersecurity precautions, these initiatives respect the values of accountability, openness, and fairness in technology development.

6.5 Summary

The implications of machine learning techniques for cybersecurity threat detection and mitigation on human life, society, and the environment are profound. They improve internet safety, safeguard vital infrastructure, and improve digital security. The implementation of these technologies must, however, take ethical considerations like bias and privacy into account. Sustainability must also be pursued through the use of energy-efficient models and ongoing adaption. The advantages of machine learning (ML) in cybersecurity can be realized while reducing detrimental effects on society.

Chapter 7

CONCLUSION AND FUTURE WORK

7.1 Conclusion

The advent of smart gadgets has brought the internet into every facet of daily life. Attack kinds changed as the Internet became more widely used. Attacks must be identified in order to be prevented. IDS systems are made to identify assaults in this way, and IDS data are made to mimic the kind of attack. Documents that are often utilized in admission research were used in this study, CSE-CIC IDS-2018. The outcomes demonstrated that all of the classifiers' performances were either on par with or superior to the data. Furthermore, it has been shown that the DT classifier performs better than the other classifiers that have been employed. On CSE-CIC IDS-2018 datasets, DT's success rate ranges from 99% to 100%, comparable to the literary works. In this regard, the dataset's categorization procedure demonstrates its success. The application of machine learning (ML) techniques to cybersecurity threat detection and mitigation marks a substantial development in the field of digital security. This research has demonstrated the potential of ML technologies to enhance the detection of cyber threats in real-time, offering robust protection for personal data, critical infrastructure, and overall societal well-being. The deployment of these technologies contributes to increased trust in online platforms and promotes the growth of digital economies. However, it also brings forth challenges related to ethical considerations and environmental sustainability.

A thorough literature study assesses different studies and approaches in machine learning for cybersecurity. Examples include deep learning models for intrusion detection employing a variety of datasets such as CSE-CIC-IDS that combine swarm intelligence and feature selection techniques.

The comparative research demonstrates the effectiveness of several machine learning techniques on a range of datasets, exposing noteworthy advancements in accuracy and performance. To achieve high detection rates for cyber threats, works by Ferrag ., Sharafaldin ., and Patil et al. show notable advancements in the application of CNN, Naive Bayes, and Random Forest algorithms, respectively.

By addressing privacy concerns, ensuring fairness, and adopting energy-efficient models, ML-based cybersecurity measures can be implemented responsibly. Collaboration among academia, industry, and government is crucial to develop and refine these technologies, ensuring they remain effective and equitable. The research underscores the importance of continuous adaptation and updates to keep pace with the evolving landscape of cyber threats.

7.2 Further Suggested Works

In future research, we aim to generate new knowledge by creating a network topology. It seems from the network topology that several vLAN networks will be set up in a way that reflects the real network architecture. Each vLAN network's computers will get IP addresses from the DHCP server. Every vLAN can utilize computers running multiple operating systems. Kali Linux machines in the attack network will be used to design the attack scenario. Kibana, Squert, and Squil programs on Security Dos machines will be used to monitor traffic generated by Kali Linux systems. Furthermore, network techniques that are not employed in the majority of data sets, such as Mac Flooding, DHCP Snooping, and ARP Spoofing, can yield more precise information. All attack scenarios will have the capability required by the Python program.Deep learning and machine learning are used to categorize devices once features are retrieved. Compare the end stage dataset with the preexisting dataset to assess the efficacy of the latter. Consequently, the goal is to offer general and up-to-date knowledge in this sector.

Attack types including Mac flooding, DHCP spying, ARP spoofing, and others that are often employed in typical networks were not included in the data that was checked. One of the problems with the IDS material as it stands now is this. Future research is expected to include new attacks as well as the attacks used in many documents. However, while eliminating the nature of these attacks, it is unclear whether these attacks were attempted or not. It is considered that the above situation will harm the future performance of the brand.

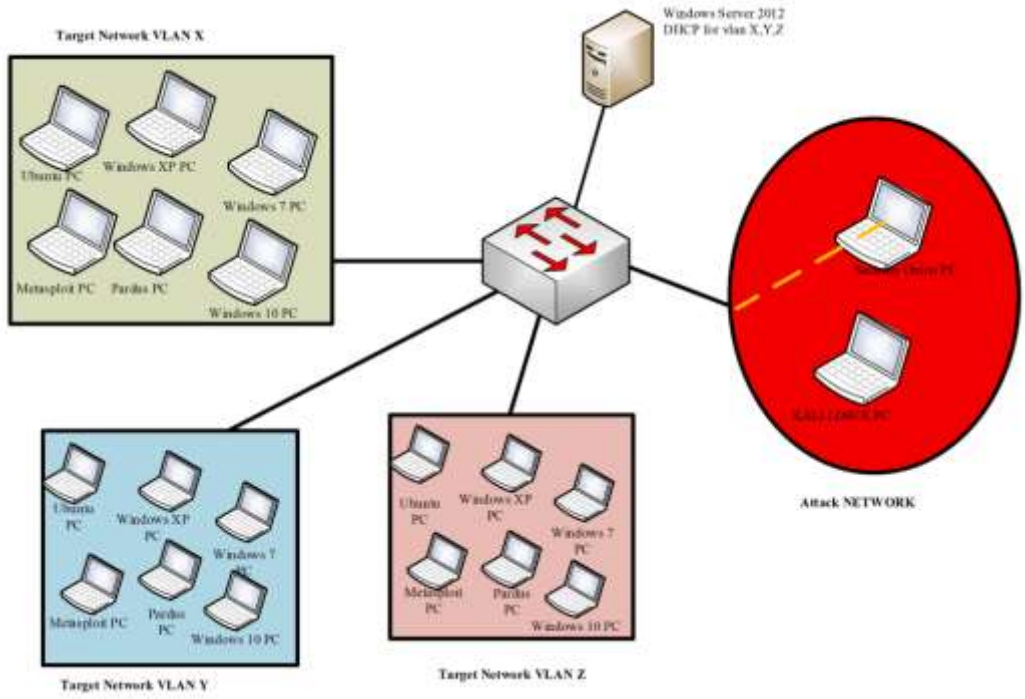


Figure 7.2.1: Planed Network Topology.

To advance the field of machine learning (ML) for cybersecurity, several avenues for future research are proposed. First, there is a need for developing more sophisticated ML algorithms that can detect complex and subtle cyber threats, potentially through the integration of deep learning. Privacy-preserving methods such as federated learning and differential privacy should be further explored to enhance the privacy capabilities of ML models while maintaining their efficacy. Addressing bias in ML algorithms is also critical; comprehensive studies are required to identify, measure, and mitigate bias, ensuring that all users receive equitable protection. Energy efficiency remains a significant concern; thus, research should focus on creating and optimizing ML models with lower energy consumption, utilizing specialized hardware, algorithmic improvements, and cloud computing resources. Finally, fostering interdisciplinary collaboration is essential; combining expertise from cybersecurity, machine learning, ethics, and environmental science can lead to holistic solutions that address the multifaceted challenges of cybersecurity. These efforts collectively aim to create robust, equitable, and sustainable ML-based cybersecurity measures that can adapt to the ever-evolving threat landscape.

7.3 Limitations

Despite the promising advancements in machine learning (ML) for cybersecurity, several limitations must be acknowledged. One significant limitation is the dependency on data availability; the effectiveness of ML models hinges on access to diverse and high-quality datasets, and limited data can impede these models' functionality and ability to be used generally. Furthermore, the use of ML models for real-time threat detection demands substantial computational resources, posing a challenge for organizations with limited infrastructure and budget constraints. The constantly evolving threat landscape further complicates matters, as static ML models may become obsolete quickly, necessitating regular updates and retraining, which can be resource-intensive. Navigating ethical and legal constraints, especially concerning data privacy and security regulations, presents another challenge. Ensuring compliance while maintaining model efficacy requires a careful balance. Finally, user adaptation is crucial; the successful implementation of ML-based cybersecurity measures relies on users being adequately trained and aware of the new systems. Without proper education, users may fail to fully leverage these technologies or may inadvertently compromise security. Addressing these limitations is essential for the continued advancement and practical application of ML in cybersecurity.

References:

[1] Li, S., Xu, L. D., & Zhao, S. (2014). The internet of things: a survey. Information Systems Frontiers, 17(2), 243–259. https://doi.org/10.1007/s10796-014-9492-7

[2] Sun, N., Zhang, J., Rimba, P., Gao, S., Xiang, Y., & Zhang, L. Y. (2018). Data-driven cybersecurity incident prediction: A survey. IEEE Communications Surveys & Tutorials, 1–1. https://doi.org/10.1109/comst.2018.2885561

[3] McIntosh, T., Jang-Jaccard, J., Watters, P., & Susnjak, T. (2019). The Inadequacy of Entropy-Based Ransomware Detection. Communications in Computer and Information Science, 181–189. https://doi.org/10.1007/978-3-030-36802-9_20

[4] Alazab, M.; Venkatraman, S.; Watters, P.; Alazab, M. Zero-day malware detection based on supervised learning algorithms of API call signatures. In Proceedings of the Ninth Australasian Data Mining Conference (AusDM'11), Ballarat, Australia, 1–2 December 2011.

[5] Shaw, A. Data breach: From notification to prevention using PCI DSS. Colum. JL Soc. Probs. 2009, 43, 517.

[6] Gupta, B. B., Tewari, A., Jain, A. K., & Agrawal, D. P. (2017). Fighting against phishing attacks: state of the art and future challenges. Neural Computing and Applications, 28(12), 3629–3654. https://doi.org/10.1007/s00521-016-2275-y

[7] Geer, D., Jardine, E., & Leverett, E. (2020). On market concentration and cybersecurity risk. Journal of Cyber Policy, 5(1), 9–29. https://doi.org/10.1080/23738871.2020.1728355

[8] Buecker, A.; Borrett, M.; Lorenz, C.; Powers, C. Introducing the IBM Security Framework and IBM Security Blueprint to Realize Business-Driven Security; International Technical Support Organization: Riyadh, Saudi Arabia, 2010.

[9] Fischer, E.A. Cybersecurity Issues and Challenges: In Brief; Library of Congress: Washington, DC, USA, 2014.

[10] Chernenko, E.; Demidov, O.; Lukyanov, F. Increasing International Cooperation

in Cybersecurity and Adapting Cyber Norms; Council

on Foreign Relations: New York, NY, USA, 2018.

[11] Papastergiou, S.; Mouratidis, H.; Kalogeraki, E.M. Cyber security incident handling, warning and response system for the

european critical information infrastructures (cybersane). In Proceedings of the International Conference on Engineering

Applications of Neural Networks, Crete, Greece, 24–26 May 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 476–487.

[12] Tolle, K. M., Tansley, D. S. W., & Hey, A. J. G. (2011). The Fourth Paradigm: Data-Intensive Scientific Discovery [Point of View]. Proceedings of the IEEE, 99(8), 1334–1337. https://doi.org/10.1109/jproc.2011.2155130

[13] Benioff, M. Data, data everywhere: A special report on managing information (pp. 21–55). The Economist, 27 February 2010.

[14] Anwar, S., Mohamad Zain, J., Zolkipli, M. F., Inayat, Z., Khan, S., Anthony, B., & Chang, V. (2017). From Intrusion Detection to an Intrusion Response System: Fundamentals, Requirements, and Future Directions. Algorithms, 10(2), 39. https://doi.org/10.3390/a10020039

[15] Saxe, J.; Sanders, H. Malware Data Science: Attack Detection and Attribution; No Starch Press: San Francisco, CA, USA, 2018.

[16] Rainie, L.; Anderson, J.; Connolly, J. Cyber Attacks Likely to Increase; Pew Research Center: Washington, DC, USA, 2014.

[17] Craigen, D.; Diakun-Thibault, N.; Purse, R. Technology Innovation Management Review Defining Cybersecurity; Technology Innovation

Management Review: Ottawa, ON, Canada, 2014.

[18] Hajisalem, V., & Babaie, S. (2018). A hybrid intrusion detection system based on

ABC-AFS algorithm for misuse and anomaly detection. Computer Networks, 136, 37–50. https://doi.org/10.1016/j.comnet.2018.02.028

[19] Intrusion response systems: Foundations, design, and challenges. (2016). Journal of Network and Computer Applications, 62, 53–74. https://doi.org/10.1016/j.jnca.2015.12.006

[20] García-Teodoro, P., Díaz-Verdejo, J., Maciá-Fernández, G., & Vázquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. Computers & Security, 28(1-2), 18–28. https://doi.org/10.1016/j.cose.2008.08.003

[21] Ferrag, M. A., Maglaras, L., Moschoyiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. Journal of Information Security and Applications, 50, 102419. https://doi.org/10.1016/j.jisa.2019.102419

[22] Sharafaldin, I., Habibi Lashkari, A., & Ghorbani, A. A. (2018). Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. Proceedings of the 4th International Conference on Information Systems Security and Privacy. https://doi.org/10.5220/0006639801080116

[23] Patil, R., Dudeja, H., & Modi, C. (2019). Designing an efficient security framework for detecting intrusions in virtual network of cloud computing. Computers & Security, 85, 402–422. https://doi.org/10.1016/j.cose.2019.05.016

[24] Kim, J., Shin, Y., & Choi, E. (2019). An Intrusion Detection Model based on a Convolutional Neural Network. Journal of Multimedia Information System, 6(4), 165–172. https://doi.org/10.33851/JMIS.2019.6.4.165

[25] Kanimozhi, V., & Jacob, Dr. T. P. (2019). CALIBRATION OF VARIOUS OPTIMIZED MACHINE LEARNING CLASSIFIERS IN NETWORK INTRUSION DETECTION SYSTEM ON THE REALISTIC CYBER DATASET CSE-CIC-IDS2018 USING CLOUD COMPUTING. International Journal of Engineering Applied Sciences and Technology, 04(06), 209–213. https://doi.org/10.33564/ijeast.2019.v04i06.036

[26] Khammassi, C., & Krichen, S. (2020). A NSGA2-LR wrapper approach for feature selection in network intrusion detection. Computer Networks, 172, 107183. https://doi.org/10.1016/j.comnet.2020.107183

[27] Ring, M., Wunderlich, S., Scheuring, D., Landes, D., & Hotho, A. (2019). A survey of network-based intrusion detection data sets. Computers & Security, 86, 147–167. https://doi.org/10.1016/j.cose.2019.06.005

[28] Kanimozhi, V., & Jacob, T. P. (2019). Artificial Intelligence based Network Intrusion Detection with hyper-parameter optimization tuning on the realistic cyber dataset CSE-CIC-IDS2018 using cloud computing. ICT Express. https://doi.org/10.1016/j.icte.2019.03.003

[29] Moustafa, N., & Slay, J. (2016). The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set. Information Security Journal: A Global Perspective, 25(1-3), 18–31. https://doi.org/10.1080/19393555.2015.1125974

[30] Moustafa, N., Slay, J., & Creech, G. (2017). Novel Geometric Area Analysis Technique for Anomaly Detection using Trapezoidal Area Estimation on Large-Scale Networks. IEEE Transactions on Big Data, 1–1. https://doi.org/10.1109/tbdata.2017.2715166

[31] Zhang, J., Ling, Y., Fu, X., Yang, X., Xiong, G., & Zhang, R. (2020). Model of the intrusion detection system based on the integration of spatial-temporal features. Computers & Security, 89, 101681. https://doi.org/10.1016/j.cose.2019.101681

[32] Gottwalt, F., Chang, E., & Dillon, T. (2019). CorrCorr: A feature selection method for multivariate correlation network anomaly detection techniques. Computers & Security, 83, 234–245. https://doi.org/10.1016/j.cose.2019.02.008

[33] Khammassi, C., & Krichen, S. (2017). A GA-LR wrapper approach for feature selection in network intrusion detection. Computers & Security, 70, 255–277. https://doi.org/10.1016/j.cose.2017.06.005

[34] Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic

Characterization DOI:10.5220/0006639801080116

Appendix A

Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

Title: MACHINE LEARNING APPROACHES FOR DETECTING AND MITIGATING CYBERSECURITY THREATS

Student ID: 201-15-13895

CO Description for FYDP

CO	CO Descriptions	PO
Phase -I		
CO1	Integrate recently gained and previously acquired knowledge to identify a **pneumonia detection** problem for the Final Year Design Project (FYDP)	**PO1**
CO2	Analyze different aspects of the goals in designing a solution for this FYDP	**PO2**
CO3	Explore diverse problem domains through a literature review, delineate the issues, and establish this goals for the FYDP	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP	**PO11**
Phase -II		
CO5	Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with public health and safety standards, as well as considering cultural, socioeconomic, and environmental factors in this FYDP	**PO3**
CO6	Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction and modeling, while adhering to relevant constraints in this FYDP	**PO5**
CO7	Analyze societal, health, safety, legal, and cultural considerations, along with associated responsibilities, in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding.	**PO6**
CO8	Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks.	**PO7**
CO9	Implement ethical principles and adhere to professional standards and norms in this FYDP	**PO8**
CO10	Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP.	**PO9**

CO11	Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design documentation, as well as provide and receive clear instructions throughout this FYDP.	**PO10**
CO12	Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology, and possess the readiness and capability to engage in lifelong learning endeavors.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	CO	Justification (with Knowledge Profile)	References
1.	**EP1: Depth of Knowledge required**	Yes	CO1, CO2, CO3, CO5, CO6, CO7 and CO8	This project demonstrates **fundamental engineering (K3)** principles by employing Machine Learning diverse SVM, KNN, DT architectures for data processing and classification tasks. The project demonstrates **specialist knowledge (K4)** by conducting transfer learning with advanced Machine Learning architectures like SVM and KNN, DT enhancing pneumonia detection accuracy in CSE-CIC-IDS2018 on AWS dataset	**Page no:** [14-17] **Page no:** [1-3]
				The project applies **engineering practice & design (K5)** by the figure of process of experiments. The project addresses **engineering practice & technology (K6)** by Machine Learning .	**Page no:** [17] **Page no:** [20]

				This project ensures to **K8 (Research Literature)** by synthesizing insights from recent studies, to advance pneumonia detection using deep learning, showcasing a comprehensive understanding of current methodologies.	**Page no:** [1-3]
2.	**EP2: Range of Conflicting Requirements**	Yes	CO2, and CO7	This project addresses **EP-2** by recognizing the hurdles in malware detection, including limitations of traditional methods and the complexities of integrating transfer learning and Machine Learning. Through comparative analysis, it confronts challenges in understanding spatial distributions, offering insights for refining diagnostic methodologies.	**Page no:** [4-5]
3.	**EP3: Depth of analysis required**	Yes	CO2, and CO6	This project addresses **EP-3** by meticulously comparing experimental outcomes, highlighting Transfer Learning as the chosen significant solution for enhancing threats detection multiple potential approaches.	**Page no:** [24-27]
4.	**EP4: Familiarity of Issues**	Yes	CO8	This project's interdisciplinary approach extends beyond computer science and engineering, impacting industry threats, contributing to advancements in Cyber Crime **EP-4**.	**Page no:** [9-11]
5.	**EP5: Extends of application codes**	No	CO5	N/A	N/A

6.	**EP6: Extends of stakeholders involved and conflicting .**	No	CO8	N/A	N/A
7.	**EP7: Interdependence**	Yes	CO5	This project's comprehensive approach addresses high-level problems by integrating various components across data collection, statistical analysis, and proposed methodology, ensuring a holistic solution to complex challenges in Cybersecurity which ensures **EP-7**.	**Page no:** [13-15]

Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:

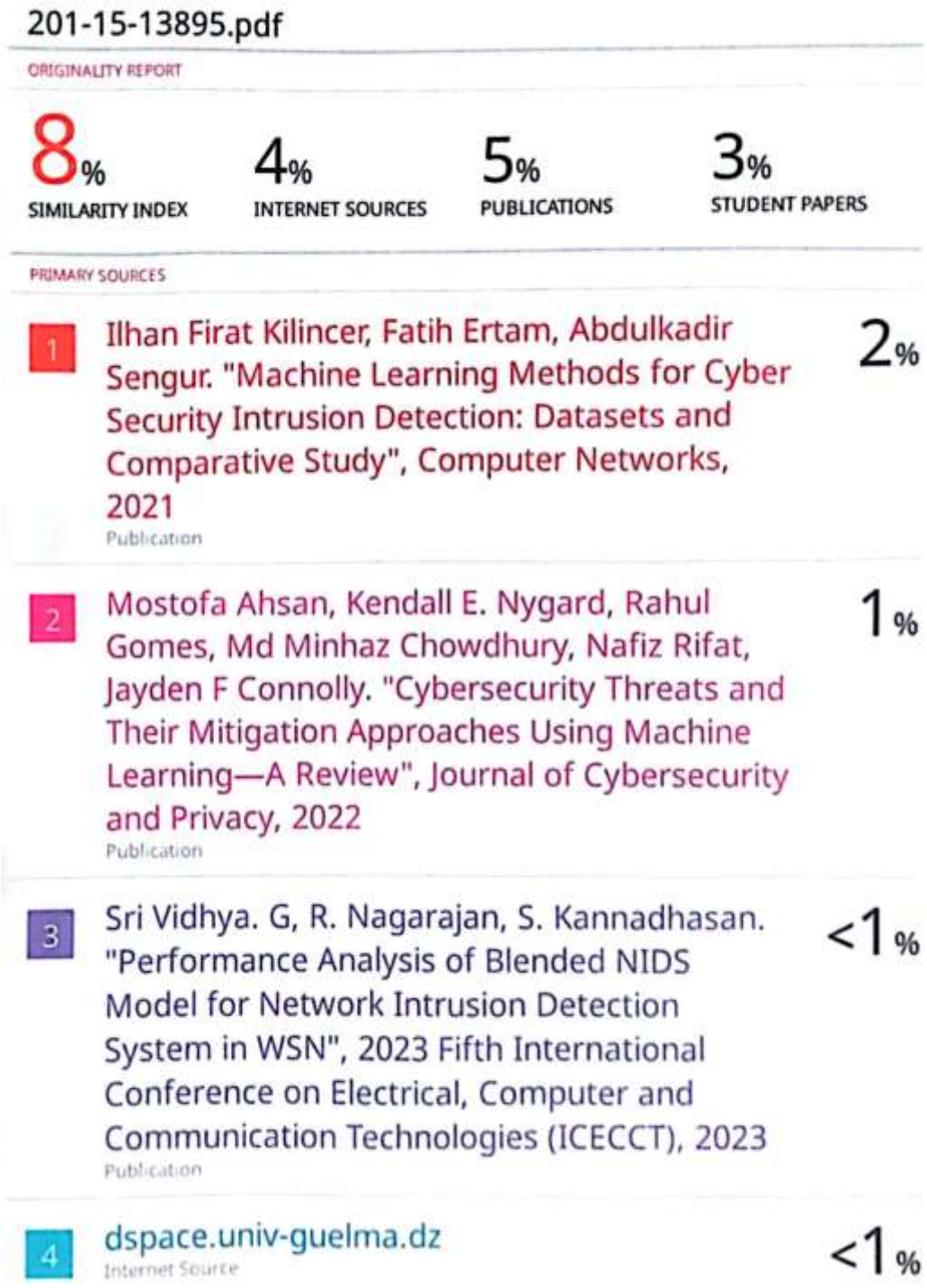
SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	Yes	CO11	Our project utilizes diverse resources such as high-performance computing infrastructure, GPUs, Machine learning frameworks, annotated datasets, and ethical considerations to ensure systematic research and contribute to advancements in threats detection through machine learning.	**Page no:** [17-19]
2.	**EA2: Level of interaction**	No		N/A	N/A

3.	**EA3: Innovation**	No		N/A	N/A
4.	**EA4: Consequences for society and the environment**	Yes		This project contributes to society by improving Machine Learning through advanced threats detection methods, while also promoting environmental sustainability by employing efficient computational resources and adhering to ethical guidelines for person's data privacy.	**Page no:** [28-29]
5.	**EA-5: Familiarity**	Yes		This project expands upon existing research by examining a novel approach in threats detection through machine learning demonstrated through preliminary terminologies and a comprehensive comparative analysis, offering new insights into the field.	**Page no:** [5-6]

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References
1	CO4	Yes	This project addresses **CO4** by integrating effective project management and financial oversight, ensuring meticulous planning, resource allocation, and budget estimation for optimal resource utilization throughout the research lifecycle.	**Page no:** [3-4]
2	CO9	Yes	The project demonstrates adherence to ethical principles by prioritizing patient privacy, obtaining informed consent, and transparently documenting the research process, ensuring responsible knowledge	**Page no:** [28-29]

			dissemination and societal well-being through the ethical application of advanced internet of things technologies which comply with **CO9**.	
3	CO10	No	N/A	N/A
4	CO12	Yes	The project's dedication to continuous learning **(CO12)** and adaptation within the dynamic technological landscape is reflected in its comprehensive data collection, rigorous statistical analysis, meticulous methodology development, and thorough experimental results and analysis, showcasing a commitment to staying updated and refining techniques to address modern challenges.	**Page no:** [14-17, 25-27]



201-15-13895.pdf

ORIGINALITY REPORT

8%	4%	5%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Ilhan Firat Kilincer, Fatih Ertam, Abdulkadir Sengur. "Machine Learning Methods for Cyber Security Intrusion Detection: Datasets and Comparative Study", Computer Networks, 2021 Publication	2%
2	Mostofa Ahsan, Kendall E. Nygard, Rahul Gomes, Md Minhaz Chowdhury, Nafiz Rifat, Jayden F Connolly. "Cybersecurity Threats and Their Mitigation Approaches Using Machine Learning—A Review", Journal of Cybersecurity and Privacy, 2022 Publication	1%
3	Sri Vidhya. G, R. Nagarajan, S. Kannadhasan. "Performance Analysis of Blended NIDS Model for Network Intrusion Detection System in WSN", 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2023 Publication	<1%
4	dspace.univ-guelma.dz Internet Source	<1%