AN ENHANCED MODEL FORINTRUSION DETECTION IN A CLOUD COMPUTING ENVIRONMENT

ABSTRACT

Intrusion is an important issue in computer networks especially in cloud computing where all the services are served using the internet. The fully distributed and open structure of cloud computing and services have made it an even more attractive target for potential intruders. The more sophisticated hackers and attackers get, the more there is work for the defense to prevent such attacks. A cloud computing system can be exposed to threats which include the integrity, confidentiality, and availability of its resources, its data, and the virtualized infrastructure can be vulnerable. The problem becomes bigger when an internal intruder misuses a cloud with massive computing power and storage capacity as a malicious party. This research developed an enhanced model for intrusion-detection that monitors and analyzes data in a cloud environment and detects intrusion in the system or network. The model can detect intrusions from external and malicious internal (authorized and unauthorized) users by normalizing and classifying all data packet using machine learning techniques. The developed system is an enhanced model of Zhang by combining it with two machine learning techniques: Support vector machine and Bayesian network to aid in the classification of normal data and intrusion data to detect intrusions. The developed model is evaluated and found to be able to make strong predictions, detect attacks, and still maintain the efficiency of the network. The system, when implemented, can detect intruders by classification of data packets and also improve the existing system in terms of providing more accurate and more efficient intrusion detection. It also provides worthwhile information about malicious network traffic, helping to identify the source of the incoming probes or attacks, collecting forensic evidence that can be used to identify intruders, and alerting security personnel that a network invasion maybe in progress.

Keywords: Intrusion detection, machine learning technique, Support Vector Machine, Bayesian Network,Cloud computing.

1. INTRODUCTION

Cyber security industry is one of the most growing industry today. Trends such as cloud computing, virtualization, and IoT (Internet of Things) have made data not only the most lucrative asset, but also the most vulnerable and easy to attack asset. As networks get larger, the attack surface and facilities becomes more prone to attacks from hackers thereby increasing cyber risk which is a prevalent problem. As hackers and attacks get sophisticated, the defense to prevent such attacks must be sophisticated as well.There are many common cyber solutions that exist today. The initial line of defence against a cyberattack is comprised of programmes like firewalls and honeypots, which we can discuss. Conversely, firewalls that filter network traffic using a set of rules stop hosts from connecting from outside the internal network to a secure end system. A persistent attacker can easily get past a firewall and penetrate the company network because of its inability to preserve state. Honeypots work as a kind of trap for hackers; by indicating that they might contain sensitive data, they draw them in and force them to try to access the honeypot before being barred from the network. However, honeypots are only successful if they can bait the attacker. Honeypots, however, are only effective if they can draw in the attacker. If the attacker learns that the honeypot is trying to trick them, they can ignore it and continue to attack the network. Thus, a demand has been developed for a system that can learn the structure of network data and discriminate regular from anomalous network traffic. In order to accomplish this, an improved model for cloud computing environment intrusion detection has been created. A system that monitors and analyses data to find any

intrusions into a system or network is called an intrusion detection system (IDS). There are two types of IDS: network-based IDS (NIDS) and host-based IDS (HIDS). A HIDS is a piece of software that operates on a host computer or a centralised controller to track file system access, confirm system call chains, etcAn NIDS, on the other hand, often operates on edge routers or switches and analyses traffic as it moves through a network. Trying to simulate the behaviour of both regular and aberrant traffic is the biggest obstacle that network-based detection systems have to overcome. The biggest development in technology today is machine learning, or ML. With the use of an ML model, complex issues without apparent solutions can be solved. In order for the model to produce accurate predictions, it will acquire the features of a dataset. Modern IDS research uses a wide range of ML approaches to investigate whether they may be applied in an IDS to secure enterprise networks.

2. RELATED WORKS

Sudhanshu and Bichitrananda(2023) discussed intrusion detection system using machine learning techniques using the KDD CUP '99' Intrusion detection dataset for training and validating machine learning models.

Alamin *et al* (2023) the authors introduced a hybrid machine learning model to enhance network intrusion detection by combining machine learning and deep learning to increase detection rates while securing dependability. Synthetic minority oversampling technology (SMOTE was used for data balancing and XGBoost for feature selection.

PanelChunying Zhang *et al.* (2022) The authors summarizes the application and research of machine learning in network intrusion detection systems from three categories: traditional machine learning, ensemble learning, and deep learning by comparing and analyzing some common machine learning algorithms in intrusion detection field in recent years. The current model has the challenges of how to preprocess when faced with different datasets.

Abdallah.E., *et al.* (2022) investigated the subject of intrusion detection using supervised machine learning algorithms methods based on a study of four popular data sets KDD'99, NSL-KDD, CICIDS2017, and UNSW-NB15 and a taxonomy for linked intrusion detection system was provided. However, data imbalance is still a major concern.

Adel *et al*. (2022) identified the power of various machaine learning (ML) algorithms and analysed the effect of ML algorithms for intrusion detection.

Stephen *et al.* (2022) the authors presented a review of hybrid deep learning models for network intrusion detection, its concepts, characteristics and A taxonomy of deep learning approaches was presented taking into account the deep networks for discriminative or supervised learning, generative or unsupervised learning, and finally hybrid learning that can be used to design a variety of Network intrusion detection systems.

Ayesha and Manivannan. (2021) presented a comprehensive survey of machine learning based approaches as presented in literatures for ten years which would serve as a supplement to other general surveys on intrusion detection as well as reference to recent work done in the area for researchers working in Machine Learning-based intrusion detection systems.

Kathryn *et al,* (2021) The authors presented an extensive overview, implementation, and cross comparison of state of the art machine learning based methods available for intrusion detection, analyzes some of the current state of the art intrusion detection methods and discusses their advantages and disadvantages. Four Machine learning algorithm was used to classify attacks to detect if traffic are benign or an attack.

Zeeshan *et al.* (2021) discussed the cyber security technology trends in intrusion detection utilizing ML (Machine Learning) and DL (Deep Learning) methods. However, the present work does not cover all the methods in the intrusion detection domain; furthermore, the authors use few benchmark datasets for the

model, and the analysis is not uniform. None of the work covers a deep and insightful analysis of the performance of the model.

Zhang *et al*. (2019). Proposed a system to detect intrusion using Deep Generative Neural Network (DGNN) that performs Adversarial learning with Data Augmentation in intrusion detection. With the use of data augmentation in intrusion detection, the detection rate and precision was better. But when there is not enough test data for the system to train with, the system experiences data scarcity and imbalance.

Enamul *et al* (2018) proposes a novel approach for intrusion detection system based on sampling with Least Square Support Vector Machine (LS-SVM). The authors discussed a new algorithm of the optimum allocation-based least square support vector machine (OA-LS-SVM) for IDS that can be used both for static and incremental data.

3. MATERIALS AND METHODS

3.1 DATA SOURCE AND COLLECTION

The dataset used in this paper was obtained by the use of a data packet receiver, indicating the time, source, destination, protocol length of the packet and other information and this was used for the analysis. This dataset consists of features and instances. The features i.e., class values areNormal Data (ND) or Intrusion Data (ID).

3.2 EXPERIMENTAL SET UP

Data were collected and moved to the Data pre-processing stage, where they were normalized, to reduce the height and volume of the data, and then they were classified into Normal Data (ND) or Intrusion Data (ID) using the two machine learning techniques (support vector machine and Bayesian network). Both the Normal Data and Intrusion data are sent to the next stage which is the Data Partition where the Normal Data were now classified as Normal Network Request while the Intrusion data were classified as Network Intrusion. The Normal Network request was sent through the Date management and Data aggregation process to the shadow learning, while the Network Intrusion (Intrusion Data) went through the Data Augmentation process of Zhang (Zhang 2019) and finally to the Deep Learning Training Model. The above description constitutes the Training Phase for the models.

3.3 EXPERIMENTAL TOOLS

All the experiments carried out are computed using an open-source python library and Python programming language with Jupyter Notebook IDE. PHP (Pre Hypertext Processor), JavaScript, HTML and CSS (Cascading Style Sheet) was used to code. A text editor called notepad++ and WAMP Server 2.1 which runs on Apache Server 2.2.17, MySQL 5.5.8 and PHP 5.3.5 interpreter was employed to execute incremental development while actual coding was done on Notepad++, local deployment of code functionality was carried out on WAMP Server 2.1.

3.4 MODEL SELECTION

Zhang, (2019) model was adopted and enhanced by combining two machine learning techniques (support vector machine and Bayesian network) to aid in the classification of normal data and intrusion data.

3.5 SYSTEM DESIGN

The new system adopted the Zhang model and enhanced it by combining two machine learning techniques (support vector machine and Bayesian network) to aid in the classification of normal data and intrusion data. From figure 1. Data are collected and moved to the Data pre-processing stage, where they are normalized, to reduce the height and volume of the data, and then they are classified into Normal Data (ND) or Intrusion Data (ID) using the two machine learning technique (support vector machine and

Bayesian network). Both the Normal Data and Intrusion data are sent to the next stage which is the Data Partition where the Normal Data are now classified as Normal Network Request while the Intrusion data are classified as Network Intrusion. The Normal Network request are sent through the Date management and Data aggregation process to the shadow learning, while the Network Intrusion (Intrusion Data) go through Data Augmentation process of Zhang (Zhang 2019) and finally to the Deep learning Training Model. The above description constitute the Training Phase for the models.

The testing phase include feeding the model with test data, which are then normalize and classified and a result of weather they are intrusion Data (ID) or Normal Data (ND) is made based on the result from the shadow learning process or Deep learning process of the trained model.

3.5.1 MODEL OF THE SYSTEM

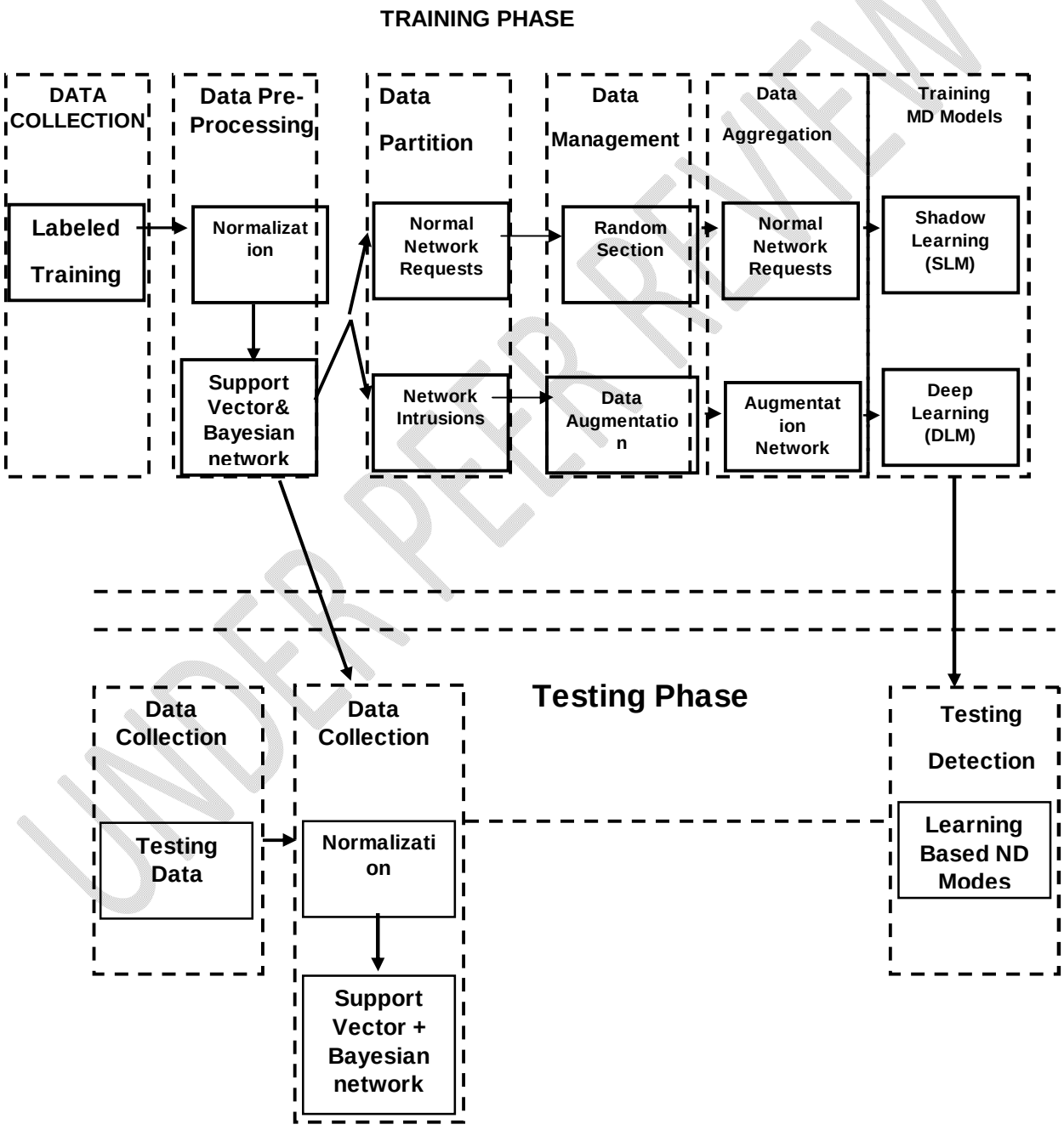

Figure 1. Model View of the New System

3.6. HIGH LEVEL MODEL OF THE NEW SYSTEM

The high level model explains the architecture that would be used for developing the automated system. The high level diagram shown in figure 2 provides an overview of the entire system, identifying the main components that would be developed and their interfaces.

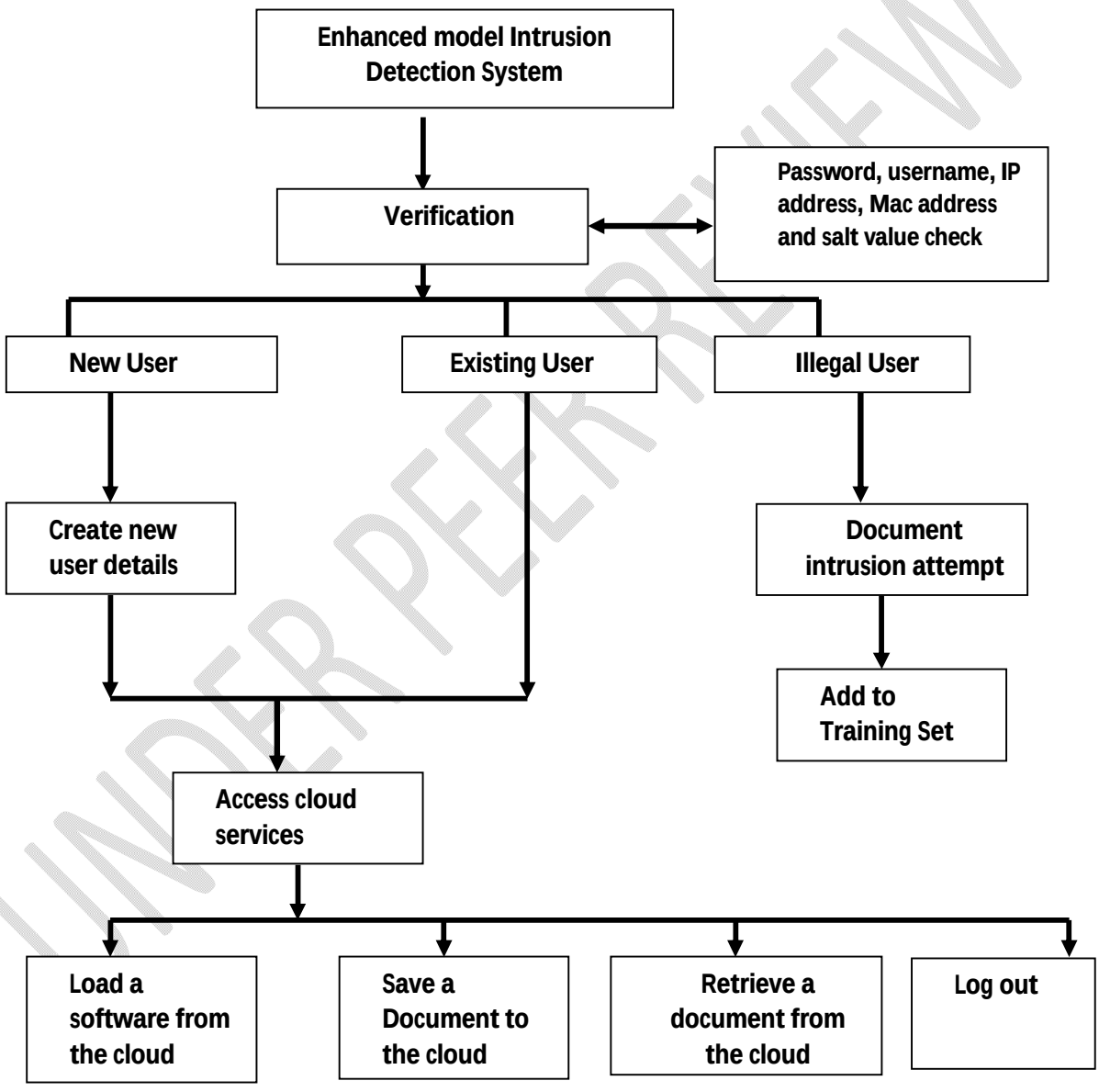

Figure 2. High level Model view of the new system

4. RESULTS AND DISCUSSION

A test data was obtained with the use of data packet receiver and classification of the intrusions as shown in table 1.The result gotten from the classification of the intrusions was used to test for true positive and true negative as shown in table 2. And from this, a comparison/performance evaluation was done to ascertain if the enhanced intrusion detection model was better and why it was better than the existing one as shown in table 3.

Table 1 Classification of the intrusions used to test for true positive and true negative

TRUE POSITVE **Reality:** An intrusion Attack occurs **Enhanced IDS:** Detects an Attack occurs **Output:** Record the attack as TP	**FALSE POSITVE** **Reality:** No intrusion Attack occurs **Enhanced IDS:** Detects an Attack occurs **Output:** Record the attack as FP
FALSE NEGATIVE **Reality:** An intrusion Attack occurs **Enhanced IDS:** Does not Detects an Attack **Output:** Record the attack as FN	**TRUE NEGATIVE** **Reality:** No intrusion Attack occurs **Enhanced IDS:** Does not Detects an Attack **Output:** Record the attack as TN

Table 2. Showing the Expected and Actual result

ID	frame	byte	Result
1	Frame 1:	208	2.05807365439093
2	Frame 2:	42	1
3	Frame 3:	42	1
4	Frame 4:	208	2.05807365439093
5	Frame 5:	208	2.05807365439093
6	Frame 6:	208	2.05807365439093
7	Frame 7:	208	2.05807365439093
8	Frame 8:	208	2.05807365439093
9	Frame 9:	208	2.05807365439093
10	Frame 10:	208	2.05807365439093

ID	frame	byte	Result
11	Frame 11:	208	2.05807365439093
12	Frame 12:	92	1.31869688385269
13	Frame 13:	92	1.31869688385269
14	Frame 14:	208	2.05807365439093
15	Frame 15:	92	1.31869688385269
16	Frame 16:	80	1.24220963172805
17	Frame 17:	158	1.73937677053824
18	Frame 18:	66	1.15297450424929
19	Frame 19:	66	1.15297450424929
20	Frame 20:	54	1.07648725212465
21	Frame 21:	54	1.07648725212465
22	Frame 22:	72	1.19121813031161

4.1. PERFORMANCE EVALUATION

Table 3. Performance evaluation carried out on the system

Number of Packet Frame Analyzed	Number of True Positive	Number of True Negative
12362	1908	10454
6120	954	5166
12240	1908	10332
6006	945	5061
5945	945	5000
5823	943	4880
11829	1888	9941

From table 3. It shows that the new enhanced intrusion detection system was more reliable than that of Zhang, as seen in the high rate of true positive and true negative value.

5. CONCLUSION

Networks security problems vary widely and can affect different security requirements including authentication, integrity, authorization, and availability. Intruders can cause different types of attacks on systems in an organization. These attacks need to be detected as soon as possible to prevent further damages to organizations sensitive data which may cause financial loss. This modelbrought about a new method of detecting intruders by adopting the Zhang model and enhancing with the combination of two machine learning techniques to aid in the classification of normal and intrusion data. The hybridaztion of these models enhanced the system by providing a better result in terms of accuracy for intrusion detection. It also provided worthwhile information about malicious network traffic; helping to identify the source of the incoming probes or attacks; collecting forensic evidence that can be used to identify intruders and alerting security personnel that a network invasion maybe in progress.

Disclaimer (Artificial intelligence)

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of manuscripts.

Option 2:

Author(s) hereby declare that generative AI technologies such as Large Language Models, etc have been used during writing or editing of manuscripts. This explanation will include the name, version, model, and source of the generative AI technology and as well as all input prompts provided to the generative AI technology

Details of the AI usage are given below:

1.

2.

3.

REFERENCES

Adel *et al*. (2022) identified the power of various machaine learning (ML) algorithms and analysed the effect of ML algorithms for intrusion detection.

Abdallah.E., *et al.* (2022) investigated the subject of intrusion detection using supervised machine learning algorithms methods based on a study of four popular data sets KDD'99, NSL-KDD, CICIDS2017, and UNSW-NB15 and a taxonomy for linked intrusion detection system was provided. However, data imbalance is still a major concern.

Alamin *et al* (2023) the authors introduced a hybrid machine learning model to enhance network intrusion detection by combining machine learning and deep learning to increase detection rates while securing dependability. Synthetic minority oversampling technology (SMOTE was used for data balancing and XGBoost for feature selection.

Ayesha &Manivannan. (2021) presented a comprehensive survey of machine learning based approaches as presented in literatures for ten years which would serve as a supplement to other general surveys on intrusion detection as well as reference to recent work done in the area for researchers working in Machine Learning-based intrusion detection systems.

Enamul *et al* (2018) proposes a novel approach for intrusion detection system based on sampling with Least Square Support Vector Machine (LS-SVM). The authors discussed a new algorithm of the optimum allocation-based least square support vector machine (OA-LS-SVM) for IDS that can be used both for static and incremental data.

Kathryn *et al,* (2021) The authors presented an extensive overview, implementation, and cross comparison of state of the art machine learning based methods available for intrusion detection, analyzes some of the current state of the art intrusion detection methods and discusses their advantages and disadvantages. Four Machine learning algorithm was used to classify attacks to detect if traffic are benign or an attack.

PanelChunying Zhang *et al.* (2022) The authors summarizes the application and research of machine learning in network intrusion detection systems from three categories: traditional machine learning, ensemble learning, and deep learning by comparing and analyzing some common machine learning algorithms in intrusion detection field in recent years. The current model has the challenges of how to preprocess when faced with different datasets.

Stephen *et al.* (2022) the authors presented a review of hybrid deep learning models for network intrusion detection, its concepts, characteristics and A taxonomy of deep learning approaches was presented taking into account the deep networks for discriminative or supervised learning, generative or unsupervised learning, and finally hybrid learning that can be used to design a variety of Network intrusion detection systems.

Sudhanshu and Bichitrananda (2023) discussed intrusion detection system using machine learning techniques with the use of KDD CUP '99' Intrusion detection dataset for training and validating machine learning models.

Zhang He, Xingrui Yu, Peng Ren, Chumbo Luo, Geyong Min (2019). Deep Adversial learning in intrusion detection. A data Augmentation EhancedFramework.ArXiv: 1901.07949v3 [CS.CR]

Zeeshan *et al.* (2021) discussed the cyber security technology trends in intrusion detection utilizing ML (Machine Learning) and DL (Deep Learning) methods. However, the present work does not cover all the methods in the intrusion detection domain; furthermore, the authors use few benchmark datasets for the model, and the analysis is not uniform. None of the work covers a deep and insightful analysis of the performance of the model.