2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA) | 978-1-6654-2829-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICAECA52838.2021.9675595

# Intrusion Detection and Prevention in Networks Using Machine Learning and Deep Learning Approaches: A Review

1st Jitti Annie Abraham
Research Scholar, SOCS
Mahatma Gandhi University
Kottayam, Kerala, India
jittiannie265@gmail.com

2nd Bindu V. R.
Professor, SOCS
Mahatma Gandhi University
Kottayam, Kerala, India
binduvr@mgu.ac.in

***Abstract*— As network services are used more and more, security is considered one of the main and important concerns of the network. Many computers connected to the network play important roles in business and other applications that provide services over the network. Thus we must seek out the most effective ways to safeguard the system. The Intrusion Detection System (IDS) is one of the most important information security technologies that use machine learning and deep learning algorithms to detect network anomalies. The accuracy of an intrusion detection system determines its performance. To reduce false alarms and increase detection rates, intrusion detection accuracy must be improved. The main function of an intrusion detection system is to analyze huge amounts of network traffic data. The main objective of this paper is to survey in-depth learning and machine learning methods for intrusion detection by reviewing literature and providing background information on either deep learning or machine learning algorithms on intrusion detection systems. The study also includes a performance comparison of different machine learning classification methods on the DARPA dataset.**

***Keywords—accuracy, deep learning, intrusion detection system, intrusion prevention system, machine learning***

## I. Introduction

Network techniques and technologies have recently evolved rapidly, hence why networks are growing rapidly both in terms of performance and total traffic volume. The progress made in network and related service technology has brought significant growth in data traffic. Along with thus the harmful repercussions of cyber-attacks have also increased. However, the damage resulting from cyber-attacks increases, not only because of the increase in the number of cyber-attacks and just because of that the attackers are skilled and their methods are much more developed. Nowadays, defending a network entirely from malicious hackers is quite impossible [1]. Cyber security is some of methods and processes advanced to shield computer systems, networks, programs, and information from assaults and unauthorized get admission to, change, or destruction. The network security systems includes a computer security system that includes a network security system and firewalls, anti-virus software, and intrusion detection systems [2].

Intrusion Detection and Prevention System (IDPS) is a system or software application that monitors network or system activity and detects malicious activity and provides a timely response. IDS can gather detailed information on the most common types of attacks as well as the attackers. The intrusion detection technique can detect a wide range of attacks and their prevention will result in intrusions being blocked. Although rule-based techniques are easy to implement and quick, they cannot replace enough or noisy data and are difficult to update. To solve these problems, statistically based systems for processing imprecise information have been proposed, but such methods have a high computing costs and limited data handling capacity.

Recently, Machine Learning (ML) based techniques have gained prominence due to their capacity to detect complicated intrusion patterns by using complex inference models trained on larger datasets [3]. Machine learning is a class of methods that become more accurate in predicting results without explicitly programming software applications. Model is an algorithm that applies to any data. Deep learning (DL) is a machine learning process that involves many layers of layers, many of which are capable of conducting processes concurrently and producing high-level characteristics from low-level ones. The creation of a neural network that simulates the human brain for quantitative learning is the fundamental basis for DL. It accepts data such as images, audio, and texts using a system that is similar to the human brain. Deep learning provides a short training period and a high accuracy rate using a unique learning method for BigData [4]. As a result, the implementation of deep learning based IDPS systems has emerged.

The following section is positioned out as follows. The survey on machine learning and deep learning approaches for network intrusion detection and prevention system is given in Section II. This section looks at a few papers that contribute to the knowledge of machine learning and deep learning techniques, as well as intrusion attack detection, analysis, and prevention. The classification algorithms and intrusion datasets used are discussed in Section III. The accuracy of machine learning algorithms is presented in Section IV. Finally, in Section V, we bring our paper to an end.

## II. Related Works

The following are the broad concepts of relevant work for network IDPS according to machine learning and deep learning approaches. The rapid use of Internet services in all

978-1-6654-2829-3/21/$31.00 ©2021 IEEE

industries around the world in recent decades has led to rapid advances in technology and networks. Counterfeiting has proliferated, and many modern systems had compromised, therefore building information security solutions to identify new attacks has become critical.

M. Mithem, Et. Al used an innovative intrusion detection system with excellent network performance to detect unknown attack packages detected using deep neural network [4]. Binary classification and multiclass classification are two methods for detecting attacks. In terms of high accuracy, the proposed approaches had promising results.

Because of big volume of data is exchanged every hour in each domain, security is a major problem these days. The work in [5] explains how to use a neural network to prevent and detect intrusions for data security. In this study, various IDS and Intrusion Prevention Systems (IPS) were studied and compared. Furthermore, a comparative examination of several approaches is conducted. Here also various intrusion detection models and techniques are discussed. The study demonstrates how neural networks can be used to solve related problems.

The study [6] discusses the differences between IDS and IPS, as well as the advantages and disadvantages of using IDSs, IPSs, and hybrids. It also explains current developments in machine learning and how it is being used to improve these systems, as well as the problems they cause and potential solutions.

The hybrid IDS proposed in [7] is the combination of two J48 DT and SVM learning machines. On KDD CUP data set the Particle Swarm Optimization (PSO) technique is applied for feature selection. WEKA is used for implementing the KDD CUP dataset classification. The data settings for training and testing purposes are divided by 60:40, 70:30, and 80:20. According to the findings, the presented method has a high detection accuracy while also having a low false alarm rate.. The main drawback is that the system's training time was not compared with other hybrid approaches.

The work [8] aims to implement deep learning-based intrusion detection and prevention methods can detect and stop attacks like DOS, R2L, and U2R in real time. An in-depth learning model, which is a multi-level comprehension trained with high precision in the kddcup99 dataset, was used to visualise the intrusion. The appropriate network data is collected and saved as a CSV file in the Display Deep Learning model, which detects the result, for real-time estimation. Background scripts prevent the intrusion in the second step. The script's goal is to complete the prevention phase by recommending various preventive measures for various types of attacks. The data collected by the Multi-Layer Perceptron classification part may be used to make the decision. For faster, more effective intrusion detection and prevention, specialized intrusion detection and intrusion prevention systems are integrated into a single system.

The work in [9] proposes a multifunctional modular deep neural network model to reduce the false-positive rate of anomaly-based intrusion detection systems. The model is made up of three modules: a feed forward module, a limited Boltzmann machine module, and two repeatable modules, the output weights of which are fed into the aggregator module to generate the model answer. Experiments are conducted using the CSE-CICIDS2018 dataset, and the final model can be used in intrusion detection systems to generate alerts and prevent new attacks. When compared to previous work, experimentalresults show an improvement in detecting certain types of network attacks, with network-level attacks having up to 100% accuracy.

The study [10] explains IDS and then presents a categorization based on the most common methods in the development of Network-based IDS (NIDS) systems, machine learning and deep learning approaches. A comprehensive review of current NIDS-based studies is presented, with an emphasis on proposed solutions, strengths, and limitations. Following a discussion of recent trends and improvements in ML and DL-based NIDS the proposed method, evaluation criteria, and dataset selection are discussed. Many research difficulties, as well as the future potential for investigation in strengthening ML and DL-based NIDS, were highlighted using the shortcomings of the presented approaches.

According to the research [11], current network intrusion, detection, and prevention systems (NIDPSs) have a number of flaws in terms of detecting or blocking growing unwanted traffic, as well as several dangers in high-speed environments. To improve intrusion detection and prevention performance, a unique quality of service (QoS) architecture is created.

After completing the study of related works, it is planned to attempt ML-DL approaches towards network intrusion detection and prevention techniques. The following section covers the implementation and analysis of some machine learning classification algorithms.

## III. Classification Algorithms and Dataset

This section examines the experimental performance of various machine learning classification algorithms on the intrusion set dataset named DARPA data set. The two most often used machine learning approaches are supervised and unsupervised learning. For training algorithms, labelled examples such as an input with a chosen output are used. Unsupervised learning is used to train instances that have no pre-existing labels. The two goals of unsupervised learning are to find some structure in the data and to explore the data. In addition to these methods, semi-supervised learning and reinforcement learning are performed [12].

### *A. Machine Learning Classification Algorithms*

The process of classifying object involves recognizing, comprehending, and grouping them into pre-determined categories or "sub-populations." The ML algorithms use pre-categorized training datasets and a variety of algorithms to classify future datasets. Classification algorithms in machine learning use training data to predict whether new data will fall into one of the descriptive data. In a nutshell, classification is a kind of 'pattern recognition,' which uses classification methods in training data to detect a pattern within the data sets [13]. The classification algorithms that were employed here for identification and classification of intrusive attacks are:

i. AdaBoost
ii. Extra Trees
iii. Gradient Boost
iv. Linear Regression
v. Multilayer Perceptron
vi. Random Forest

AdaBoost is an abbreviation for Adaptive Boosting. Ada Boosting was, in theory, the first truly successful boosting technique created for binary classification. Furthermore, current boosting methods, most likely stochastic gradient boosting methods, are based on AdaBoost. AdaBoost is typically used with small decision trees which following the creation of the first tree, the performance of the tree on each training instance is used.

Extra trees are different from standard decision trees in the way they are built. Random sections are generated for each of the maximum randomly picked features, and the best split among them is favoured when deciding on the optimal split to divide a node's sample data into numerous groups. When features max is set to 1, this results in the construction of a completely random decision tree. Extra-trees should only be used in combination with ensemble approaches. The Extra-Tree approach was proposed with the primary goal of further randomising tree generation in the presence of numerical input data, when the optimal cut-point selection explains for a considerable part of the induced tree's variance.

Gradient boosting is a machine learning technique that generates a predictor for regression and classification problems by combining weak prediction models, such as decision trees. It works in a similar way to previous boosting approaches in that it develops the model step by step and then generalises it by permitting the optimization of any differentiable loss function. Gradient boosting is a valuable approach for building predictive models.

Regression is a method that use independent predictors to predict a target value. This method is commonly used for forecasting and determining cause-and-effect relationships. The number of independent variables and the type of link between the independent and dependent variables are the most important distinctions in regression processes. Linear regression is one of the most well-known and well-understood algorithms in statistics and machine learning.

Linear regression is a supervised learning-based machine learning technique. It carries out a regression task, in which it uses independent variables to model the desired prediction value. It is mostly utilized in forecasting and determining the relationship among variables.

Many perceptron constitute a multilayer perceptron (MLP). An input layer receives the signal, an output layer makes a judgement or makes a prediction about the input, and the MLP's true computational engine is an arbitrary number of hidden layers. Using MLPs and a single hidden layer, any continuous function may be approximated. The multilayer perceptron is often used to solve supervised learning challenges.

Random forests, also known as random decision forests or random forests, are a popular ensemble method for modelling classification and regression issues. To improve prediction results, ensemble techniques employ several learning models. In the case of a random forest, for example, the model creates a forest of random uncorrelated decision trees to arrive at the best possible result.

### *B. DARPA Dataset*

The "DARPA" dataset was used to lead the test. In 1998, the MIT Lincoln Laboratory generated a set of raw PCAP files that includes full, training, and test sets. The DARPA datasets DARPA 1999 and DARPA 2000 are newer versions based on the 1998 edition. One of the most extensively used intrusion detection datasets is this one, nonetheless, it is widely regarded as obsolete and including inconsistencies [14]. DARPA IDS assessment dataset is valuable for testing interruption discovery frameworks in that great execution against it is a fundamental yet not adequate condition to showing the capacities of a propelled IDS.

This dataset was built for system security examination purposes. Analysts scrutinized DARPA because of issues related to the counterfeit infusion of assaults and benevolent traffic. DARPA incorporates exercises, for example, send and get mail, peruse sites, send and get documents utilizing FTP, the utilization of telnet to sign into remote PCs and perform work, send and get IRC messages, and screen the switch remotely utilizing SNMP [15].

The following are some of the attack categories identified in the DARPA dataset [16]:

- Denial of Service (DoS): DoS is a type of intrusion assault that involves overloading network resources and making them unavailable to legitimate users.
- User to Root (U2R): This is an intrusion attack that compromises the user's legitimacy by granting the invader root access.
- Remote to Local (R2L): It is an intrusion attempt that occurs when the network's integrity is compromised, allowing the attacker access to the local network.
- Probe: A probe is an intrusion activity that involves scanning the network and collecting all network-related information concerning network activities.

## IV. EXPERIMENTAL DISCUSSION

Researchers have contributed to machine learning, and several techniques have been implemented. There are six algorithms used in this work, and they are AdaBoost, Extra Trees, Gradient Boost, Linear Regression, Multilayer Perceptron, and Random Forest. These algorithms are described in the preceding section.

TABLE 1: Accuracy Comparison between Different Machine Learning Classification Algorithms

| Algorithm | Accuracy | Training Time(sec.) | Prediction Time(sec.) |
|---|---|---|---|
| **AdaBoost** | 75.37 | 0.0064 | 0.0030 |
| **Extra Trees** | 76.12 | 0.0092 | 0.0010 |
| **Gradient Boost** | 81.73 | 0.0379 | 0.0009 |
| **Linear Regression** | 60.52 | 1.825 | 0.0020 |
| **MLP** | 74.58 | 0.1011 | 0.0010 |
| **Random Forest** | 78.47 | 0.0019 | 0.0058 |

The DARPA dataset has now been used to compare the performance of different algorithms. There are 41 features in total, divided into four classes: DoS, R2L, U2R, and Probe. Python software is used to pre-process and label the data. All of these algorithms were offered the same information. On the basis of accuracy, training time, and prediction time, the algorithms were compared. The above table TABLE 1 shows the experimental results.

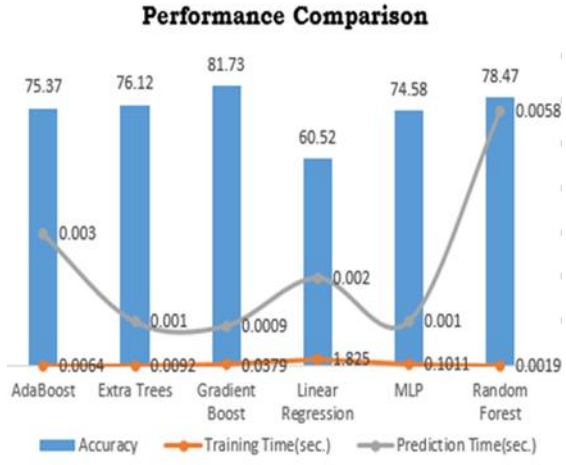


Fig. 1. Intrusion Detection Accuracy of Machine Learning Algorithms

Based on the results in Fig. 1, it is clear that the Gradient Boost classification has a high accuracy measurement performance. However, the performance of algorithms is generally determined by the domain and data set to which they are applied. Under certain circumstances, machine-learning algorithms may outperform each other in terms of performance.

## V. Conclusion

In the cybersecurity domain, an intrusion detection and prevention system is critical for preventing network attacks. Due to continuous change in behavioural patterns of attacks, the traditional statistical methods get not enough to detect and prevent intrusions. Thus deep learning and machine learning approaches have evolved as most recent training and classification methods. This paper reviews research on foundational machine learning and deep learning methods for intrusion detection and prevention systems. On the DARPA dataset, some machine learning classification algorithms have also been tested experimentally.

Since some machine learning algorithms have low accuracy levels, future works need to deal with some deep learning methods or hybrid approaches. Furthermore, the data set used here is limited and does not reflect real-world network activity, the absence of false-positive cases, and anomalies in attack data. Therefore, the DARPA dataset should be replaced with either KDDCup99 or NSL KDD datasets.

## References


[1] Wooseok, Wooguil Pak, “Real-time Network Intrusion Prevention System Based on Hybrid Machine Learning”, DOI: 10.1109/ACCESS.2021.3066620, IEEE Access, Volume 9, 2021J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Anish Halimaa A, Dr. K.Sundarakantham, “Machine Learning-Based Intrusion Detection System”, Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019), 978-1-5386-9439-8/19/$31.00 ©2019, IEEE, 2019.

[3] Gozde Karatas, Onder Demir, Ozgur Koray Sahingoz, “Deep Learning in Intrusion Detection Systems”, International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism Ankara, Turkey, 3-4 Dec, 2018, 978-1-7281-0472-0/18/$31.00 ©IEEE, 2018.

[4] Mohammed Maithem, Dr.Ghadaa A. Al-sultany, "Network Intrusion Detection System Using Deep Neural Networks”, doi:10.1088/1742-6596/1804/1/012138, Journal of Physics: Conference Series, ICMAICT 2020.

[5] Vaishali Bhatia, Shabnam Choudhary, K.R Ramkumar, “A Comparative Study on Various Intrusion Detection Techniques Using Machine Learning and Neural Network”, 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) Amity University, Noida, India. June 4-5, 2020, 978-1- 7281-7016-9/20/$31.00 ©IEEE, 2020.

[6] Keturahlee Coulibaly, “An overview of Intrusion Detection and Prevention Systems”, Research Gate, April 2020.

[7] Anku Kumari, Ashok Kumar Mehta, “A Hybrid Intrusion Detection System Based on Decision Tree and Support Vector Machine”, 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), Galgotias University, Greater Noida, UP, India. Oct 30-31, 2020, IEEE, 2020.

[8] Akhil Krishna, Hari M, Et.Al “Intrusion Detection and Prevention System Using Deep Learning”, Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020), IEEE, 2020.

[9] Ramin Atefini, Mahmood Ahmadi, “Network Intrusion Detection Using Multi Architectural Modular Deep Neural Network”, The Journal of Supercomputing https://doi.org/10.1007/s11227-02003410-y, Springer, August 2020.

[10] Adnan Shahid Khan,Et.Al “Network Intrusion Detection System: A Systematic Study Of Machine Learning And Deep Learning Approaches”, Transactions on Emerging Telecommunications Technologies, DOI: 10.1002/ett.4150 ©WILEY, 2020.

[11] Waleed Bul'ajoul, Anne James,Siraj Shaikh “A New Architecture for Network Intrusion Detection and Prevention”, DOI: 10.1109/ACCESS.2019.2895898, IEEE Access, Volume 7, 2019.

[12] Aditya Phadke, Mohit Kulkarni,Pranav Bhawalkar and Rashmi Bhattad, A Review of Machine Learning Methodologies for Network Intrusion Detection, Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019), 978-1-5386-7808-4/19/$31.00 ©2019 IEEE.

[13] Venkata Ramani Varanasi, Shaik Razia,” Intrusion Detection using Machine Learning and Deep Learning”, ISSN: 2277-3878, Volume-8 Issue-40, IJRTE, November 2019.

[14] Dilara Gümü¸sba¸s, Tulay Yıldırım,” A Comprehensive Survey of Databases and Deep Learning Methods for Cybersecurity and Intrusion Detection Systems” IEEE SYSTEMS JOURNAL, IEEE 2020.

[15] Markus Ring, Sarah Wunderlich, Deniz Scheuring, “Survey of Network-based Intrusion Detection Data Sets”, 1903.02460v2 [cs.CR] 6 Jul 2019.

[16] Cunningham RK, Lippmann RP, Fried DJ, Garfinkel SL, Graf I, Kendall KR, et al. Evaluating intrusion detection systems without attacking your friends: The 1998 DARPA intrusion detection evaluation. Massachusetts Institute of Technology Lexington Lincoln Lab; 1999. Communication Control and automation (ICCUBEA) (pp. 1-7).

[17] Jitti Annie Abraham, Susan M George. "A Survey on Preventing Crypto Ransomware Using Machine Learning", 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), 2019.

[18] Zeeshan Ahmad, Adnan Shahid Khan, Cheah Wai Shiang, Johari Abdullah, Farhan Ahmad. "Network intrusion detection system: A systematic study of machine learning and deep learning approaches", Transactions on Emerging Telecommunications Technologies, 2020.