Intesab Aalami
ISSN: 2348-5035
Vol. 13 / No. 1 / January 2025
UGC Approved Sr. No. 597 (UGC - CARE)

# MACHINE LEARNING FOR NETWORK INTRUSION DETECTION

**Dr Parveen Kumar Bansal**
**parveen.bansalin@gmail.com**

**Abstract**
Modern cyber threat is undoubtedly one of any state's pressing concern and it is a defense aspect that requires an innovative and integrated approach. More attention is being paid to the need to reinforce scientific and technical capabilities, actual armed forces and law-enforcement agencies as well as the relevant ministries and departments with modern technology. Cyber threats should be included into strategic defense and security planning. Cyber security which concerns a strategy for both defense and offense was pointed out as the tasks of information security. Active measures of information security concern the negative counteraction to criminals or attacks. Such programs do not just protect the information system from the attacks. They engage with criminals' activities. The purpose of this type of information security program is to look for and in a short time neutralize the perpetrator before it can do any damage.

**Keywords:** Machine Learning, Network Intrusion Detection, Cybersecurity, Intrusion Detection System (IDS), Deep Learning

## Introduction

The process of which modern computer networks are designed has led to there being more technologies in computer science, automation, the internet of things, 5G, and digital communication. In this modern society today, almost everyone, in every organization they belong to, interacts with different systems where they store, process and provide access to information. The connection and communication of these technologies have improved speeds in information sharing and accessibility although, also increased the attack surface for intruders and other cybercriminals. For this reason, networks are always in a state of threat from viruses, ransomware, piracy, DoS attacks, insiders, and brokers. Ensuring the security of network infrastructure has thus turned into a major concern for all concerned parties including, businesses, governments, universities, and individuals. one of the system in place that helps in detecting malicious activities in the network traffic is NIDS or Network Intrusion Detection System. “The earlier detection systems primarily have a dependence on signature-based detection methods in which the data or traffic is checked against a signature database that shows how a particular attack was carried out in the past. However, these technologies have been falling short most of the times due to their inability to identify new or unknown cyber threats commonly referred to as zero-day attacks. Moreover, the arrangement and the modifications of the signature base is very laborious and does not take into consideration modern cyber defense trends.

The limits are said to have been there for more than a decade. And the Machine Learning is one of the major breakthroughs that was put in place to counter all these challenges. Algorithms applied in machine learning can well distinguish between normal network behaviour and harmful network behaviour at ease. Importantly, such machines are smart in a way that they can learn, improve, and critically improve the detection process over time. In the case of the attack signature, the machine learning process well identified the complex forms of attacks, decreasing the false alarm and enabling very fast action against attacks. Different intelligent techniques are used for network intrusion detection, such as: focused learning, based on known patterns and anomalous behavior, semi-focused learning, reinforcement learning for semantic networks. Supervised methods of the classificative training called Decision Trees, Random Forests, Support Vector Machines (SVM), Naïve Bayes, and Logistic Regression models, require training sets with annotation to classify network traffic. Unsourced techniques, for example, K-

Means Clustering and Autoencoders may find known/unknown -periodic or non-periodic shapes without reference to the past, so useful in areas where unknown attack plays important role. In a more recent development, deep learning has not boosted the capabilities of intrusion detection through models like Biological Rhythms (ANNs), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long-Short Term Memory (LSTMs) among others, by ingesting very large network traffic data sets and learning intricate patterns of intrusions efficiently.

The rate at which machine learning-based intrusion detection systems operate is subject to a number of factors such as the quality of data, the selection of features, model selection and methods of evaluation. KDD Cup 1999, 2019 NSL-KDD, UNSW-NB15, CICIDS2017, and CSE-CIC-IDS2018 are some of the publicly available datasets which are quite often used for the training and evaluation procedures of machine learning models. Most activities of security, like classification, evaluation of classifiers using metrics such as accuracy, precision, recall, f1-score and roc metrics. Transparent assessments of classifiers using metrics such as accuracy, precision, recall, f1-score, and eventually roc metrics. Pre-processing and feature engineering are vital for enhancing the accuracy of the model and reasonableness of machine learning models. With great stride in technology again, the implementation of machine learning for the detection of network intrusion still possesses some complexiness. A variety of factors including high dimensional network data, class imbalanced data, encrypted traffic, malicious defences such as adversarial machine learning, scaling issues, and time constraints still degrade the performance of the system. For any advancement, the performance requirements motivated the need to ensure the optimality of privacy of data and each and every model to be explainable within the scope of machine learning. Cyber security is being dramatically changed by machine learning enabling systems that automatically adapt and detect any malicious access to a network. As the nature of cyber threats continues to evolve and become more sophisticated, it is clear that use of complex machine learning technologies within existing intrusion detection systems is the most pragmatic approach to improving network security. It is expected that further research about this issue will improve the quality of detection, reduce response time and increase the resistance of digital infrastructure to cyber threats.

**Network Intrusion Detection**

A network intrusion detection system, commonly referred to as NIDS, is continuously analyzing network data packets simply for the purpose of identifying packets that are in some way suspicious, often in the context of cyberattacks such as virus infections or SSL but not limited to only these, DoS, smishing, breaches of data and unauthorized uses, and infecting the system. The focus of the network intrusion detection systems is to protect all the resources on the network, which is why it raises alarms and advises all the necessary administrative measures to be taken to restrain any catastrophic damage done.

With the increase in the use of cloud computing, Internet of Things (IoT), mobile applications development, and distributed systems, the security of network connection has become a number one priority. Network modernization and IoT system further all activity not only to the high injury but also to high technicality and reduce the density against complex cyber threats. While traditional cybersecurity solutions, firewalls and antivirus, for example, are rather effective against known incidents, the novel types of assaults or even zero days attacks may pass unnoticed. Intrusion detection systems in networks serve the aforementioned purpose — they are an additional security layer that watches the activities on a network at all times and can detect any kind of threat be it known or unknown. Network IDSs usually come in two types of detection mechanisms. These mechanisms include the defense mechanisms for the network traffic of the organization as employable presentation: Signature-based detection and Anomaly-based detection. Signature based with network traffic patterns flagged in

its database as attacks when seen gives the very best accuracy for the already known threats. Yet, such designs are not suitable for new or altered patterns of attacks. With abnormal intrusion detection, a normal pattern of networking is created and any deviations from it such as increased bandwidth consumption or unauthorized access may enable the detection of malicious intent. The defense is one of the most sophisticated defenses against network intrusion in the modern era due to its ability to recognize changes in normal operations in real-time and take appropriate action in case of an attack. In order to to learning and identifying behaviour automatically in real-time as knowledge is acquired without any prior knowledge used alone against intrusions, machine learning applied in network-based intrusion prevention completely revolutionizes IPS or IDS performance. Among the modern appliances that need no or minimal human intervention are intrusion preventions or intrusion detectio that are based on machine learning prionciple. And various algorithms such as Decision Trees, Random Forests, Support Vector Machines (SVM), Artificial Neural Networks (ANN) and so on are used in creating Deep Learning based systems which can classify the incoming traffic locally among other functionalities electrifying intelligence to the network. The system increases detection performance, system detection precision, reduces false positives, and allows for detecting threats in real time enabling intervention. Therefore, advanced ML-based IDS without these drawbacks have all reasons to be an obligatory component of modern cyber security infrastructures in respect to malware detection in the whole intelligence framework.

**Evolution of Intrusion Detection Systems**

The concept of intrusion detection has evolved considerably over the past four decades in response to the increasing sophistication of cyber threats. Early computer security mechanisms focused primarily on user authentication and access control. However, as computer networks expanded during the 1980s and internet connectivity became widespread, organizations recognized the need for systems capable of continuously monitoring network activities for suspicious behavior. The first generation of Intrusion Detection Systems (IDS) relied on manual log analysis, where security administrators reviewed system logs to identify abnormal events. This approach was labor-intensive, time-consuming, and ineffective for large-scale networks. To overcome these limitations, signature-based intrusion detection systems were introduced. These systems used databases of known attack signatures to identify malicious activities accurately. While highly effective against known threats, they required frequent signature updates and were unable to detect previously unseen attacks.

The increasing frequency of zero-day attacks, polymorphic malware, and sophisticated cyber intrusions led to the development of anomaly-based intrusion detection systems. Instead of relying solely on predefined attack signatures, anomaly-based systems learned normal network behavior and detected deviations that could indicate malicious activities. Although these systems improved the detection of unknown threats, they often generated higher false-positive rates due to variations in legitimate network traffic. The rapid advancement of artificial intelligence and machine learning has transformed intrusion detection into a more adaptive and intelligent process. Modern IDS solutions employ supervised learning, unsupervised learning, and deep learning algorithms capable of analyzing massive volumes of network traffic in real time. These systems automatically extract meaningful features, recognize complex attack patterns, and continuously improve detection performance through model training and optimization. intrusion detection systems are increasingly integrated with cloud security platforms, Software-Defined Networking (SDN), Internet of Things (IoT) environments, and Security Information and Event Management (SIEM) solutions. Emerging technologies such as explainable artificial intelligence (XAI), federated learning, and edge intelligence are expected to

further enhance intrusion detection by improving transparency, preserving data privacy, reducing response times, and strengthening cybersecurity resilience against evolving cyber threats.

**Fundamentals of Machine Learning**

ML or machine learning constitutes an important sphere of AI which lets computers understand and make use of data in a way that they can self-recognize features, forecast outcomes or even make decisions without any explicit instructions as to what needs to be done. It's not like classical programming as with that, instructions need to be defined what to do, rules need to be manually implemented. When it comes to machine learning healthcare, unlike hard and fast programming models, the learning process of these programs is represented by the usage of prior data in order to ascertain the first loaded data for relevant patterns within the model variables. Thus, this is why at every step these models engage in analyzing any new piece of operating data that arrives enabling them to further perfect themselves upon considering the relationships between various interacting components as detected in the past data. Such kind of approach makes machine learning appropriate for such applications as handling significant amount of information which is constantly changing and making complex decisions.

Machine Learning belongs to a fundamental class of algorithms which are most popularly subdvided into four main types – supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. In supervised learning, there exist normally labeled datasets, using which models are trained and used for the classification and prediction purposes. When it comes to classification of attacks, algorithms like Decision Trees, Random Forests, Support Vector Machines (SVM), Logistic Regression, and Naïve Bayes, are extensively used within over 90-95% of networks because they are able to conclude with a very high degree of confidence the exact category of traffic being transferred as normal or malicious. In the context with unsupervised learning, note that this kind of learning works on unlabelled data and is able to discover underlying patterns as well is outliers. There are also clustering algorithms such as K-Means Clustering, and Autoencoders which are very effective in dealing with such threats, however, not all the threats are known for which these are useful. Semi-supervised learning is a machine learning technique that combines a limited amount of labeled data with vast amounts of unlabeled data to increase learning efficiency in cases where labeled data is scarce. In its turn, reinforcement learning is used christomathicly to assist in causal inference. An agent executes an action and then observes the outcome of the given action in the scenario with at least one decision point. Later, the reasoning and inferential process surrounding the action are evaluated. Both pure and hybrid historic, transactional and behavior datasets regardless of the nature of the transaction are used below. In fact, such a data-driven approach is still considered innovative for protection systems that are expected to put up a fight when faced with unknown attacks. Typical stages of machine learning include steps such as data collection, data cleaning, variables rather features and process design, processing of the data—time series or other structure type of data, developing a predictive model, validating the model, testing and putting the model forward. Data cleaning removes the erroneous measures, eliminates the missing values and converts the raw data into the form where it is ready for the learning algorithm. Feature selection aids in identifying the most significant characteristics of the network, which could help to improve the overall detection accuray and reduce the computational cost of the systems. Once the model is trained, accuracy testing metrics such as precision, accuracy, recall, after Receiver Operating Characteristic (ROC) curves are checked and only after that it’ll be tested in users' environment or the system. The progress of significant network data

accross speed and power allows machine learning introduction in many areas, such as medicine, economics, manufacturing, road transport and cyber security.

| Aspect | Description | Examples/Applications |
|---|---|---|
| **Definition** | Machine Learning (ML) is a branch of Artificial Intelligence (AI) that enables computers to learn from data and make predictions or decisions without explicit programming. | Fraud detection, spam filtering, intrusion detection |
| **Objective** | To identify patterns in data and improve decision-making through experience and continuous learning. | Predictive analytics, classification, forecasting |
| **Learning Process** | Involves data collection, preprocessing, feature selection, model training, testing, evaluation, and deployment. | Network traffic analysis, image recognition |
| **Supervised Learning** | Uses labeled datasets to train models for classification and prediction tasks. | Decision Tree, Random Forest, Support Vector Machine (SVM), Logistic Regression, Naïve Bayes |
| **Unsupervised Learning** | Discovers hidden patterns and relationships in unlabeled datasets. | K-Means Clustering, Hierarchical Clustering, Autoencoders |
| **Semi-Supervised Learning** | Combines a small amount of labeled data with a large amount of unlabeled data to improve model performance. | Medical diagnosis, intrusion detection with limited labeled data |
| **Reinforcement Learning** | Learns optimal actions through trial and error by receiving rewards or penalties from the environment. | Autonomous systems, cybersecurity response automation |
| **Feature Engineering** | Process of selecting and transforming relevant variables to improve model accuracy and efficiency. | Network packet features, protocol information, traffic statistics |
| **Training Dataset** | Historical data used to train machine learning models. | NSL-KDD, UNSW-NB15, CICIDS2017 datasets |
| **Testing Dataset** | Independent dataset used to evaluate the performance of trained models. | Validation of intrusion detection accuracy |
| **Model Evaluation Metrics** | Measures the effectiveness of machine learning models. | Accuracy, Precision, Recall, F1-Score, ROC-AUC |
| **Advantages** | Learns automatically, detects complex patterns, adapts to new data, and improves prediction accuracy. | Cybersecurity, healthcare, finance, manufacturing |
| **Limitations** | Requires quality data, high computational resources, careful feature selection, and may suffer from overfitting or bias. | False positives, class imbalance, adversarial attacks |

| Aspect | Description | Examples/Applications |
|---|---|---|
| **Applications in Cybersecurity** | Detects cyber threats by analyzing network behavior and identifying malicious activities in real time. | Intrusion Detection Systems (IDS), malware detection, phishing detection, anomaly detection |

**Role of Machine Learning in Cybersecurity**

One of the benefits of engineering in recent times is the ability to study the abnormal representation in cyber space. Designing tools that can perform this function is an aspect of machine learning. The use of machinges to improve the outlook of information is called machine learning and this function can be enabled to work in corroboration with an array of other practical devices with built-in intelligence. Typical security solutions require human intervention during operations. This means these solutions are effective only when the under attack threat is known and even the possible actions the opponent could undertake. As more and more cyber incidents evolve, the intelligent learning approach will be rather instrumental in such organizations as it allows for uncovering all forms of both known and unknown attacks, whether they follow a pattern of prior attack behavior or they instigate altogether new attack behavior. To such purposes is machine learning very useful alike the security access control and online fraud detection technologies. In this case, these detections do not require the signature based detection method but rather the behavior of the user in the system as against the typical behavior she used to exhibit which was legitimate. This also involves the use of machine leaning in detecting malware as opposed to the usual method in which malwares are identified through the use of signatures. In the same way, phishing detection uses the artificial intelligence strategies that help in the scanning of the content, the urls, the attachments of the emails and conducts a detailed examination of users behavior in order to find impostor emails.

Another critical use case is the identification of fraud in financial and bank systems. In such systems, ML engines can be used to critically evaluate migration's various patterns which assist in the recognition of abnormal activities typical of hacks or any sort of financial fraud. Machine learning can be applied in User and Entity Behavior Analytics (UEBA) systems to develop usual activity of individual equipment and users to discover any unusual account usage and any unauthorised access threats. It can also be used in conglomeration with Security Information and Event Management (SIEM) platforms as it is able to sift through many logs from many places all at once while ensuring all related items are attended to. For the security dimension for example in the field of cloud computing and in the Internet of Things (IoT), machine learning helps in monitoring the various decentralized nodes and noticing any abnormalities in the communication patterns and providing effective responses in the event of any form of threat. However, as protective tool, technology, the weaknesses include the threat to machine learning systems, problems of data used and ingested into the machine that is, data reliability or data accuracy, inappropriate class distribution, transparency of the model and ease of the computation. However, these detrimental factors cannot render machine learning keeping its usage as a main pillar for enhancing the current cyber-safe systems.

**Types of Network Intrusion Detection Systems**

Network Intrusion Detection Systems (NIDS) are designed to monitor network traffic and identify malicious activities that may compromise the confidentiality, integrity, or availability of information systems. Based on their deployment and detection methodology, intrusion detection systems are classified into several categories.

**1. Network-Based Intrusion Detection System (NIDS)**

A Network-Based Intrusion Detection System monitors traffic flowing across an entire network segment. It is typically installed at strategic locations such as gateways, routers, or network switches to inspect incoming and outgoing data packets. NIDS analyzes network traffic for suspicious patterns, unauthorized access attempts, malware communication, denial-of-service attacks, and other malicious activities. The primary advantage of NIDS is its ability to monitor multiple devices simultaneously without requiring software installation on individual hosts. However, encrypted network traffic and high-speed communication may reduce its detection effectiveness unless advanced analysis techniques are employed.

**2. Host-Based Intrusion Detection System (HIDS)**

A Host-Based Intrusion Detection System is installed directly on individual computers, servers, or endpoints. Unlike NIDS, HIDS monitors system logs, file integrity, application behavior, user activities, registry changes, and operating system events to detect unauthorized actions occurring within a specific device. HIDS provides detailed visibility into host-level activities and is particularly effective in detecting insider attacks, unauthorized file modifications, privilege escalation, and malware infections. However, deploying HIDS across numerous systems requires additional maintenance and computing resources.

**3. Signature-Based Intrusion Detection System**

Signature-based intrusion detection identifies attacks by comparing network traffic against a database of known attack signatures or predefined rules. This approach provides high detection accuracy for previously identified threats and generates relatively few false alarms. Its primary limitation is the inability to detect zero-day attacks, new malware variants, or previously unknown attack techniques. Regular updates to the signature database are necessary to maintain effectiveness against evolving threats.

**4. Anomaly-Based Intrusion Detection System**

Anomaly-based intrusion detection establishes a baseline of normal network behavior and identifies deviations that may indicate malicious activity. Machine learning algorithms are widely used in anomaly detection because they can automatically learn network behavior and recognize unknown attacks without predefined signatures." This approach is highly effective for detecting emerging cyber threats and advanced persistent attacks. However, it may generate higher false-positive rates if legitimate network behavior differs significantly from the established baseline.

**5. Hybrid Intrusion Detection System**

In order to optimize system performance, the Hybrid Intrusion Detection System is a mandatory strategy as it includes signature based detection and anomaly detection. The introduction of Signature based IDS introduced improves the detection speed and also serves the purpose of attacking an anomaly detection system that is always deployed whenever new forms of attacks are released. IT managers appreciate the improved results of this new approach that improves the accuracy of the detection and response to the cyber threats by lowering false positives and covering a broader scope of cyber rule of axiom.

Hybrid intrusion detection systems have seen an increase in their use, especially in enterprise networks, cloud computing platforms, as well as critical infrastructure due to the fact that they even out in the good and bad aspects of individual intrusion detection deployments. the style of implementing a particular IDS, of necessity, is determined by the size, the infrastructure, the availability of resources, as well as the threats themselves. The inclusion of machine learning techniques in such systems has advanced them considerably in terms of counterattacking APT, response time reduction, and general network security enhancement.

**Conclusion**

Network intrusion detection has experienced a complete transformation especially with the fact that machine learning has been applied in the algorithmic process of classifying cyber behaviors. This has made it possible for models that provide a much needed one stop approach in cyber threats analysis. Signature-based intrusion detection systems are becoming obsolete in the age of the machine learning systems. This paper discusses the approaches on which the machine learning techniques has had applied in intrusion detection over the years clarifying its limitations and also suggesting possible control measures. Intrusion detection system processes have been made efficient in order to effectively manage sophisticated cyberattacks as well as reduce false alarms and improve the efficiency of detecting threats. Additionally, the authors confirmed that the eff|c|ency of machine learning based intrusion detection hinges on several issues such as the kind of data in use. The efficiency of machines depends on the available data that is most appropriate with regard to feature selection, most appropriate with regard to algorithm selection and also promotes evaluation of robustness. Every freely available website has contributed in the development of effective intrusion detection software that is capable of securing present day networks. Moreover, combining signature-based detection with abnormality detection allowing the development of hybrid intrusion detection systems, which are still being investigated at the time of this research is cost effective. However, there are number of challenges which require further research or advanced solutions; they are such as while addressing operational environment, normative and interpretive requirements (interpretability) cannot be ignored, while concentrating on the same: secure network architecture design, encrypted networks, adversarial attacks, class imbalance, computational complexity, and model interpretability. Advancements in deep learning, eXplainable Artificial Intelligence (XAI), as well as newer technologies such as federated learning and edge computing, are expected to address the limitations and improve the design of the ID systems described. It is therefore, noted that the applications of machine learning in the network intrusion detection constitute a milestone in information security. Intrusion detection systems will remain very valuable in the field as they will prevent the growth of the cyber threats to computer systems.

**References**


- Belavagi, M. C., & Muniyal, B. (2016). Performance evaluation of supervised machine learning algorithms for intrusion detection. *Procedia Computer Science, 89*, 117–123. https://doi.org/10.1016/j.procs.2016.06.016
- Cardoso, L. S. (2007). Intrusion detection versus intrusion protection. In *Network security: Current status and future directions* (pp. 99–115). IEEE Press.
- Javaid, A., Niyaz, Q., Sun, W., & Alam, M. (2016). A deep learning approach for network intrusion detection system. In *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (BIONETICS)* (pp. 21–26).
- Kemmerer, R. A., & Vigna, G. (2002). Intrusion detection: A brief history and overview. *Computer, 35*(4), 27–30. https://doi.org/10.1109/MC.2002.1012428
- Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: Techniques, datasets and challenges. *Cybersecurity, 2*(1), 1–22. https://doi.org/10.1186/s42400-019-0038-7
- Saranya, T., Sridevi, S., Deisy, C., Chung, T. D., & Khan, M. A. (2020). Performance analysis of machine learning algorithms in intrusion detection system: A review. *Procedia Computer Science, 171*, 1251–1260. https://doi.org/10.1016/j.procs.2020.04.133

- Scarfone, K., & Mell, P. (2007). *Guide to intrusion detection and prevention systems (IDPS)* (NIST Special Publication No. 800-94). National Institute of Standards and Technology.
- Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 data set. In *Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications* (pp. 1–6). https://doi.org/10.1109/CISDA.2009.5356528
- Yang, L., Moubayed, A., Hamieh, I., & Shami, A. (2019). Tree-based intelligent intrusion detection system in Internet of Vehicles. In *Proceedings of the 2019 IEEE Global Communications Conference (GLOBECOM)* (pp. 1–6). https://doi.org/10.1109/GLOBECOM38437.2019.9013255
- Yin, C., Zhu, Y., Fei, J., & He, X. (2017). A deep learning approach for intrusion detection using recurrent neural networks. *IEEE Access, 5*, 21954–21961. https://doi.org/10.1109/ACCESS.2017.2762418