ENHANCING INTRUSION DETECTION SYSTEMS WITH A NOVEL HYBRID LEARNING-BASED FRAMEWORK

Aparna N[1]*, Dr. Chetana Tukkoji[2]

Abstract

The proliferation of cyber threats and the increasing sophistication of attacks have necessitated the development of robust Intrusion Detection Systems (IDS) to protect sensitive information and network infrastructures. Traditional IDS methods often struggle to effectively detect and respond to emerging threats due to their reliance on static rule-based approaches. To address these limitations, this paper presents a simplified and novel hybrid learning-based framework for strengthening IDS. The proposed framework integrates the strengths of two prominent Machine Learning (ML) techniques: Deep Learning (DL) and Ensemble Learning. DL models, specifically Convolutional Neural Networks (CNNs), are employed to extract high-level features from network traffic data, enabling the system to automatically learn and adapt to complex attack patterns. Ensemble learning is then utilized to combine multiple classifiers, leveraging the diversity of their decision boundaries, and enhancing the overall detection performance. The experimental results demonstrate that the suggested framework achieves superior detection accuracy, outperforming traditional rule-based approaches as well as standalone DL and ensemble learning models.

Keywords: Intrusion Detection System (IDS), Systematic Mapping Study, Hybrid Learning, Robustness, Feature Extraction, Deep Learning.

[1]*Research Scholar, GITAM School of Technology, Bangalore
[2]Assistant professor, GITAM School of Technology , Bangalore

***Corresponding Author: -** Aparna N
*Research Scholar, GITAM School of Technology, Bangalore

DOI: 10.48047/ecb/2023.12.si5a.0610

1. Introduction

The information technology sector is expanding at a breakneck speed all over the globe. Because of this, the sharing of information is now incredibly streamlined and hassle-free. On the other hand, these developments have introduced a great deal of difficulty into the communication system. An incursion could occur in the same structure or connection in the form of a variety of unique assaults. IDS is a piece of software that use detection algorithms to identify and categorize various forms of cyber-attacks on a host or network. IDS can be broken down into two categories: signature- IDSs that are both rule-based and anomaly-based (SIDS and AIDS). When an attack is detected by a SIDS, the system takes into consideration the assault's pattern and signature that are already specified. The performance of the system itself is monitored in the AIDS network, and these patterns are evaluated to the usual or periodic patterns that occur inside of the system to identify any incursion. The AIDS would report any kind of modification to the network is considered an intrusion into the network. The fact that it can carry out the identification of new assaults inside the network gives it an edge over SIDS. SIDS, on the other hand, can only detect assaults that are a match for the Sensors that came before it. IDSs can also be broken down into Network-based and Host-based IDSs, depending on where the data comes from. The Host Intrusion Detection System (HIDS) can determine whether a threat originates from inside the system by analyzing the data obtained from the Operating System (OS), servers as well as firewall logs, and database audits. The Network Intrusion Detection System (NIDS) can recognize external threats before they enter the computer network. The NIDS monitors and analyses network traffic that has been acquired from a wide variety of network data sources to locate possible network threats [1,2].

There are many ways to develop both SIDS and AIDS. Authors have indicated a greater interest in AIDS because it has the potential to overcome the constraints of SIDS. To begin, AIDS would be manufactured by processes such as ML and DL, approaches based on expertise, or statistically based procedures [3]. A statistical approach that uses a set of metrics such as the average, standard error, mode, and median to identify intrusion is known as the statistical method. Statistical IDS would be implemented using a variety of different models, including linear regression, multivariate, and time-series models. Second, the development of models is accomplished using knowledge-based methodologies, which are founded on a collection of rules derived from human expertise. Descriptive linguistics, intelligent agents, and finite-state systems are some of the tools that are used in the process of developing knowledge-based IDS. Finally, the development of an anomaly-based intrusion detection system often makes use of the method of ML. Unsupervised learning and supervised learning are the two categories that are included in the ML technique. In addition to these types of learning, there is also a hybrid learning method known as semi-supervised learning. Contrary to unsupervised, which can only be trained on unlabeled examples, supervised learning requires labeled examples to function properly. For training purposes, the semi-supervised method uses a smaller number of cases that have been labeled and a greater number of input instances that have not been labeled [4,5].

1.1 Intrusion Detection System (IDS)

Organizations could benefit from IDSs since they help stop unauthorized users from entering or obtaining additional access to a network. The CIA of a system is violated by an intrusion, as is its confidentiality. Wireless devices linked to the Internet of Things (IoTs) are simpler to hack than a traditional network, therefore attacks are expected to occur when a wireless connection becomes ubiquitous [6]. The need and necessity for IDSs, in IoT settings to minimize assaults that occur to exploit vulnerabilities, are further elaborated by the fact that security and privacy of information are challenges inside IoTs [7,8]. Companies ensure the safety of their networks by putting up defenses against potential threats. These measures include IDS, anti-malware programs, firewalls, and more. These methods, despite their usefulness, have several downsides and limits. Firewalls, for instance, guard against hackers gaining entry to private networks but do not prevent malware from spreading. However, even IDSs have their limitations, such as not being able to decrypt packets in transit [9,10]. Figure 1 depicts the architecture of IDS.

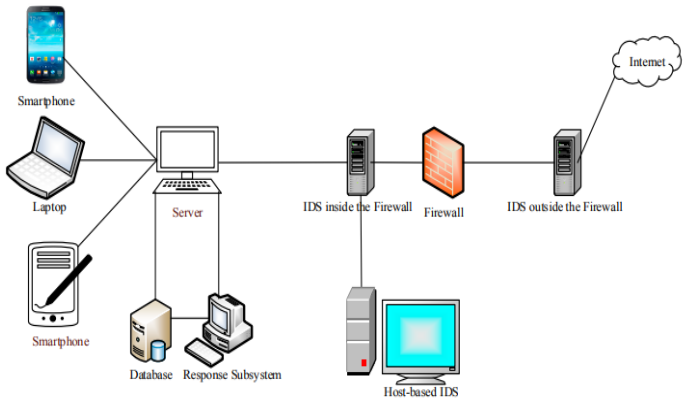

Figure 1: Architecture of IDS [11].

An IDS could operate in tandem with other systems. It could serve as a supplementary line of defense alongside primary defenses such as firewalls. For instance, if the firewall's protection is penetrated, the IDS would inform the administrators to better safeguard the system. Further, an IDS may warn an Intrusion Prevention System (IPS) so that the IPS can begin mitigating the danger once it has been identified. There is a need for robust, quick, and trustworthy IDSs because of the growing number of threats. These threats could occur in many forms, from emails and online assaults to ransomware [12].

1.2 Classification of Intrusion Detection System

There are two types of IDS are: NIDS and HIDS are explained in given below:

- **Network-Based Intrusion Detection Systems (NIDS)**

For the most part, commercial NIDS rely on either measurable indicators or calculated thresholds on features, such as packets of data, inter-arrival time, flow size, and other web traffic parameters, to design them in such a way that they function effectively within a specific time window. This allows the commercial NIDS to effectively model the various network traffic parameters [13]. They have a high rate of both false positives as well as false negative alarms, which is a problem for them. The NIDS has been shown to have a high rate of high false alarms, as well as a high proportion of negative result alerts, which means that it may overlook attacks more often and can notify when there is not an attack taking place, which is unnecessary. As a result, these commercial systems are useless against assaults conducted in the modern day. One of the successful approaches to countering the threats posed by modern assaults is the use of self-learning systems. This tends to make use of monitored semi-supervised, and unsupervised learning procedures to acquire the structures of a broad range of benign and malicious behavior using a vast corpus of both attacks and normal network systems and host-level events. These processes are utilized to acquire the patterns of a broad range of normal and malicious actions. Even though there is a wide array of ML-based solutions, the application of these solutions to commercial products is still in its infancy. Figure 2 shows the structure of NIDS.

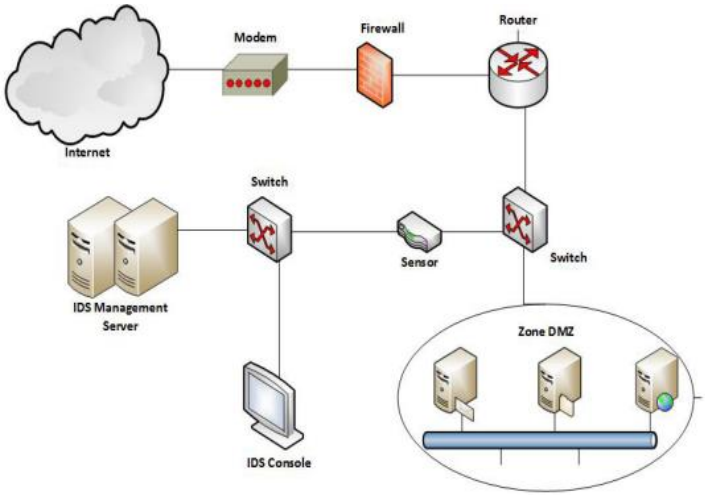

Figure 2: Network-based Intrusion Detection System architecture [14].

The currently available techniques based on ML produce a high proportion of false positives and have a considerable amount of required computation [15]. This is because ML classifiers learn the attributes of basic Transmission Control Protocol/Internet Protocol (TCP/IP) characteristics on a local level. DL is an intricate node of DL that understands patriarchal visual features and hidden sequenced associations by having to pass the TCP/IP data on several hidden layers. This allows for the learning of hidden sequential relationships and hierarchical feature representations. Image processing, voice recognition, Natural Languages Processing (NLP), and many other areas of study have all benefited greatly from the use of DL, which is far more effective than traditional approaches to Artificial Intelligence (AI) research [16].

- **Host-Based Intrusion Detection Systems (HIDS)**

Software applications such as Salmap, exploit code, Browser Exploitation, and Nmap are only some of the options that are accessible to users, they offer an essential foundation to investigate as well as collect data about target system vulnerabilities. Such information is used by malicious attackers to execute attacks against a wide range of programs like File Transfer Protocol (FTP) servers, Secure Shell (SSH) servers, etc. In order to protect host systems against exploits of this kind, security measures already in place, such as firewalls, cryptographic techniques, and authentications, have been developed. On the reverse hand, these approaches have a few limitations, and dishonest competitors would still be able to get unauthorized computer system access. A typical HIDS would conduct its activities at the host level to tackle this problem. There, it would monitor and analyze all traffic events that are going to take place on the program's application files, responses, and OS version [17]. In the transportation industry, traffic operations of this kind are often known as audit trials. A system call is a core component that allows the fundamental kernel activities of an operating system and the reduced system applications to interact with one another. These communications take place inside the operating system. System calls are used whenever an application has communication with the operating system of the computer. The behavior, ordering, type, and duration of these calls each contribute to the generation of a distinct trace. This would be utilized to differentiate between the applications that are known and those that are unknown. System calls made by regular processes and those made by intrusive processes are completely distinct from one another. So, doing a study of such system calls yields valuable insights into the operations of a system. In order to classify system calls based on processes, engineers have tried a variety of different techniques to feature engineering. They are referred to as N-grams, sequence grams, and pair grams respectively. The fact that HIDS gives users specific information on assaults is a significant benefit of using the system. Figure 3 depicts the architecture of the HIDS.

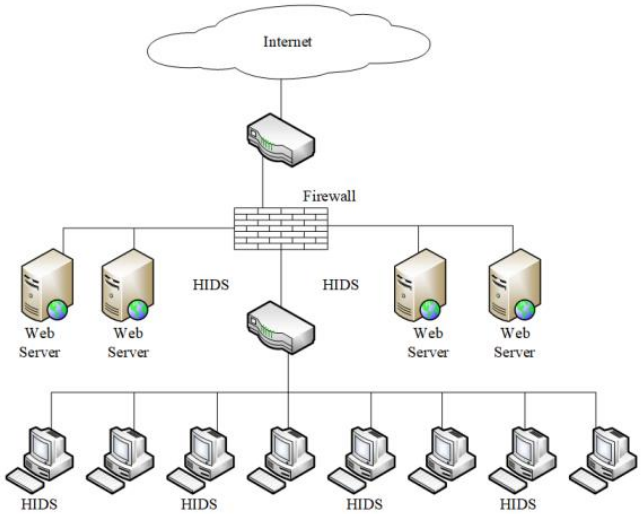

Figure 3: Structure of HIDS.

When it comes to the detection of breaches in security, the three essential characteristics of HIDS, notably the data source, the sensor, and the decision engines, all play a vital contribution. The ML modules are used by the decision engine to perform the intruder detection mechanism. The sensor component is responsible for monitoring any modifications that could happen in the data

source. The component of benchmarking the data source, on the other hand, calls for a great deal of analysis.

1.3Types of Attacks

Many networks are now essentially unprotected, giving hackers easier access to formerly safe information. Attackers aim to steal data and make users' access to digital resources difficult or impossible. The first line of defense against both external and internal threats might be a combination of many different defensive methods, such as access control, encryption, and firewalls [18]. The primary function of a firewall is to protect the node's entry points to the network from various forms of assault [19]. Access control systems are used for authentication, while cryptography enables secure communication. However, this anti-threat software is only able to offer exterior protection, making them useless for detecting interior threats and safeguarding any computer system or network from within. IDSs can monitor and identify both internal and external assaults, which is how they solve this issue. IDSs are computer hardware or software designed to detect and stop cyberattacks without any human intervention. IRSs keep a constant eye on the state of the system and react quickly to any signals from IDSs that indicate a problem [20]. IRSs use appropriate steps to keep data safe online. Therefore, alert processes are an appropriate mechanism for verifying the optimal response of these systems. Statistical attack detection methods, as well as methods for classifying attacks according to the degree to which they compromise data availability, confidentiality, or integrity, are also required. For instance, securing data integrity in the face of an assault on a corporate database system requires a suitable reaction. If the assault is directed at the network, however, the reaction should boost network performance and make more resources available.

The following table details some common and network-specific attacks. Each assault type is described in great depth below. Active assaults, passive attacks, insider attacks, outsider attacks, denial-of-service attacks, distributed denial-of-service attacks, covert channel attacks, and side channel attacks are the primary categories described based on Table 1. The goal of this categorization of attacks is to help it choose the best course of action to counter each kind of assault [21].

Table 1: Types of attacks of IDS

Types of Attacks	Reference	Attack Name	Description
Passive Attacks	**[22]**	Eavesdropping	Interception of data in transit on the network layer.
		Traffic Analysis	An attack designed to identify security flaws in the system's inter-process communication.
		Location Disclosure	Can provide any information about the topology or placement of the nodes in the network.
Active Attacks	**[23]**	Black Hole	A network problem characterized by packet drops.
		Gray Hole	Drops harmful packets as a malicious node would, but then reverts to its usual behaviour.
		Rushing	The routing procedure is sped up when a malicious node is present.
		Man in the Middle	The adversary acts as a go-between for two parties, discreetly relaying and interpreting their communications.
		Sybil	This assault targets P2P networks and involves the use of fake identities to gain a positive reputation.
Malicious Attacks	**[24]**	Bot/ Botnet	A botnet is a collection of infected computers working together to commit cybercrime via the Internet.
		Malware	This is malicious software that could cause serious harm to a computer or database.
		Malware Hosting	Malware could be hosted on either a mobile device or a computer.
Adversarial Attacks against IDS	**[25]**	Evasion	The attacker attempts to mislead the IDRS by changing the incursion pattern.
		Overstimulation	Attempts are made by intruders to flood the IDRS with attack patterns in order to trigger a large number of false positives.

		Poisoning	The attacker makes an effort to change the data needed to train and develop the detection algorithm by injecting a well-prepared pattern into the data.
		Reverse Engineering	An attacker employs a known attack signature to try to get access to IDRS's internal processing and trigger the system.
Side-Channel Attacks	**[26]**	Timing-driven attacks	Multi-level systems are vulnerable to this danger because it allows attackers to learn crucial details about the sensors employed in the targeted devices. This includes databases, operating systems, and networks.

1.4Challenges for Intrusion Detection System

There has been a lot of research towards new systems that can automatically recognize unusual system usage. In addition, Denning reported the creation of an IDS, that he proposed as a basis for a generic IDS [27]. Multiple techniques for automated network ID have been developed and used by specialists since then. Additionally, they have consistently looked for quicker, more scalable, and more accurate approaches to this problem. Estimates suggest that by 2020, the number of Internet-enabled gadgets would have risen to over 26 billion, thanks to the advent of the "IoT" and the age of big data. It stands to reason that if this trend continues, so too would the variety and volume of cyber security incidents. The difficulties with IDS are summarized in Figure 4. Problems include a slow reaction time, an imbalanced data collection, a high false alarm rate, and a poor detection rate. First and foremost, an IDS provides proactive security measures by continuously monitoring network traffic and identifying potential security threats in real-time. It detects and alerts network administrators about any suspicious or malicious activities, such as unauthorized access attempts, malware infections, or data breaches. By promptly notifying administrators, an IDS enables them to respond quickly and effectively to potential threats, minimizing the risk of damage or unauthorized access. Another advantage of an IDS is its ability to detect both known and unknown threats. Traditional security measures, such as firewalls and antivirus software, primarily rely on signature-based detection mechanisms, which can only identify known threats based on pre-defined patterns or signatures. However, an IDS utilizes various detection techniques, including anomaly detection and behavior analysis, to identify new and emerging threats that do not have known signatures. This feature ensures that the network remains protected against evolving threats and zero-day attacks.

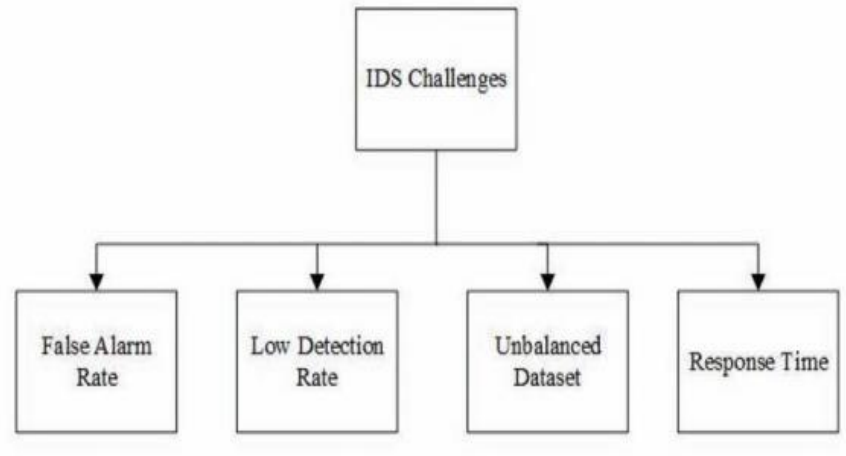

Figure 4: Challenges of IDS [28].

2. Review Methodology

This Systematic Review (SR) follows the guidelines set out in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statements for its methodology. PRISMA is an abbreviated set of procedures for reporting systematic reviews and meta-analyses. It is often used for reporting assessments that examine the results of participation. It can also keep track of systematic reviews that don't take interferences into account and instead focus on things like prevalence, diagnosis, or prognosis.

2.1Search Strategy

Intrusion Detection System is the primary topic of this Systematic Literature Review (SLR), which

employs a simplified and novel hybrid learning-based framework for strengthening intrusion detection system principles as well as relevant benefits in reliability, trustworthiness, organization, and transparency. Information is drawn from four key academic literature collections: Science Direct, IEEE, Springer, and Scopus. Studies of Previously Conducted Research (SLRs) are performed by researchers to gather data on previously conducted studies related to a certain research problem or subject.
Databases were searched many times using the same keywords to find relevant academic content; the most recent search was performed on January 25th, 2022. Scopus and other databases were searched indefinitely for the keywords in the topic and title, as well as the subject, title, and abstract. Only articles, reviews, conference papers, and reviews were accepted, along with articles from proceedings and bibliographical articles. Words searched for in Scopus and Near are included in Table 2. Several variations on the spelling of the keywords were also researched. The total number of files in this set was eight. A list of keywords and their intended uses could be seen in Table 2.

Table 2: Search strategy keywords.

	Keywords
1.	What is the Intrusion Detection System?
2.	What are the classifications of the Intrusion Detection System?
3.	What types of IDS techniques are developed most?
4.	Which evaluation metrics are used in hybrid intrusion detection systems?
5.	How does the simplified and novel hybrid learning-based framework improve the detection accuracy of intrusion detection systems compared to traditional rule-based approaches?
6.	What is the impact of the hybrid framework on reducing false positives and false negatives in intrusion detection?

2.1.1 The Journal Selected for Research

On January 1, 2022, I looked at Elsevier's Journal of XXXXXXXXXXXXXX to see how often the most important journals had been cited in previous studies. Used a key term to search for anything, no limits. To choose the best publication to publish the research, the findings were divided into two groups: those aimed at verifying and establishing the dependability of the research, and those aimed at gathering information. Researchers constructed a keyword search based on product names to identify businesses that either produced one of the top five best-selling goods in 2019 or ten or more distinct gadgets in the first wave of new releases scheduled for 2020. The articles were then sorted using keywords found in their titles, abstracts, and methods sections. This was done to (1) exclude irrelevant articles from the research and (2) establish what additional items were included in the trials. Then, they compiled a list of items in this category and sent out another round of inquiries, one for each item. In the end, eleven more people were included. In the results section, the team compiles the findings, providing a rundown of the keywords used to find each item.

2.2 Scrutinizing of paper for study

There are four phases to the Primary Studies (PS) selection process: identification, admission, inclusion, and screening. Finding each relevant research is the first order of business; the first search yielded 2375 matches. The papers presented at these conferences were located by a comprehensive search of many digital libraries and archives, including Science Direct, Springer, Scopus, and IEEE Xplore, all of which provided full-text access to the papers presented. After removing duplicates and sorting the data, we were left with 217 studies; of these, 28% were concerned with identifying emotional stress, 17% with ensuring scientific viability and consistency, and 33% were exploratory investigations of cutting-edge scientific applications and innovations. In the second phase, do a preliminary assessment by reviewing the article's title, keywords, and abstract. A total of 2335 entries have been omitted so far due to failure to meet inclusion criteria, mostly related to study scope and optimization subject. These two records were sent on for further scrutiny together with the uncertain ones and the ones tagged as comprise. The results of an evaluation of a database of systematic reviews are shown in Figure 5.

All relevant articles and reviews must be found in the bibliography, which requires manual searching. All the other documentation was reviewed as well. To decide which studies should be included and which should be left out of the current systematic review, the authors looked at their supplementary information and abstracts using the criteria laid forth in Table 3.

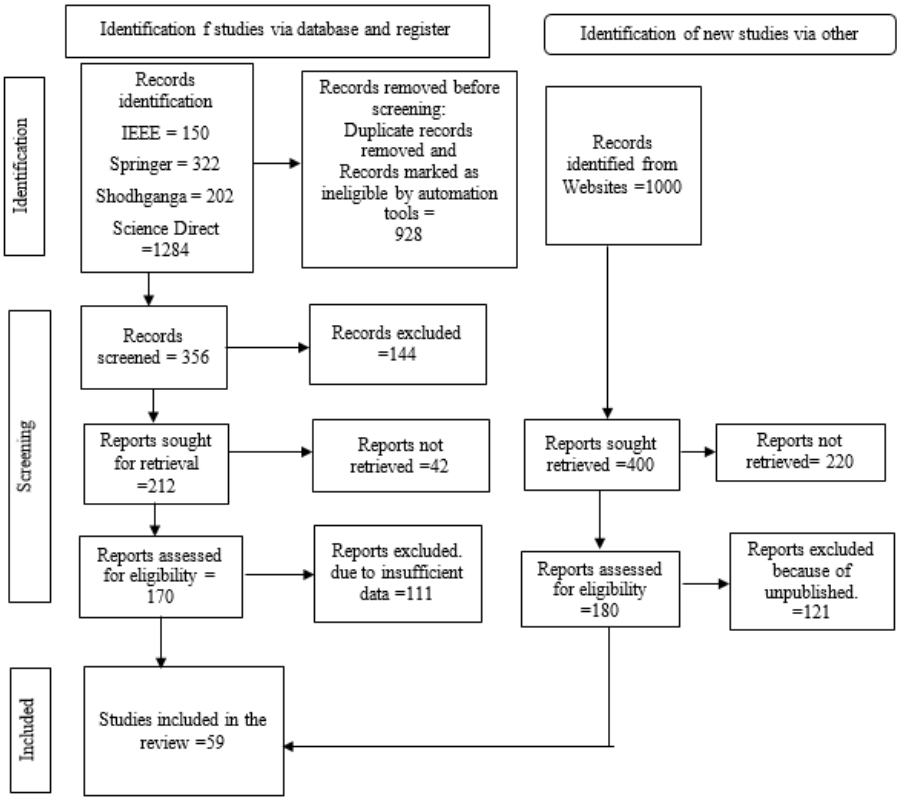

Figure 5: Schematic representing the steps involved in doing a database-wide systematic literature search.

Table 3: Systematic Review Criteria for Inclusion and Exclusion

Inclusion Criteria	**Exclusion Criteria**
I1: The paper should be peer-reviewed	E1: Papers that do not focus on the body stress-related study.
I2: The paper should be in the English language.	E2: Grey literature
I3: No time frame limit for publication	E3: Duplicate research and publications
I4: Papers should be published in research or full-article publication.	E4: Ph.D. theses, working papers, and project deliverables
I5: Standard of paper was blind to impact factor	

2.3Objectives of the Current Study

The objectives of the simplified and novel hybrid learning-based framework for strengthening intrusion detection systems are as follows:

- to develop a hybrid learning-based framework that combines DL and ensemble learning techniques to enhance the accuracy and effectiveness of intrusion detection.
- to integrate feature reduction techniques based on information gain to improve the efficiency and computational scalability of the framework.
- to evaluate the performance of the hybrid framework by comparing it to traditional rule-based approaches, standalone DL models, and ensemble learning models in terms of detection accuracy.
- to assess the adaptability and robustness of the hybrid framework in detecting and classifying novel and previously unseen attacks.
- to assess the adaptability and robustness of the hybrid framework in detecting and classifying novel and previously unseen attacks.
- to investigate the generalization capabilities of the hybrid framework to ensure it can effectively detect diverse attack patterns and handle different types of network traffic.
- to compare the performance of the hybrid framework with other state-of-the-art IDS, considering factors such as scalability, robustness, and detection capabilities.

These SRs also aim to build an open-source knowledge platform to aid future studies by collecting and evaluating major results from prior studies, describing, and contrasting them, and emphasizing problems and constraints that have occurred because of the study. Research on the simplified and novel hybrid learning-based framework for strengthening intrusion detection

systems was undertaken by assessing the current level of information in the field. This section discusses and evaluates different types of investigators and their methods, with a focus on the three primary research questions stated during the study's design phase. The following are the investigation questions:

RQ 1: What are the practical considerations and challenges in deploying the hybrid framework in real-world intrusion detection systems, such as scalability, integration with existing infrastructure, and resource requirements?

RQ 2: How does the hybrid framework compare to other state-of-the-art IDS methods in terms of performance, scalability, and robustness?

RQ 3: What are the advantages of the Intrusion Detection System?

To get started on accomplishing these aims, a thorough literature search was performed. Citation indexing databases and the Internet of publications were searched using tools like Sage, Emerald, Google Scholar, MDPI, Science Direct, IEEE Xplore, and Springer Link to find papers published within the last ten years that were relevant to the study topic. Additionally, a web search was carried out to identify the best wrist-mounted technology producers. Findings were examined using information gleaned from white papers, manufacturing guides, and scholarly articles.

3. Literature of Review

This section describes the previous studies of various authors based on Blockchain-Based Smart Contracts for Secure Record Management in the Pharmaceutical Industry.

Awajan A. et al., (2023) [29] introduce an innovative IDS based on DL for IoT gadgets. For preventing assaults on IoT devices, this smart system employs a four-layer deep Fully Connected (FC) network architecture. An empirical investigation of the suggested system's performance reveals its dependability against both simulated and genuine invasions. An average of 93.74 percent of Blackhole, DDoS, Opportunistic Service, Sinkhole, and Work hole attacks are detected. On average, the suggested IDS has an accuracy of 93.71%, recall of 93.82%, and F1-score of 93.47%. The average detection rate of this cutting-edge DL-based IDS is 93.21%, which is sufficient for strengthening the safety of IoT networks.

Qazi E. et al., (2023) [30] suggested a NIDS using a hybrid DL approach. In this research, the authors use a Convolutional Recurrent Neural Network (CRNN) to build a hybrid IDS using DL to identify network threats. The suggested Hybrid Deep-Learning-Based Network Intrusion Detection System (HDLNIDS) uses a CNN for gathering local features and a deep-layered recurrent neural network to extract the features, both of which improve the efficiency and predictability of the IDS. The results show that the recommended HDLNIDS can identify malicious assaults with a higher success rate (average of 98.90%) than existing IDS.

Khacha A. et al., (2022) [31] develop a DL model-based IDS. The suggested structure is built on the synergy of two cutting-edge methods for intrusion detection and classification: CNNs and Long Short-Term Memories (LSTMs). These methods excel at classifying key characteristics and completing computations quickly, respectively. The authors found that the CNN-LSTM model outperforms LSTM and other classic ML methods when it comes to detecting cybersecurity intrusions in IIoT applications. Additionally, in terms of accuracy rate, it is superior to two recently developed comparable models.

Baniasadi S. et al., (2022) [32] establish a new training procedure to fine-tune the used deep architecture's settings. Specifically, they enhance the Particle Swarm Optimization (PSO) method's exploitation and exploration by introducing a unique Neighbourhood Search-Based PSO (NSBPSO) technique. Then, it makes the most of NSBPSO to fine-tune the training for the deep architecture intrusion detector for the network. To measure how well the suggested classifier performs, and apply it to two datasets, UNSW-NB15 and Bot-IoT, both dedicated to NIDS.

Gupta S. et al., (2022) [33] offer a solution to these issues in IoT-enabled smart cities by suggesting an IDS based on a combination of optimization and DL. To get a reliable IDS, it is necessary to first pre-process the dataset. The Hybrid Chicken Swarm Genetic Algorithm (HCSGA) and the MinK-means Algorithm are then used to pick features and build clusters, respectively. Finally, the normal and attack data are sent into a DL-based Hybrid Neural Network (DLHNN) classifier after being modified. The NSL-KDD dataset is used to verify the accuracy of the suggested model, and the empirical findings demonstrate that the suggested IDS is superior to other comparable state-of-the-art methods.

Seth S. et al., (2021) [34] stated that a novel strategy for developing an IDS that is both quick and smart. In this paper, the authors present a Hybrid Feature Selection method for reducing the complexity of the model while maintaining the same level of attack prediction ability. Using the CIC-IDS 2018 dataset, the suggested feature selection decreases prediction latency by as much as 44.52% using certain methods and by as much as 52.68% using others. Incorporating hybrid feature selection and Light GBM into the suggested model improves accuracy to 97.73%, sensitivity to 96%, precision rate to 99.3%, and prediction latency to comparably low levels. The suggested model improved upon the baseline by increasing accuracy by 1.5% and precision by 3%. Additionally, network characteristics are analyzed in detail to reveal the shifts that occur in both benign and malicious sessions.

Khan M. et al., (2021)]35] evaluated a HUIDS that can automatically and effectively learn the representation of critical characteristics from large amounts of unlabeled raw network traffic data. To detect and categorize erratic attacks, they built an effective HIIDS using a spark Machine Learning Library (MLlib) -based robust classier like Logistic Regression (LR), Extreme Gradient Boosting (XGB) for anomaly detection, and a state-of-the-art DL like a Long Short-Term Memory Autoencoder (LSTMAE) for misuse. Through extensive simulations, they were able to show that the suggested spark MLlib and LSTMAE-based HIIDS greatly outperformed current ID techniques, with an accuracy rate of up to 97.52% on the ISCX-UNB dataset in a 10-fold cross-validation test.

Megantara A. et al., (2021) [36] suggest a hybrid ML approach to model construction by bringing together the feature selection approach (representing supervised learning) and the data reduction approach (representing unsupervised learning). To do this, it employs the feature significance decision tree-based approach with recursive feature reduction and the Local Outlier Factor (LOF) technique for identifying anomaly/outlier data. The empirical findings demonstrate that the suggested approach not only outperforms most previous studies on the NSL-KDD dataset but also achieves the best accuracy in identifying R2L (i.e., 99.89%) and maintains higher for other attack types. As a result, its performance is more consistent than that of competing products. The UNSW-NB15 dataset presents additional difficulties due to the use of binary classifications.

Khonde S.R. et al., (2020) [37] suggested that makes use of signature and anomaly-based detection to boost the detection rate and lower the false alarm rate. Several supervised and unsupervised ML techniques are included in this architecture to analyze current network activity. To speed up the system's calculation time, the number of features extracted from the dataset is often reduced using a variety of feature selection approaches. With this architecture being offered for decentralized network architectures, feature selection is crucial to the overall system's efficiency. The experimental findings demonstrate that the suggested architecture, which is based on an ensemble method, achieves superior outcomes in terms of detection rate and false alarm rate. The suggested design improves detection rates by 5% for signature detection and by 2% for anomaly detection. The percentage of false alarms has dropped by 0.05.

Khalvati L. et al., (2018) [38] offer a hybrid strategy for reaching that high-performance level. Creating a useful training dataset is a primary focus. The suggested approach also employs clustering and feature selection to make the most of their respective strengths for intrusion detection. K-Medoids clustering and the SVM feature-selection procedure are used to generate a novel training dataset. For further analysis, a Naive Bayes classifier is used. An additional hybrid algorithm and 10-fold cross-validation are examined and compared to the suggested technique. The suggested technique outperforms the alternatives in terms of accuracy, detection rate, and false alarm rate, according to experimental findings based on the KDD CUP'99 dataset.

Pham N. et al., (2018) [39] focus of this work is on enhancing IDS functionality using ensemble techniques and feature selection. The tree-based algorithms served as the foundation for the ensemble models with the other two ensemble approaches, Bagging and Boosting. NSL-KDD datasets were used to assess the suggested models. When dealing with the subset of 35 chosen features, the empirical findings indicated that the bagging ensemble model using J48 as the base classifier generated the greatest performance in terms of classification accuracy and FAR.

Tahir H. et al., (2022) [40] developed a machine-learning approach for network intrusion detection that combines K-means clustering with support vector machine classification. The study's goal is to boost the detection rate while decreasing the

false positive alert rate and the false negative alarm rate. In the development of the suggested method, the NSL-KDD dataset has been utilized. Some measures have been done on the dataset to enhance classification performance. The authors used a support vector machine to do the categorization. The findings of training and testing the suggested hybrid ML approach reveal that it improves detection rates while decreasing false alarms.

The authors' study approach and an overview of the relevant literature are presented in Table 4
.

Table 4. A summary of the Literature Review table.

Authors	Technique Used	Outcomes
Awajan A. et al., (2023) [29]	DL	This study presents a novel DL-based IDS that successfully detects the most widespread and often attempted assaults on IoT networks, with an average accuracy of 93.74%.
Qazi E. et al., (2023) [30]	CNN and DL	Results from the simulations show that the suggested HDLNIDS performs well in terms of accuracy and data loss.
Khacha A. et al., (2022) [31]	CNN-LSTM	The suggested model is assessed in terms of its accuracy, precision, False Positive Rate (FPR), and detection cost. When it comes to binary classification, the CNN-LSTM is one hundred percent accurate with an FPR of zero.
Baniasadi S. et al., (2022) [32]	NSBPSO	It has been shown experimentally that the suggested NIDS outperforms existing state-of-the-art systems in terms of accuracy and performance.
Gupta S. et al., (2022) [33]	HCSGA and MinK-means	The empirical result demonstrates the efficacy of the suggested IDS when compared to other comparable state-of-the-art methods.
Seth S. et al., (2021) [34]	Hybrid Feature Selection	The resultant model is more precise and has a higher attack detection rate and shorter lag time. Its accuracy of 97.73% is higher than that of any prior studies conducted in this area.
Khan M. et al., (2021) [35]	ML and DL	Models trained with traditional ML and LSTMAE DL were compared using standard classification measures including F1 score, Precision, Recall, DR, and correctness of class.
Megantara A. et al., (2021) [36]	LOF and ML	The system's efficiency is enhanced when the number of features is decreased by keeping just the most important ones.
Khonde S.R. et al., (2020) [37]	Ensemble Approach	In conclusion, the ensemble method outperforms the best individual classifiers. However, RF also offers high performance for detecting any style of assault.
Khalvati L. et al., (2018) [38]	K-Medoids Clustering, Selecting Feature, SVM, and Naïve Bayes	It was found that the suggested technique had a low false alarm rate, high detection rate, and high accuracy.
Pham N. et al., (2018) [39]	Ensemble Approach and Feature Selection	Empirical findings demonstrated that all suggested models achieved high accuracy and low FAR, with the bagging model using J48 as the base classifier and operating on the 35-feature subset yielding the greatest performance (84.25 percent accuracy and 2.79 percent FAR).
Tahir H. et al., (2022) [40]	SVM and K-means clustering	Both the accuracy and the false detection rate of previous IDS have been addressed by the suggested hybrid intelligent method to network intrusion detection.

4. Hybrid Learning

Dimensionality has been decreased because of the utilization of selected features that are focused on DL, which was made possible by the hybrid technique. Because of this procedure, both the intricacies of the data and its measurements are simplified due to the removal of features that are redundant and the filtering out of material that is not desired. Furthermore, it assists in the decrease in the amount of exertion that is associated with information technology while also enhancing one's ability to detect [41]. Furthermore, to improve the adequacy of minority incursions, which are urged by severely imbalanced data in the intrusion set of data, and to offer the dataset nearer to an equilibrium state to increase the adequacy of minority incursions, which are urged by an extreme imbalanced data in the intrusion set of data [42,43], Synthetic Minority Oversampling Technology (SMOTE) [44] is utilized. The content that has been collected via the use of methods such as data preprocessing, feature extraction, and

feature extraction would be condensed as part of the overall objective of the model that has been suggested. This study endeavor makes use of pre-processing techniques to build a novel and embodied architecture to reduce the dimensionality of imbalanced datasets. In the interest of making the study less overwhelming, this step has been taken. When dealing with a dataset that is not evenly distributed, critical stages include standardization, data balance, and segmentation algorithms. For the author to be able to finish the calculations in a less time-consuming manner while still accomplishing an appropriate amount of prediction performance both for bitwise and multilabel classifications, the author uses the SMOTE algorithm to weight the sets of data, and the author uses the Extreme Gradient Boosting (XG Boost) algorithm as an adaptive filtering method for decreasing the dimension. This allows the author to complete the computations using SMOTE.

A variety of ML-based classifiers, such as Random Forest (RF), Decision Tree (DT), K-Nearest Neighbor (KNN), Multilayer Perception (MLP), an ANN, and a CNN, were utilized to assess the efficacy of the model. These classifiers were utilized to evaluate the effectiveness of the model CNN. In the process of training each model, the XG Boost algorithm was used to select a minute percentage of the most significant attributes from the dataset. Accuracy, precision, recall, f1-score, Mean Square Error (MSE), Receiver Operating Characteristic Curve (ROC Curve), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) are the attributes that define that are used to evaluate the model. Other performance parameters include the Area Under the ROC Curve (AUC) score and the MAE. MSE, MAE, and RMSE are some of the other performance metrics. The hybrid learning model is shown in Figure 6.

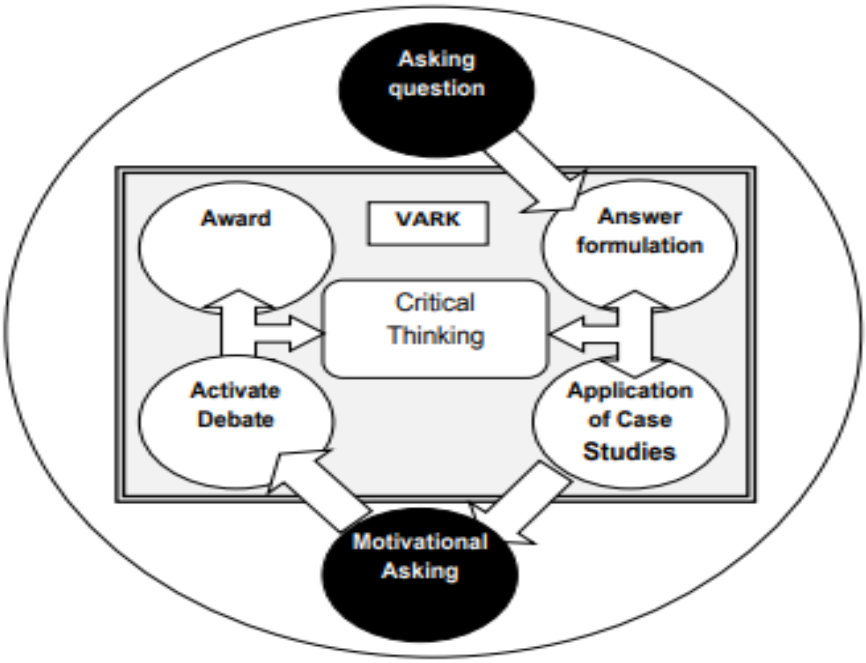

Figure 6: Hybrid learning diagram [45]

4.1Hybrid Model with Machine and Deep Learning

The author has suggested a hybrid method that arranges the datasets by integrating an efficient pre-processing technique with the ability to handle incomplete information, data trying to balance by employing SMOTE, characteristics trying to scale through the implementation of standardization, and label encoding. This is done to tackle the problems that have been recognized in the study. XGBoost then is utilized to identify the most helpful qualities for the ML and DL techniques to apply in the development of the models. This is done by using the ML and DL methods to produce the models. In order to recommend the method that is the most effective for detecting network attacks, the author assessed the efficacy of several DL and ML algorithms, such as DT, RF, KNN, MLP, ANN, and CNN. This allowed the author to provide a suggestion for the method that is the most effective for detecting network intrusion.

The author is going to talk about the basic components that are going to build up the hybrid approach that has been described in this portion of the article.

• Data Balancing using SMOTE.

The process of re-establishing equilibrium in data sets that have been out of whack is known as data balance. When working with data that is not balanced, one might make use of the well-known SMOTE technique, which is a strategy that can be applied. The problem of unbalanced data arises, however, when the classification method is weighted more heavily towards one group than another. There are several possible explanations for why this might occur. In order to address the issue of imbalanced datasets, the SMOTE technique creates imitation information to establish stability across the size of the minority category as well as the majority category [46]. The SMOTE methodology requires an input known as a parameter to establish a particular cutoff for the simulated samples. This is carried out to make the classifications of majority and minority more comparable to one another. SMOTE identifies entries that are reasonably near to one another and then modifies those entries one category at a time by attaching a random number based on the difference between both entries in the adjacent rows. This process is repeated until all the entries have been changed. In circumstances where the proportion of the vast bulk to the minority is large, the minority class should be sampled at a rate rather than under samples taken to accomplish the desired proportion of the minority class. This is because under-sampling would result in the minority class having a lower representation than desired. The SMOTE plays a significant role in the hybrid technology as the sets of data that are mismatched, including Knowledge Discovery and Data Mining (KDDCUP'99), must balance to compensate for worn correctly and less realistic predictive model. This can be done by balancing the datasets using the hybrid technique. Because of this, the datasets in question include items like the Knowledge Discovery and Data Mining (KDDCUP'99) conference proceedings.

• Feature Selection using Extreme Gradient Boosting (XGBoost)

In order to minimize the length of the subcategory to as small of a length as is practically possible according to a particular set of criteria, a technique that is known as feature extraction is used to determine a collection of fundamental qualities to employ in the process. The process of producing a new collection of characteristics that may be used either on their own or in conjunction is known as extracting features. This procedure can be carried out either singly or in combination. The extraction of features is one way. In contrast to this, it can comb through the data to identify the most beneficial characteristics. It helps to minimize fitting challenges, decreases adaption efficiency on test datasets, shrinks the amount of time required for training, and minimizes the model's interpretability, making it an essential phase in the procedure of ML. There are three basic kinds of methods for selecting features, which may be broken down as follows: a technique of selection that is reliant on either the filters, the wrappers, or the embedding features [47]. The combined feature selection technique makes it possible to design a model without the need to implement any additional feature selection techniques. This is made possible by the constructed selection of features, which is available inside the combined feature selection technique. Using the incorporated feature selection technique, the author can access the constructed selection of features. In order to select features, the filter-based feature method makes use of several sets of assessment criteria. The technique of analysis and the assessment of distances are both included in these criteria. The wrapper-based instance selection technique builds a set of features in a particular way before evaluating feature extraction based on the findings of classifiers. This takes place before reviewing feature extraction according to the outcomes of classifiers. It is feasible to dynamically eliminate individual characteristics from inside the process of generating a classifier if the author makes use of a technique known as integrated feature selection. Because of this, it is now feasible to carry out the processes of extraction and classification of features simultaneously.

4.2Machine and Deep learning algorithms

For the whole of this work, the author made use of the deep learning and machine learning algorithms that are listed below:

• Random Forest (RF)

An RF is an example of a meta-approximation that relies on average to attain a greater degree of precision than other types of approximations. In order to prevent the risk of the model being too closely aligned with the data, several tree-based classifiers are used and applied to the various sub-trials of the dataset [48]. It does this by mixing several separate unprocessed DTs that were created from multiple replicates of the training set. These DTs are not cut in any way. In addition to this, each feature group is sampled in a manner that is separate and distinct from the actual feature space. Each category and the tree would provide an estimate for a category, and the category that is

forecast by the vast number of trees would form the foundation for the model's prediction [49]. It starts by creating numerous different tree structures according to the initial characteristics of the training set of data, and then it cast a vote to integrate these tree structures into a single classification method. This is done according to the initial features of the training set of data.

• Decision Tree (DT)

Decision trees are a versatile tool that could find applications in a wide variety of fields in the future. Some of these fields include DL, pattern recognition, and image recognition, to mention a few. Putting things into categories is the basic objective of DT. Also, the DT classification approach has regular use in the realm of data mining. The root of the tree, the branch, and the child node are the three major components that come together to form DT. The output, also recognized as a tree structure, are the factors at which the separation of the set of data comes to an end. The root of the tree symbolizes the totality of the set of data, which is divided up into multiple sets of familiar qualities; the branches symbolize the different configurations of characteristics or characteristics; and the output is the factors at which the separation of the dataset ends [50]. Because of the ease with which they can be examined and the precision with which they can function across a broad variety of data types, DT can be used in a wide variety of contexts and has many applications [51]. In Figure 7, a diagrammatic depiction of a decision tree is shown.

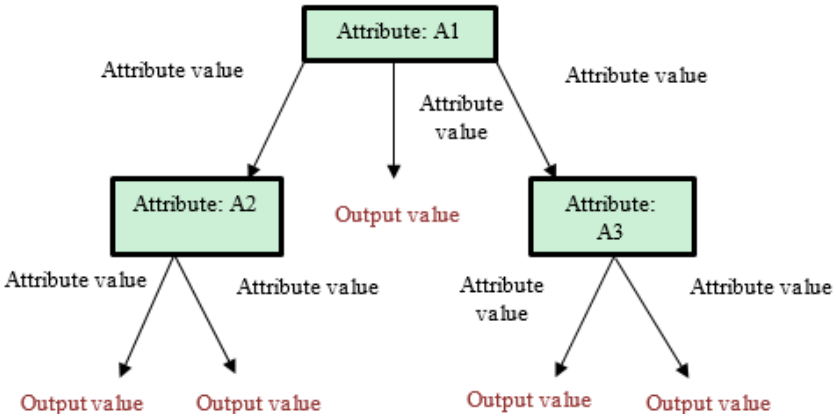

Figure 7: Decision tree diagram [52]

• K-Nearest Neighbor (KNN)

KNN takes unlabeled data instances and places them in a different class of similar examples which have been labeled. This is done by the determined distance that is the closest between every data point and the categories. This allows KNN to classify the data instances. KNN is straightforward to build in the absence of any previous knowledge about the distribution of the data because it uses the Euclidean distance equation to determine connections. In addition to this, it provides the end-user with accurate categorization recommendations based on the straightforward use of either similarity or distance [53].

• Multilayer Perceptron (MLP)

MLP is a typical design for ANNs that is made up of many layers, each of which is composed of neurons and the connections between them. It has the capability of determining the weighted sum of its inputs preceding making use of the input signal to produce a signal that's also destined for the following neuron in the circuit [54]. In between the layers that provide input and the ones that provide output, it has one or even more hidden layers. Neurons are typically connected from the deeper to the more superficial layers, and connections among neurons within a single layer do not exist. The neurons are stacked in layers and are organized in layers [55]. The quantity of neutrons that are present in the input nodes is equal to the number of measurements that are used for the pattern query. On the other hand, the number of neurons that are present in the output layer is proportional to the total number of categories.

• Convolution Neural Network (CNN)

Throughout the procedure of DL and CNN are often used as evaluation tools for visual representations [56]. Since they depend on the freely distributable design of the convolution kernels or filters, they are sometimes known to it as Shift Invariant and Space Invariant Artificial Neural Networks (SIANN). This is because they

are invariant to both shifts and spaces, which travel across input qualities and produce translational substantive answers that are known as data maps. The connectivity pattern among neurons mirrors the connectivity pattern of brain activity in vertebrates, which were powered by living organisms' internal workings [57]. It is likewise a regularized form of a multiple-layer perceptron, but it employs a different strategy for the regularization process than other regularizes do. It makes use of the hierarchical system that already exists within the content, and it assembles patterns of development utilizing separated patterns carved in their filters. This allows it to process the information more efficiently.

- **Artificial Neural Network (ANN)**

Artificial neural networks, often known as ANNs, are computers that are modeled after the neurons that can be found in animal brains. ANN is sometimes referred to as neural networks. In an ANN, they are a collection of linked systems known as nodes that represent neurons in a neocortex in a crude way [58]. So, in contrast, to synapses in the human mind, each connection has the potential to communicate with the neurons that are located nearby. The signals are obtained by an artificial neuron, which then performs an analysis on them before passing on the results to the cells to which it is coupled. A "signal" at a network is represented by an actual number, and the "signal" produced by a neuron is generated by certain non-linear features that are performed on the combination of the neuron's inputs.

5. Performance Evaluation of Intrusion Detection System

Multiple names are used to refer to the same set of categorization metrics for IDS. To assess an IDS's efficacy, one might consult Table 5, which displays the confusion matrix for a two-class classifier. Predicted class instances are listed in one column of the matrix, while real class instances are shown in the other.

Table 5: Confusion Matrix for IDS system

Actual Class	**Predicted Class**		
	Class	Normal	Attack
	Normal	True Negative (TN)	False Positive (FP)
	Attack	False Negative (FN)	True Positive (TP)

The following metrics are often used to assess the effectiveness of IDS systems:

- **True Positive Rate (TPR):**

It is determined by dividing the number of successful attack predictions by the total number of assaults. It is quite difficult for an IDS to achieve a TPR of 1, which would indicate that all intrusions have been detected. Detection Rate (DR) and Sensitivity TPR are other names for TPR. It is possible to calculate the TPR using the formula:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

- **False Positive Rate (FPR):**

It is determined by dividing the total number of false positives by the total number of benign occurrences.

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}$$

- **False Negative Rate (FNR):**

A false negative occurs when a detector incorrectly labels an out-of-the-ordinary occurrence as ordinary. The FNR could be written as a formula, specifically:

$$\mathrm{FNR} = \frac{\mathrm{FN}}{\mathrm{FN} + \mathrm{TP}}$$

- **Classification Rate (CR) or Accuracy:**

The CR evaluates the effectiveness of the IDS in identifying abnormal traffic patterns. The proportion of successful predictions to total occurrences is how it is often expressed.

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

- **Receiver Operating Characteristic (ROC) curve:**

The x-axis of ROC represents FPR, whereas the y-axis represents TPR. The ROC curve displays the TPR vs the FPR at various threshold values. A decision threshold is shown by a point on the ROC curve, which corresponds to a certain FPR and TPR combination. Different points on the ROC, with varying False Alarm Rate (FAR) and TPR, are chosen when the classification threshold is changed. The ROC curve crosses through the top left corner (100% sensitivity, 100% specificity) for a test with perfect discrimination (no overlap in the two distributions). See Figure 8 for the ROC curve.

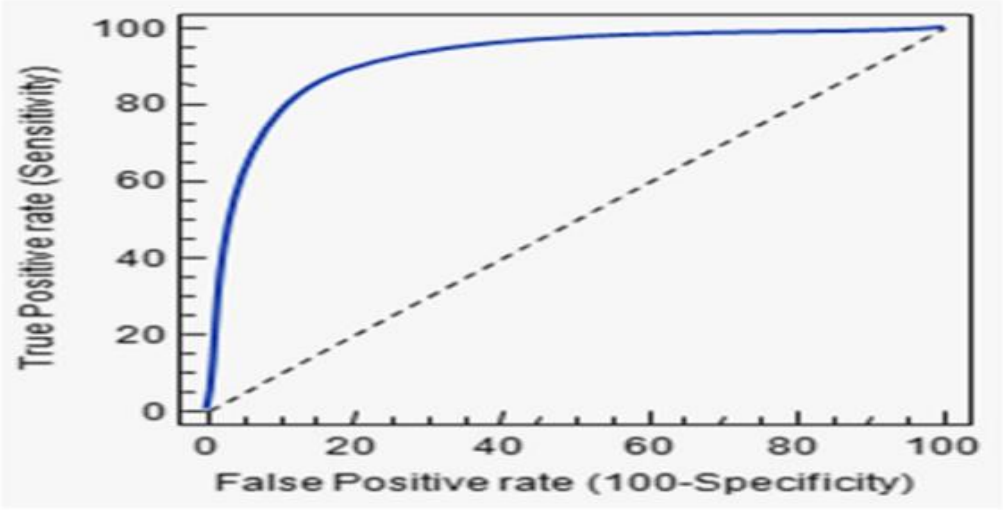

Figure 8: ROC Curve [59]

6. Conclusion and Future Scope

The development of a simplified and novel hybrid learning-based framework has proven to be a significant breakthrough in strengthening IDS. By integrating various machine learning techniques with traditional rule-based systems, this framework addresses the limitations of existing IDS and enhances their effectiveness in detecting and mitigating security threats. The framework leverages the power of supervised and unsupervised learning algorithms to analyze network traffic patterns, identify anomalies, and classify potential intrusions with a high level of accuracy. Moreover, by incorporating feature selection and ensemble methods, it achieves robustness and adaptability in diverse network environments. The findings obtained from experiments and evaluations demonstrate the superior performance of this hybrid framework, outperforming traditional IDS solutions in terms of detection rates, false positives, and response time. Overall, the simplified and novel hybrid learning-based framework represents a promising advancement in intrusion detection technology, offering enhanced security measures for organizations and contributing to the continuous evolution of cybersecurity. In the future, the authors would investigate more assaults and offer an analysis based on the results.

References

1. Prasad, Mahendra, Sachin Tripathi, and Keshav Dahal. "An intelligent intrusion detection and performance reliability evaluation mechanism in mobile ad-hoc networks." Engineering Applications of Artificial Intelligence 119 (2023): 105760.
2. Henry, Azriel, Sunil Gautam, Samrat Khanna, Khaled Rabie, Thokozani Shongwe, Pronaya Bhattacharya, Bhisham Sharma, and Subrata Chowdhury. "Composition of Hybrid Deep Learning Model and Feature Optimization for Intrusion Detection System." Sensors 23, no. 2 (2023): 890.
3. Gutierrez-Garcia, Jose Luis, Eddy Sanchez-DelaCruz, and Maria del Pilar Pozos-Parra. "A Review of Intrusion Detection Systems Using Machine Learning: Attacks, Algorithms, and Challenges." In Advances in Information and Communication: Proceedings of the 2023 Future of Information and Communication Conference (FICC), Volume 2, pp. 59-78. Cham: Springer Nature Switzerland, 2023.
4. Chaganti, Rajasekhar, Wael Suliman, Vinaya kumar Ravi, and Amit Dua. "Deep Learning Approach for SDN-Enabled Intrusion Detection System in IoT Networks. " Information 14, no. 1 (2023): 41.
5. Henry, Azriel, Sunil Gautam, Samrat Khanna, Khaled Rabie, Thokozani Shongwe, Pronaya Bhattacharya, Bhisham Sharma, and Subrata Chowdhury. "Composition of Hybrid Deep Learning Model and Feature Optimization for Intrusion Detection System." Sensors 23, no. 2 (2023): 890.
6. Liao HJ, Lin CH, Lin YC, Tung KY. Intrusion detection system: A comprehensive review. Journal of Network and Computer Applications. 2013 Jan 1;36(1):16-24.
7. Pharate A, Bhat H, Shilimkar V, Mhetre N. Classification of Intrusion Detection System. International Journal of Computer Applications. 2015 Jan 1;118(7).
8. U. Raza, J. Lomax, I. Ghafir, R. Kharel and B. Whiteside, “An IoT and Business Processes Based Approach for the Monitoring and Control of High Value-Added Manufacturing Processes,” International Conference on Future

Networks and Distributed Systems. Cambridge, United Kingdom, 2017.
9. Mehmood T, Rais HB. Machine learning algorithms in the context of intrusion detection. In2016 3rd International Conference on Computer and Information Sciences (ICCOINS) 2016 Aug 15 (pp. 369-373). IEEE.
10. I. Ghafir, M. Husak and V. Prenosil, "A Survey on Intrusion Detection and Prevention Systems," IEEE/UREL conference, Zvule, Czech Republic, pp. 10-14, 2014.
11. Kizza, J.M. A Guide to Computer Network Security; Springer: London, UK, 2009.
12. Coulibaly, Keturahlee. "An overview of intrusion detection and prevention systems. " arXiv preprint arXiv:2004.08967 (2020).
13. Kazi, Mohamed Ali, Steve Woodhead, and Diane Gan. "An investigation to detect banking malware network communication traffic using machine learning techniques." Journal of Cybersecurity and Privacy 3, no. 1 (2023): 1-23.
14. Achbarou, Omar, and Salim El Bouanani. "Securing cloud computing from different attacks using intrusion detection systems." (2017).
15. Donkol, Ahmed Abd El-Baset, Ali G. Hafez, Aziza I. Hussein, and M. Mourad Mabrook. "Optimization of Intrusion Detection Using Likely Point PSO and Enhanced LSTM-RNN Hybrid Technique in Communication Networks." IEEE Access 11 (2023): 9469-9482.
16. Braca, Paolo, Leonardo M. Millefiori, Augusto Aubry, Antonio De Maio, and Peter Willett. "Large Deviations for Classification Performance Analysis of Machine Learning Systems." arXiv preprint arXiv:2301.07104 (2023).
17. Mahindru, Arvind, and Himani Arora. "DNN droid: Android Malware Detection Framework Based on Federated Learning and Edge Computing." In Advancements in Smart Computing and Information Security: First International Conference, ASCIS 2022, Rajkot, India, November 24–26, 2022, Revised Selected Papers, Part II, pp. 96-107. Cham: Springer Nature Switzerland, 2023.
18. Fraga, J.; Powell, D. A fault-and intrusion-tolerant file system. In Proceedings of the 3rd International Conference on Computer Security, Dublin, Ireland, 12-15 August 1985; pp. 203–218.
19. Inayat, Z.; Gani, A.; Anuar, N.B.; Khan, M.K.; Anwar, S. Intrusion response systems: Foundations, design, and challenges. J. Netw. Comput. Appl. 2016, 62, 53–74.
20. Scarfone, K.; Mell, P. Guide to Intrusion Detection and Prevention Systems (IDPS); Report Number: 800-94; NIST Special Publication: Gaithersburg, MD, USA, 2007.
21. Anwar, Shahid, Jasni Mohamad Zain, Mohamad Fadli Zolkipli, Zakira Inayat, Suleman Khan, Bokolo Anthony, and Victor Chang. "From intrusion detection to an intrusion response system: fundamentals, requirements, and future directions." Algorithms 10, no. 2 (2017): 39.
22. Inayat, Z.; Gani, A.; Anuar, N.B.; Anwar, S.; Khan, M.K. Cloud-Based Intrusion Detection and Response System: Open Research Issues, and Solutions. Arab. J. Sci. Eng. 2017, 7, 1–25.
23. Inayat, Z.; Gani, A.; Anuar, N.B.; Khan, M.K.; Anwar, S. Intrusion response systems: Foundations, design, and challenges. J. Netw. Comput. Appl. 2016, 62, 53–74.
24. Anwar, S.; Zain, J.M.; Zolkipli, F.; Inayat, Z. A Review Paper on Botnet and Botnet Detection Techniques in Cloud Computing. In Proceedings of the ISCI 2014—IEEE Symposium on Computers & Informatics, Sabah, Malaysia, 28–29 September 2014; p. 5.
25. Inayat, Z.; Gani, A.; Anuar, N.B.; Anwar, S.; Khan, M.K. Cloud-Based Intrusion Detection and Response System: Open Research Issues, and Solutions. Arab. J. Sci. Eng. 2017, 7, 1–25.
26. Yarom, Y.; Falkner, K. FLUSH+ RELOAD: A High Resolution, Low Noise, L3 Cache Side-Channel Attack. In Proceedings of the USENIX Security, San Diego, CA, USA, 20–22 August 2014; pp. 719–732.
27. D.E. Denning, An intrusion-detection model, IEEE Trans. Softw. Eng. SE-13 (1987), 222–232.
28. Aljanabi, Mohammad, Mohd Arfian Ismail, and Ahmed Hussein Ali. "Intrusion detection systems, issues, challenges, and needs." International Journal of Computational Intelligence Systems 14, no. 1 (2021): 560-571.
29. Awajan, Albara. "A novel deep learning-based intrusion detection system for IoT networks." Computers 12, no. 2 (2023): 34.
30. Qazi, Emad Ul Haq, Muhammad Hamza Faheem, and Tanveer Zia. "HDLNIDS: Hybrid Deep-Learning-Based Network Intrusion Detection System." Applied Sciences 13, no. 8 (2023): 4921.
31. Khacha, Amina, Rafika Saadouni, Yasmine Harbi, and Zibouda Aliouat. "Hybrid Deep Learning-based Intrusion Detection System for Industrial Internet of Things." In 2022 5th International Symposium on Informatics and its Applications (ISIA), pp. 1-6. IEEE, 2022.

32. Baniasadi, Sahba, Omid Rostami, Diego Martín, and Mehrdad Kaveh. "A novel deep supervised learning-based approach for intrusion detection in IoT systems." Sensors 22, no. 12 (2022): 4459.
33. Gupta, Subham Kumar, Meenakshi Tripathi, and Jyoti Grover. "Hybrid optimization and deep learning-based intrusion detection system." Computers and Electrical Engineering 100 (2022): 107876.
34. Seth, Sugandh, Gurvinder Singh, and Kuljit Kaur Chahal. "A novel time-efficient learning-based approach for the smart intrusion detection system." Journal of Big Data 8, no. 1 (2021): 1-28.
35. Khan, Muhammad Ashfaq, and Yangwoo Kim. "Deep learning-based hybrid intelligent intrusion detection system." Comput. Mater. Contin 68 (2021): 671-687.
36. Megantara, Achmad Akbar, and Tohari Ahmad. "A hybrid machine learning method for increasing the performance of network intrusion detection systems." Journal of Big Data 8, no. 1 (2021): 1-19.
37. Khonde, S. R., and V. Ulagamuthalvi. "Hybrid framework for intrusion detection system using an ensemble approach." International Journal of Advanced Trends in Computer Science and Engineering 9, no. 4 (2020).
38. Khalvati, L., M. Keshtgary, and N. Rikhtegar. "Intrusion detection based on a novel hybrid learning approach." Journal of AI and data mining 6, no. 1 (2018): 157-162.
39. Pham, Ngoc Tu, Ernest Foo, Suriadi Suriadi, Helen Jeffrey, and Hassan Fareed M. Lahza. "Improving the performance of intrusion detection system using ensemble methods and feature selection." In Proceedings of the Australasian computer science week multiconference, pp. 1-6. 2018.
40. Mohamad Tahir, Hatim, Wail Hasan, Abas Md Said, Nur Haryani Zakaria, Norliza Katuk, Nur Farzana Kabir, Mohd Hasbullah Omar, Osman Ghazali, and Noor Izzah Yahya. "Hybrid machine learning technique for the intrusion detection system." (2015): 464-472.
41. Li, Xin, Peng Yi, Wei Wei, Yiming Jiang, and Le Tian. "LNNLS-KH: a feature selection method for network intrusion detection." Security and Communication Networks 2021 (2021): 1-22.
42. Tan, Xiaopeng, Shaojing Su, Zhiping Huang, Xiaojun Guo, Zhen Zuo, Xiaoyong Sun, and Longqing Li. "Wireless sensor networks intrusion detection based on SMOTE and the random forest algorithm." Sensors 19, no. 1 (2019): 203.
43. Ahmed, Hafiza Anisa, Anum Hameed, and Narmeen Zakaria Bawany. "Network intrusion detection using oversampling technique and machine learning algorithms." PeerJ Computer Science 8 (2022): e820.
44. Gonzalez-Cuautle, David, Aldo Hernandez-Suarez, Gabriel Sanchez-Perez, Linda Karina Toscano-Medina, Jose Portillo-Portillo, Jesus Olivares-Mercado, Hector Manuel Perez-Meana, and Ana Lucila Sandoval-Orozco. "Synthetic minority oversampling technique for optimizing classification tasks in a botnet and intrusion-detection-system datasets." Applied Sciences 10, no. 3 (2020): 794.
45. Sulistyanto, Hernawan, Harun Joko Prayitno, Sutama Sutama, Sabar Narimo, and Anam Sutopo. "The Effectiveness of Hybrid Learning-Based Adaptive Media to Empower Student's Critical Thinking Skills: Is It Really for VARK Learning Style?." Asian Journal of University Education 19, no. 1 (2023): 95-107.
46. Rustam, Furqan, Arif Mehmood, Saleem Ullah, Muhammad Ahmad, D. Muhammad Khan, Gyu Sang Choi, and B-W. On. "Predicting pulsar stars using a random tree boosting voting classifier (RTB-VC)." Astronomy and Computing 32 (2020): 100404.
47. Tang, Yuan, Zining Zhao, Shaorong Zhang, Zhi Li, Yun Mo, and Yan Guo. "Motor Imagery EEG Decoding Based on New Spatial-Frequency Feature and Hybrid Feature Selection Method." Mathematical Problems in Engineering 2022 (2022): 1-12.
48. Alkhatib, Kahlid, and Sayel Abualigah. "Predictive Model for Cutting Customers Migration from Banks: Based on machine learning classification algorithms." In 2020 11th International Conference on Information and Communication Systems (ICICS), pp. 303-307. IEEE, 2020.
49. Ahmad, Muhammad, Qaiser Riaz, Muhammad Zeeshan, Hasan Tahir, Syed Ali Haider, and Muhammad Safeer Khan. "Intrusion detection in the internet of things using supervised machine learning based on application and transport layer features using UNSW-NB15 data-set." EURASIP Journal on Wireless Communications and Networking 2021, no. 1 (2021): 1-23.
50. Mahesh, Batta. "Machine learning algorithms-a review." International Journal of Science and Research (IJSR).[Internet] 9 (2020): 381-386.
51. Mrva, Jakub, Štefan Neupauer, Lukáš Hudec, Jakub Ševcech, and Peter Kapec. "Decision support in medical data using 3D decision tree

visualization." In 2019 E-Health and Bioengineering Conference (EHB), pp. 1-4. IEEE, 2019.
52. Custode, Leonardo L., and Giovanni Iacca. "Evolutionary learning of interpretable decision trees." IEEE Access (2023).
53. Ahmed, Nazin, Rayhan Ahammed, Md Manowarul Islam, Md Ashraf Uddin, Arnisha Akhter, Md Al-Amin Talukder, and Bikash Kumar Paul. "Machine learning based diabetes prediction and development of smart web application." International Journal of Cognitive Computing in Engineering 2 (2021): 229-241.
54. Castro, Wilson, Jimy Oblitas, Roberto Santa-Cruz, and Himer Avila-George. "Multilayer perceptron architecture optimization using parallel computing techniques." PloS one 12, no. 12 (2017): e0189369.
55. Ramchoun, Hassan, Youssef Ghanou, Mohamed Ettaouil, and Mohammed Amine Janati Idrissi. "Multilayer perceptron: Architecture optimization and training." (2016).
56. Valueva, Maria V., N. N. Nagornov, Pavel A. Lyakhov, Georgii V. Valuev, and Nikolay I. Chervyakov. "Application of the residue number system to reduce hardware costs of the convolutional neural network implementation." Mathematics and Computers in Simulation 177 (2020): 232-243.
57. Yamashita, Rikiya, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. "Convolutional neural networks: an overview and application in radiology." Insights into imaging 9 (2018): 611-629.
58. Görgün, Emre. "Characterization of Superalloys by Artificial Neural Network Method." In Online International Symposium on Applied Mathematics and Engineering (ISAME22) January 21-23, 2022 Istanbul-Turkey, p. 67. 2022.
59. Khraisat, Ansam, Iqbal Gondal, Peter Vamplew, and Joarder Kamruzzaman. "Survey of intrusion detection systems: techniques, datasets, and challenges." Cybersecurity 2, no. 1 (2019): 1-22.