Vol. 06, No. 1 (2024) 311-320, doi: 10.24874/PES.SI.24.02.014

Proceedings on Engineering Sciences

www.pesjournal.net

# UTILIZING MACHINE LEARNING-BASED INTRUSION DETECTION TECHNOLOGIES FOR NETWORK SECURITY

Rahul Kumar Sharma[1]
Arvind Kumar Pandey
Bhuvana Jayabalan
Preeti Naval

Received 09.11.2023.
Received in revised form 26.12.2023.
Accepted 05.01.2024.
UDC – 004.85

Keywords:

*Machine Learning, Network Security, Network attack, cybersecurity, Intrusion Detection Systems (IDS), Stochastic Cat Swarm Optimized Privacy-Preserving Logistic Regression (SCSO-PPLR).*

A B S T R A C T

*Effective intrusion detection systems (IDS) are becoming essential for maintaining computer network security due to the growing complexity of cyber-attacks. Machine Learning (ML) can increase the effectiveness of intrusion detection technology, which is an essential resource to safeguard network security. A novel ML technique for intrusion information detection called Stochastic Cat Swarm Optimized Privacy-Preserving Logistic Regression (SCSO-PPLR) is proposed. We assess intrusion detection systems using KDDCup99 dataset. The dataset is pre-processed using Z-score normalization to normalize the features. Next, Features are extracted by Principal Component Analysis (PCA). By comparing the results of the SCSO-PPLR methodology with traditional methods and using assessment criteria including accuracy, precision, recall, and F1-score, the model's performance is extensively evaluated. The study reveals that SCSO-PPLR is an acceptable strategy for intrusion detection in network security and it is effective. These insights broaden IDS and groundwork for further research on reliable cybersecurity remedies.*



## 1. INTRODUCTION

An intrusion detection system (IDS) links to networks for suspicious behavior by scanning and examining them. In contrast, well-known tracking techniques are signature and anomaly-based detection, both of which the area of study on privacy has long investigated. Depending on the scope of use, intrusion detection systems can fall into several categories. For example, both host and network-based IDS, with capabilities ranging from single PCs to extensive networks, are among the most used. The ability to find undetermined dangerous code is constrained in the host-based intrusion detection system (HIDS), which is dependent on the single structure and watches critical operating system files for unusual activity (Sarker et al., (2020)). The scope of NIDS is represented in Figure 1.

[1] Corresponding author: Rahul Kumar Sharma
Email: rahulsharma.cse@niet.co.in

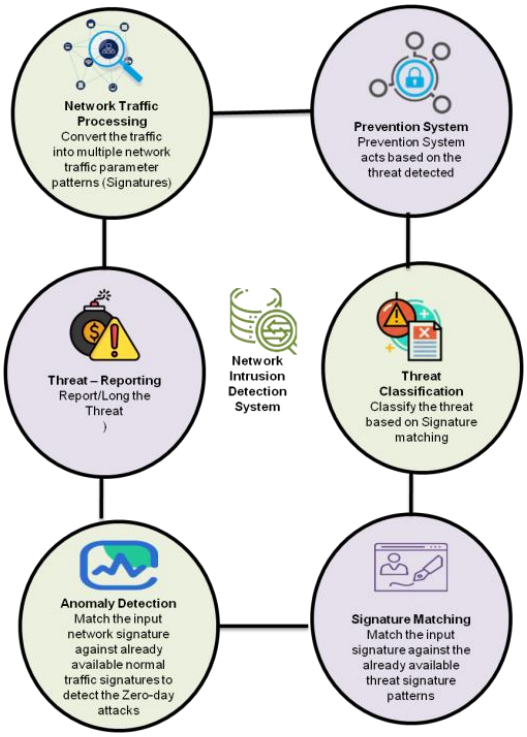


**Figure 1.** Framework of NIDS

With the widespread usage of the Internet, network security has become essential. The prevalence of data accessibility has led to the emergence of serious risks, such as viruses and network intrusions that can cause significant damage to businesses. As a result, businesses are investing money in research employing clever strategies to enhance security, such as intrusion detection technologies. It gets more important to preserve research on intrusion detection in computer networks. The deployment of the IP protocol in version 6 (IPv6) raises serious concerns about network security, specifically about intrusion detection, as the IPv6 protocol allows for connections to the Internet of Things (IoT) (Da Costa et al., (2019)). There are benefits and drawbacks regarding establishing security protection when using the security-by-design strategy, which is necessary due to the interconnected nature of the equipment. In addition, additional IoT techniques including the distributed ledger, cloud computing, machine learning (ML) and connected devices are considered. They are faced with the innate characteristic of resource and energy limitations present in IoT. Industrial control systems (ICS), smart grids, smart automobiles and medical gadgets are examples of cyber-physical systems (CPS) (Bertoli et al., (2021)). Owing to IoT's explosive growth and the advent of wireless devices, in Wireless Sensor Networks (WSN), there has been a huge rise in the targeted area, leaving the network vulnerable to cyber-attacks. To safeguard such networks, IDS techniques with a high degree of dependability, effectiveness and flexibility are needed. Currently, there are several issues with standard wireless network IDS technologies, including high false positive rates, poor precision rates and low detection accuracy. Suggesting a more accurate and efficient IDS methodology grows more important to enhance safety assurance in the setting of wireless sensor networks (Alruhaily & Ibrahim (2021)).The purpose of IoT is to gather and distribute data. Wireless technology is used to connect sensors to external devices. IoT technologies present several unheard-of chances for human connection. The network protocols that manage IoT operations under the Transmission Control Protocol TCP/IP model are HTTP, MQTT and DNS. IoT has the possibility to enhance daily life for individuals in a number of ways and aid in the creation of smart cities. A general framework for information involvement between nature, culture and business is provided by IoT. Therefore, for IoT-enabled intelligent cities to operate effectively there must be adequate communication among the amenities provided and the available assets (Gupta et al., (2022)). Technological innovation has made it possible to communicate effectively in every industry. In particular, the CPS is a unique platform that offers an improved method to share and transmit data between different locations via various communication channels. Because of the progress in communication transferring platforms made possible by this technology, the economy is developing at a rapid pace. Protection and resilience are difficult to provide, therefore procedures aimed at enhancing safety must take this into account. Component breakdowns and security are the two important elements that contribute to communication failure. Malicious action directed at the CPS system is made possible by hackers or other intruders; in the meantime, the system is developing and pervasive in contemporary society (Atul et al., (2021)). IDS is one of the most crucial security components that, when combined with antivirus programs, can manage a range of security threats. IDS plans can primarily categorize as abnormality and misuse detection approaches, which can be implemented utilizing a range of ML techniques. Systems that offer multiple categorizations and identify signature based systems largely rely on the signature of malicious actions and security threats. They are unable to identify fresh assaults, because their distinctive mark is not accessible to the IDS. The advantage of these systems is better at spotting to recognize fraudulent activity and its modifications (Lansky et al., (2021)). The study's goal is to use machine learning to detect intrusions using network security.

### 1.1 Contributions of the study

- Increase the precision of recognizing and categorizing malevolent actions in network traffic by utilizing machine learning techniques. Lower false positive and false negative results to raise intrusion detections for overall efficacy.
- Create IDS based ML to match the size and complexity of contemporary networks. Ascertain the capacity to manage substantial amounts of network data.

- KDDCup99 dataset is used to evaluate intrusion detection systems. Z-score normalization is used in the dataset's pre-processing to standardize its characteristics. For obtaining characteristics, Principal Component Analysis (PCA) is utilized.
- The SCSO-PPLR can combine optimization processes for parameter tuning to address the safety concerns of intrusion detection of computer networks and assure the security of information systems.

## 2. RELATED WORKS

Dina& Manivannan (2021) aimed at the last ten years that has seen a marked rise in network security breaches, partly because of the lucrative underlying cybercrime industry and the accessibility of advanced tools for initiating such assaults. Academic and industrial researchers has been creating systems yet proposing methods for spotting as well as avoiding these kinds of security flaws for over 40 years. The two main categories of solutions for handling network intrusions are signature-based and anomaly-based. Software- Defined Networks (SDN) as a young one that has the potential to revolutionize the network of architecture design, construction and operation. Traditional private network architecture was gradually giving way to flexible, customizable network design. Nevertheless, with new and developing security threats, this creative and enhanced technology added a further safety strain to the computer system architecture. The network's vulnerability has increased due to the focus shifting to a single point of failure, which makes the primary server an appealing target for attacks. Therefore, IDS integration into the SDN architecture must provide a network with an attack defence. Minawi et al., (2020) revealed that to provide consumers with a safer and more efficient experience, the automotive industry was innovating at an exponential rate. The innovations of Vehicle-to-Everything and autonomous cars were leading the way in defining the transportation of the future. The ability for cars to connect to a range of services has made vital network such as the Controller Is a Network (CAN), vulnerable to possible enemy manipulation. The CAN bus has several weaknesses in its conventional configuration, including inadequate bandwidth and no security. Threats can be launched via wired and wireless channels, taking advantage of entertainment, Bluetooth and monitoring connections to undermine the security, confidentiality and availability of data communication in cars. Si-Ahmed et al., (2023) assessed through the use of sensors to capture physiological data that is sent to instant computers for ongoing analysis by medical professionals, medical care sector completely changed with the help of Internet of Medical Things (IoMT). There were several advantages to these devices, including the capacity to identify diseases at an early stage and offer ongoing care to patients (Si-Ahmed et al., 2023). IoMT technology carries several serious security hazards, including the potential for patient death in the event of a privacy violation or the exposure of private information to collection assaults brought by wireless connection. Additionally, because of the diverse connectivity and the limited computing, storage and energy capacity of medical machinery, conventional safety techniques like cryptography were difficult to deploy. Amouri et al., (2020) described the use of IDS were essential for identifying malicious activity that impairs network performance. Wireless adhoc networks and WSNs were examples of networks that operate without the requirement for infrastructure, allowing for the transport of data. IoT was a more modern and inventive connectivity architecture that can be viewed as a combination of the aforementioned paradigms. Providing privacy to these networks was made extremely difficult by their dispersed nature and limited capabilities. IDS must be able to adapt to these kinds of obstacles. Alrowaily et al., (2019) examined, when it comes to data breaches or information protection, security is the most important factor. In addition, hackers were releasing a new range of cyber-attacks that prohibit users from controlling their computer systems. Because of this reason, it was very important that cyber-security research studies, including those on ID and prevention systems, continue to grow. IDS were useful defenses against malevolent intrusions. Saba et al., (2022) determined that the idea of IoT was founded with the intention of making people's lives better by providing an extensive range of intelligent, networked devices and apps across several sectors. But the main problem facing the devices in an IoT ecosystem was security risks. Although several state-of-the-art methods were safeguarding IoT gadgets, further developments are desired. ML has proven to identify patterns when other approaches have failed. One innovative method of enhancing IoT security was the use of DL. This resulted in a seamless detection process based on abnormalities. Amrollahi et al., (2020) focused on the act of network security is the process of preventing attacks that can threaten a network's availability. Furthermore, network safety must tackle the problem of illegal access to materials that were accessible through the network. Conventional detection methods' lengthy and complicated computations make them ineffective for handling large volumes of data. Large-scale data has provided with tools and methods that can be used to investigate and analyze data in IDS, hence assisting in reducing analyzing and training times. Furdek et al., (2020) introduced that operators are looking for automation of network diagnosis and management with ML to achieve cost-effective administration of intricate optical wireless networks. To facilitate cognitive, independent control of optical network security new capabilities were required. The effectiveness of ML-based methods for recognizing and determining the location of optical-layer assault,

as well as their compatibility with conventional Network Management Systems (NMSs), was the main topic of the paper. To suggest a privacy identification framework using Supervised Learning (SL), Semi-Supervised Learning (SSL) and Unsupervised Learning (UL) techniques, along with an assault distribution framework determined the precise place of compromised link or dangerous connection. This framework enables cognitive security diagnostics. Olowononi et al., (2020) obtained the potential of CPS to combine the cyber and physical worlds were one of its defining characteristics. Their implementation in vital infrastructure has proven to have the capacity to change the globe. Because of their crucial nature and the extensive consequences that cyber-attacks had on individuals, facilities and the environment, it was difficult to realize this ability. The process of transmitting data from detectors to controllers across a wireless communication network, which increased the attack surface, was one of the factors that draw cyber-worries in CPS.

## 3. METHODOLOGY

A study aimed at addressing the issue and enhancing the dependability of NIDS concerning Service Disruption, User-to-Admin, Probe and External-to-Internal attack to pre-process the data using z-score normalization. After this process, the data is extracted by principal component analysis (PCA). Figure 2 shows our concepts and recommended methods.

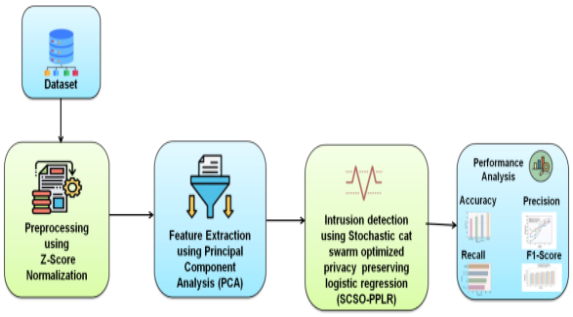


**Figure 2.** Structure of Proposed method

### 3.1 Dataset

In this research, the KDDCup99 dataset was used for the following model training and publicly accessible Packet Capture (PCAP) files from the ISTS dataset, validated the model using an assortment of real-time traffic, an assortment of PCAP files from the ISTS dataset and 10% of the KDDCup99 dataset (Kadam et al., 2020).

### 3.2 Using z-score normalization for preprocessing

Data pre-processing is to maximize the information captured and analysis by altering the data values of a particular dataset. Because of the significant contrast between the dataset's maximum and minimum values, normalizing the data makes the related processing procedure easier. The dataset features can be normalized using the normalization function, known as the z-score. Its characteristics allow it to normalize a standard distribution by standardizing a dataset of feature values. Equation (1) represents these values, with μ denotes the median and $\partial$ the mean average.

$$w'^{(j)} = \frac{w^{(j)} - \mu^{(j)}}{\partial^{(j)}} \quad (1)$$

### 3.3 Feature extraction using principle component analysis (PCA)

The study used PCA to extract features cut down on training time and testing on IDS improves detection accuracy. The outcomes demonstrate that the aforementioned strategy surpassed its rivals in terms of quality of detection and computing time. They employed PCA for feature extraction and conducted feature-based classification. As a feature extraction technique, PCA is used to produce a dataset with fewer dimensions of features and attributes. The feature classification is used to identify assaults and gauge accuracy in the IDS. By keeping elements with high variance, plenty of data as well as eliminating components with low variance and inadequate data, the PCA algorithm reduces the original image dimensions. One of the diverse analysis methods, the PCA algorithm, is based on communication theory's $K-L$ transform. This is the most effective to represent the $m$ sample data in a $(\overline{w}_1, \overline{w}_2 \ldots\ldots \overline{w}_n)$ low-dimensional in $d$ space is the primary challenge in Equation (2-3)

$$F_0(\overline{w}_0) = \sum_{j=1}^{m} \left\| \overline{w}_0 - \overline{w}_j \right\|^2 \quad (2)$$

The average of n sample points is expressed by $\overline{n}$ that is $\overline{n} = \frac{1}{m}\sum_{j=1}^{m} \overline{w}_j$the sum of distance squares $F_0(\overline{w}_0)$ of $\overline{w}_0$ to $m$ samples is calculated.

$$F_0(\overline{w}_0) = \sum_{j=1}^{m} \left\| (\overline{w}_0 - \overline{n}) - (\overline{w}_i - \overline{n}) \right\|^2 = \sum_{i=1}^{m} \left\| \left( (\overline{w}_0 - \overline{n}) \right) \right\|^2 + \sum_{i=1}^{m} \left\| \left( (\overline{w}_0 - \overline{n}) \right) \right\|^2 \quad (3)$$

From the equation (3), $(\overline{w}_i - \overline{n})$ has no connection to$\overline{w}_0$. When, $\overline{w}_0 = m$ lowest values are obtained. It demonstrates that the $d$ dimensional vector of the original $m$ samples can be best represented by the average $\overline{n}$ of $m$ samples.

### 3.4 Stochastic cat swarm optimized privacy preserving logistic regression (SCSO-PPLR)

#### 3.4.1 Stochastic cat swarm optimization (SCSO)

The SCSO-PPLR is expected to outperform the PSO in terms of efficiency. By using a novel set of

acquisition processes, the issue of the optimizer is solved by modeling the actions of a cat hunting for its prey. It's fascinating to note that the SCSO-PPLR feline and a particle in PSO are similar, with a small algorithmic variance. There is a pair of separate phases to feline actions: the searching phase and the tracking phase. The cat searches its surroundings during the searching phase and tries to go to the next location. Similar to this process, the cats chase some desired objects during the tracing phase. They enter the tracing phase right away if they locate the closest possible prey. The searching phase of the process for optimizing the problem is comparable to a wide search operation, the tracing phase is comparable to a close seek process. It is necessary to fix parameters like the mixed ratio (MR), count dimension change (CDC) and searching memory pool (SMP). The next step is to compute the fitness ratings of each potential point, after which the cats' similar fitness functions are allocated a likelihood that is similar to their own. Use equation (4) to get each suggested point's picking probability.This configuration can be applied to trace mode, where the range of fitness values must be computed. Every applicant must have an orientation and speed given to them in the monitoring mode. The cats' fitness parameters are assessed at each repetition and the outcomes are stored in database arrays based on their highest performance numbers. The SCSO-PPLR algorithm determines how a cat behaves and runs the algorithm until the finalization criterion is met. If not, the process repeats itself and the steps are repeated. Search mode, the definition of probability is as follows in Equation (4-7)

$$O_l = \frac{FS_l - FS_a}{FS_{max} - FS_{min}} \quad (4)$$

Tracing Mode, Velocity is as follows:

$$U_{l,c} = \beta \times U_{l,c} + d \times q \times (w_{best,c} - w_{l,c}) \quad (5)$$

Where dimension is,

$$c = 1,2, \ldots, N \quad (6)$$

And position is shown in equation (7):

$$w_{l,c} = w_{l,c} + U_{l,c} \quad (7)$$

$FS$ And $Pi$ stand for each application of fitness value and probability, respectively. Similarly, $q$ is any arbitrary integer between 0 and 1 yet $\beta$ is the residual mass and $d$ is the acceleration constant. The global and current positions are denoted by $w_{best,c}$,$dw_{l,}$ and $c$ respectively. Algorithm 1 provides the SCSO-PPLR algorithm.

***Algorithm 1: SCSO***

*Step 1: Cats is first placed at random locations in M dimensions, or $w_{l,c}$.*
*Step 2: Arbitrary cat velocity is initialized at $U_{l,c}$.*
*Step 3: Cats are selected arbitrarily from the community based on the mixing ratio; they are assigned to the searching mode and tracing mode.*
*Step 4: Each cat's competence must be determined and the regional location ($w_{l,c}$,$d$) and global position ($w_{best,c}$,) of each feline will be determined as a result.*
*Step 5: The fitness function compares the prior and current global best positions, saving the best one based on that comparison.*
*Step 6: The cat's location and speed will be adjusted using equations (5) and (6) for a fresh group.*
*Step 7: Verify the termination requirement; if achieved, terminate the program; if not, go back and do Steps 4–6 again.*

### 3.4.2 Privacy-preserving logistic regression (PPLR)

PPLRA seeks to improve computer performance by moving the majority of processes to the cloud while maintaining data privacy protection. Before providing a thorough description of the PPLRA, the approximate value of the sigmoid function and privacy protocols related to homomorphic cryptography are addressed. The following actions are taken by PPLRA to safeguard the privacy of the data: 1) it protects the training data sets locally; 2) it uploads the cipher text to the influence to run the LLR; 3) it transmits the confidential result locally and 4) it decodes the computed cipher text locally to get the final result. Figure 3 depicts the PPLRA procedure. It should be noted that to use homomorphic security, the sigmoid function in the LLR must be estimated by Taylor's theorem when cipher text is transmitted to the cloud for processing.

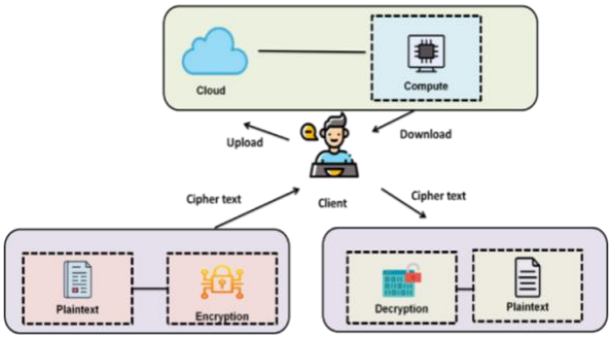


**Figure 3.** Architecture of PPLRA

The mathematical operations used in the homomorphic security approach are, multiplying, addition and subtraction. Stated differently, exponentiation and division operations are not supported. The Sigmoid factor together with separation and growing processes is the basic operation of the logistic regression algorithm.

As a result, Table 1 illustrates the operations of adding; subtracting, multiplying, dividing and exponentially increased operations are required.

**Table 1.** Functions for PPLRA.

| Operands | Homomorphic | Sample |
|---|---|---|
| + | Yes | $\sum_{j=1}^{n}(z^e - g_p(w^s))w_i^j)$ |
| - | Yes | $z^e - g_p(w^s)$ |
| x | Yes | $(z - g_p(w))ww_2$ |
| ÷ | No | $1/(1+f^{-s})$ |
| $f^y$ | No | $f^{-z}$ |

The Sigmoid function approximation requires the application of additional processes for support due to the limits of homomorphic authentication. Using Taylor's principle and quadratic summation forms, one can estimate the sigmoid relation. A fraction is eliminated in an assumption and exponential operations that allow cipher text to be directly accessed using homomorphic encryption. Taylor's theorem provides the following approximation for the sigmoid function in Equation (8-11):

$$T(w) = \frac{1}{f^{-w}+1} \tag{8}$$

$$\frac{1}{2} + \frac{1}{4}w - \frac{1}{48}w^3 + \frac{1}{480}w^5 - p(w^7) \tag{9}$$

$$\frac{1}{2} + \frac{1}{4}w - \frac{1}{48}w^3 + \frac{1}{480}w^5 \tag{10}$$

$$0.5 + 0.25w - 0.021w^3 + 0.002w^5 \tag{11}$$

Equation (8) demonstrates that calculating$x^5$, which can be changed to $x^5 = x.x.x.x.x$ is the primary challenge for homomorphic encryption. As a result, Algorithm 2 can be used to determine the forecasting factor safely. The middle variable is $o_i$ where $(i = 1,2,3....)$ and the function that calculate $T(w)$. Algorithm 2 denotes the predictive module computing in the secured cloud.

***Algorithm 2:*** *PPLR*
***Enter Data:*** *iteration, η*
***Execute Data:*** *ω*
***Local side:***
*Encryption training information along with parameters;*
*Move the information that is encoded to the server;*
***Termination local side***
*Sigmoid relation approximation in the sense of algebraic combination included in the cloud with the help of Taylor's principle;*
***for*** $iteration = 1,2,3,4 \dots, iteration_{max}$ ***do***
*execute PPLR algorithm;*
*Forward the findings to the local side;*
***Termination for***
***Termination cloud***
***Local side:***
*Decrypting the outcomes by applying a decoding process;*
*Determine the reliability of the categorization.*
***Termination local side***

## 4. PERFORMANCE ANALYSIS

In this study, several factors, including detection accuracy, recall, F1 score and precision, are necessary for assessing ML-based IDS technology's performance for network security. We suggest an overall comparison with relevant methods including Deep learning intrusion detection system (DL-IDS) (Otoum et al., (2022)), Naive Bayes (NB) (Meryem & Ouahidi (2020)) and System for detecting network intrusions using ensemble learning ELNIDS (Verma & Ranga , (2019). Key performance analysis concerns are broken out as follows:

### 4.1 Accuracy

Accuracy is a common metric used to assess a prediction model's overall accuracy. The proportion of correctly predicted true positive and true negative instances is subtracted from the total number of occurrence in the dataset and arrives at this calculation. The accuracy value is a number between 0 and 1, where 0 denotes no accurate predictions and 1 denotes perfect accuracy across all predictions. The following is the accuracy Equation (12):

$$Accuracy = \frac{True\ Positives + True\ Negatives}{True\ Positives + True\ Negatives + False\ Positives + False\ Negatives} \tag{12}$$

Figure 4 compares the accuracy of the suggested technique SCSO-PPLR, which reaches a standard of 92%, with the present method 83% of DL-IDS, 82% of NB and 87% of ELNIDS with coefficient values. Table 2 provides the suggested approach.

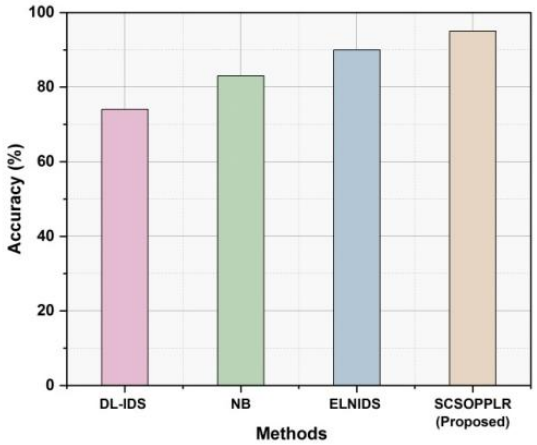


**Figure 4.** Comparisons between Accuracy
**Table 2.** Comparison of Accuracy

| Methods | Accuracy (%) |
|---|---|
| DL-IDS (Otoum et al., (2022)) | 83 |
| NB (Meryem & Ouahidi , (2020)) | 82 |
| ELNIDS (Verma & Ranga ,(2019) | 87 |
| SCSO-PPLR (Proposed) | 92 |

## 4.2 Precision

The measure used to assess a classification model's performance is precision. Its definition is the relationship between all of the favorable projections from the model and the actual positive results. Accuracy is especially important when the cost of false positives is substantial. The following formula can be used to determine the precision in Equation (13)

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (13)$$

The precision of the current approaches, which yield coefficient values for 75% of DL-IDS, 83% of NB and 86% of ELNIDS, is compared with the standard of 94% achieved by the suggested methodology, SCSO-PPLR, in Figure 5. The recommended method is shown in Table 3.

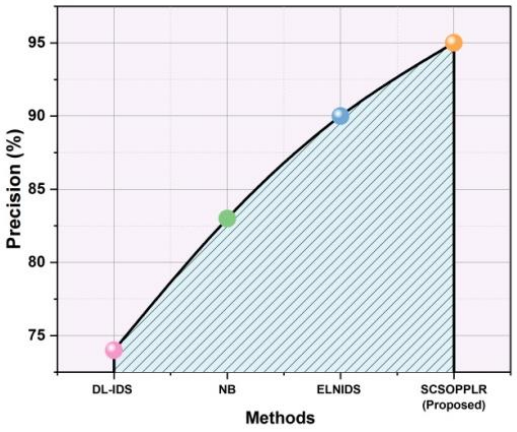


**Figure 5.** Comparison of precision

**Table 3.** Evaluation of precision.

| Methods | Precision (%) |
|---|---|
| DL-IDS (Otoum et al., (2022)) | 75 |
| NB (Meryem & Ouahidi , (2020)) | 83 |
| ELNIDS (Verma & Ranga ,(2019) | 86 |
| SCSO-PPLR (Proposed) | 94 |

## 4.3. Recall

The percentage of all discoveries in the actual class that are positively predicted is known as recall. It is referred to as sensitivity or true positive rate. It demonstrates the way a classifier can identify positive labels. Sensitivity is another name for the ratio of accurately matched all real class observations positive expected behaviors TP and FN. It is insufficient to determine whether a system can be trusted because the higher it is, the lower the true positives and false negatives as shown in Equation (14)

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (14)$$

In Figure 6, the standard of 93% attained by the recommended methodology, SCSO-PPLR, is contrasted with the recall of the present methodologies, which provide coefficient values for 79% of DL-IDS, 69% of NB and 89% of ELNIDS. In Table 4, the suggested approach is displayed.

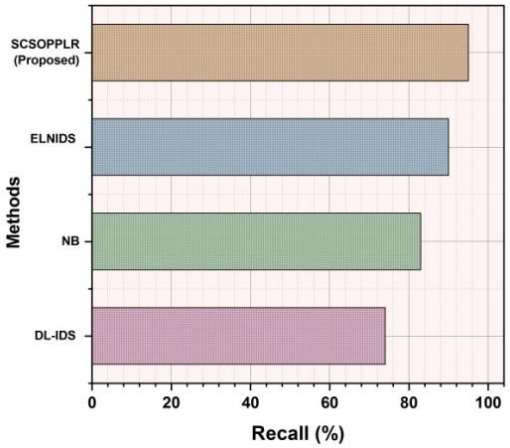


**Figure 6.** Comparison of recall

**Table 4.** Numerical outcomes of recall

| Methods | Recall (%) |
|---|---|
| DL-IDS (Otoum et al., (2022)) | 79 |
| NB (Meryem & Ouahidi , (2020)) | 69 |
| ELNIDS (Verma & Ranga, (2019)) | 89 |
| SCSO-PPLR (Proposed) | 93 |

## 4.4 F1-score

The harmonic mean of recall and accuracy is the definition of the F1-score. It is useful in uneven classifications since it balances accuracy and memory. This illustrates the connection between the collected information labels and the classifier positive labels. This score is more useful than accuracy when the costs associated with false positives and false negatives are different as shown in Equation (15).

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (15)$$

In Figure 7, the F1-score of the existing methods which produce coefficient values for 74% of DL-IDS, 83% of NB and 90% of ELNIDS is contrasted with the 95% level attained by the recommended methodology, SCSO-PPLR. In Table 5, the suggested approach is displayed.

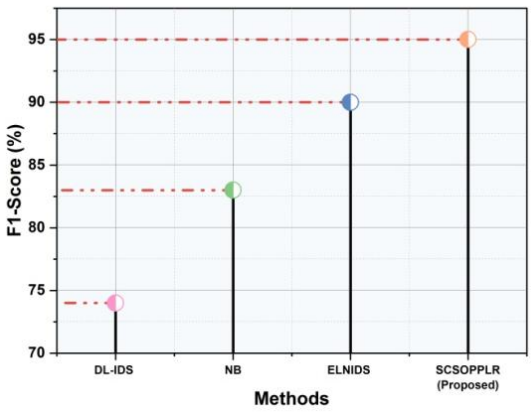


**Figure 7.** Comparison of F1-score

**Table 5.** Numerical outcomes of F1-score

| Methods | F1-Score (%) |
|---|---|
| DL-IDS (Otoum et al., (2022)) | 74 |
| NB (Meryem & Ouahidi , (2020)) | 83 |
| ELNIDS (Verma & Ranga ,(2019) | 90 |
| SCSO-PPLR (Proposed) | 95 |

## 5. DISCUSSION

In this study we discussed existing methods NB assumes that features are conditionally independent. This indicates that it is thought that the existence of one feature has no bearing on the existence of other features. This assumption might not hold because features are associated with real-world circumstances. The NB assigns a chance of zero for every class that is an identifiable variable in the test data. And it was not observed in the training data. This can result in the total forecast possibility being zero. Methods such as Laplace smoothing can be utilized to tackle this problem. DL-IDS models are sometimes described as black boxes; it can be difficult to understand the processes that arrive at certain results. A major disadvantage in security software, when it is necessary to comprehend the reasoning behind an intrusion detection decision, is when there is a lack of accessibility and simplicity. ELNIDS different models are trained and combined using ensemble methods, which can be computational and resource-intensive. Due to the network's complexity, real-time deployment and training times can be difficult, especially in high-speed network situations.

## 6. CONCLUSION

The identification of irregularities in streaming data is becoming more and more important as the number of cloud-connected devices rises. To increase a system's capability for identification, retaliation and prevention, the main application for machine learning-based detection and prevention solutions is network security. ML-based IDS methods have potential in terms of F1 score, recall, accuracy, and precision, there are issues with information quality, model accessibility and potential attacker challenges. Developing strong network security systems requires a comprehensive and integrated strategy that incorporates machine learning with other security techniques. Our hybrid ML system was successfully adopted, as demonstrated by the outcomes of our experiments. It decreased error rates and improved the cyber-attack labelling precision of 94%, accuracy of 92%, recall of 93% and F1 score of 95%. Additionally, it demonstrated its efficacy when compared to alternative machine learning algorithms that disregard the significance of updated security guidelines. The system must be monitored and updated to remain resistant to changing cyber security threats. Future research would concentrate more on methods for estimating the significance of incomplete data and minimizing the number of characteristics in network traffic assessment with information value and weight-of-evidence methods. To determine the difference in frequency at which an unknown behavior is regarded as a suspicious attempt, it could assess outlier.

## References:

Ahmed, M. R., Shatabda, S., Islam, A. M., & Robin, M. T. I. (2023). Intrusion Detection System in Software-Defined Networks Using Machine Learning and Deep Learning Techniques--A Comprehensive Survey. Authorea Preprints. https://doi.org/10.1109/EST.2017.8090413

Alrowaily, M., Alenezi, F., & Lu, Z. (2019). Effectiveness of machine learning based intrusion detection systems. In Security, Privacy, and Anonymity in Computation, Communication, and Storage: 12th International Conference, SpaCCS 2019, Atlanta, GA, USA, July 14–17, 2019, Proceedings 12 (pp. 277-288). Springer International Publishing. https://doi.org/10.1007/978-3-030-24907-6_21.

Alruhaily, N. M., & Ibrahim, D. M. (2021). A multi-layer machine learning-based intrusion detection system for wireless sensor networks. International Journal of Advanced Computer Science and Applications, 12(4), 281-288. https://doi.org/10.1036/ko.commnet.2018.011.023.

Amouri, A., Alaparthy, V. T., & Morgera, S. D. (2020). A machine learning based intrusion detection system for mobile Internet of Things. Sensors, 20(2), 461. https://doi.org/10.3390/s20020461.

Amrollahi, M., Hadayeghparast, S., Karimipour, H., Derakhshan, F., & Srivastava, G. (2020). Enhancing network security via machine learning: opportunities and challenges. Handbook of big data privacy, 165-189. https://doi.org/10.1007/978-3-030-38557-6_8.

Atul, D. J., Kamalraj, R., Ramesh, G., Sankaran, K. S., Sharma, S., & Khasim, S. (2021). A machine learning based IoT for providing an intrusion detection system for security. Microprocessors and Microsystems, 82, 103741. https://doi.org/10.1016/j.micpro.2020.103741.

Bertoli, G. D. C., Júnior, L. A. P., Saotome, O., Dos Santos, A. L., Verri, F. A. N., Marcondes, C. A. C., ... & De Oliveira, J. M. P. (2021). An end-to-end framework for machine learning-based network intrusion detection system. IEEE Access, 9, 106790-106805. https://doi.org/10.1109/ACCESS.2021.3101188.

Da Costa, K. A., Papa, J. P., Lisboa, C. O., Munoz, R., & de Albuquerque, V. H. C. (2019). Internet of Things: A survey on machine learning-based intrusion detection approaches. *Computer Networks*, *151*, *147-157*. https://doi.org/10.1016/j.comnet.2019.01.023.

Dina, A. S., & Manivannan, D. (2021). Intrusion detection based on machine learning techniques in computer networks. Internet of Things, 16, 100462. https://doi.org/10.1016/j.iot.2021.100462.

Furdek, M., Natalino, C., Lipp, F., Hock, D., Di Giglio, A., & Schiano, M. (2020). Machine learning for optical network security monitoring: A practical perspective. Journal of Lightwave Technology, 38(11), 2860-2871. https://doi.org/10.1109/JLT.2020.2987032.

Gupta, S. K., Tripathi, M., & Grover, J. (2022). Hybrid optimization and deep learning based intrusion detection system. Computers and Electrical Engineering, 100, 107876. https://doi.org/10.1016/j.compeleceng.2022.107876.

Kadam, G., Parekh, S., Agnihotri, P., Ambawade, D., & Bhavathankar, P. (2020, November). An Approach to Reduce Uncertainty Problem in Network Intrusion Detection Systems. In 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS) (pp. 586-590). IEEE. https://doi.org/10.1109/ICIIS51140.2020.9342634.

Lansky, J., Ali, S., Mohammadi, M., Majeed, M. K., Karim, S. H. T., Rashidi, S., & Rahmani, A. M. (2021). Deep learning-based intrusion detection systems: a systematic review. IEEE Access, 9, 101574-101599. https://doi.org/10.1109/ACCESS.2021.3097247.

Meryem, A., & Ouahidi, B. E. (2020). Hybrid intrusion detection system using machine learning. Network Security,2020(5), 8-19. https://doi.org/10.1109/ACCESS.2021.3097247.

Minawi, O., Whelan, J., Almehmadi, A., & El-Khatib, K. (2020, November). Machine learning-based intrusion

detection system for controller area networks. In Proceedings of the 10th ACM Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications (pp. 41-47). https://doi.org/10.1145 /341 6014. 3424581.

Olowononi, F. O., Rawat, D. B., & Liu, C. (2020). Resilient machine learning for networked cyber physical systems: A survey for machine learning security to securing machine learning for CPS. IEEE Communications Surveys & Tutorials, 23(1), 524-552. https://doi.org/10.1109/COMST.2020.3036778.

Otoum, Y., Liu, D., & Nayak, A. (2022). DL-IDS: a deep learning–based intrusion detection framework for securing IoT. Transactions on Emerging Telecommunications Technologies, 33(3), e3803. Doi:https://doi.org/10.1002/ett.3803

Saba, T., Rehman, A., Sadad, T., Kolivand, H., & Bahaj, S. A. (2022). Anomaly-based intrusion detection system for IoT networks through deep learning model. Computers and Electrical Engineering, 99, 107810. https://doi.org/10.1016/j.compeleceng.2022.107810.

Sarker, I. H., Abushark, Y. B., Alsolami, F., & Khan, A. I. (2020). Intrudtree: a machine learning based cyber security intrusion detection model. Symmetry, 12(5), 754. https://doi.org/10.3390/sym12050754.

Si-Ahmed, A., Al-Garadi, M. A., & Boustia, N. (2023). Survey of Machine Learning based intrusion detection methods for Internet of Medical Things. Applied Soft Computing, 110227. https://doi.org/10.1016/ j.asoc.2023.110227.

Verma, A., & Ranga, V. (2019, April). ELNIDS: Ensemble learning-based network intrusion detection system for RPL-based Internet of Things. In 2019 4th International Conference on Internet of Things: Smart innovation and usages (IoT-SIU) (pp. 1-6). IEEE. https://doi.org/10.1109/IoT-SIU.2019.8777504.

**Rahul Kumar Sharma**
Noida Institute of Engineering & Technology, Greater Noida, Uttar Pradesh, India
rahulsharma.cse@niet.co.in
ORCID 0000-0003-1604-6962

**Arvind Kumar Pandey**
Arka Jain University, Jamshedpur, Jharkhand, India
dr.arvind@arkajainuniversity.ac.in
ORCID 0000-0001-5294-0190

**Bhuvana Jayabalan**
Jain (Deemed to be University), Bangalore, Karnataka, India
j.bhuvana@jainuniversity.ac.in
ORCID 0000-0002-8372-6311

**Preeti Naval**
Maharishi University of Information Technology, Uttar Pradesh, India
er.preetinaval09@gmail.com
ORCID 0000-0003-2988-7082