Computer Systems Science and Engineering

Tech Science Press

DOI: 10.32604/csse.2023.039265

ARTICLE

Cybersecurity Threats Detection Using Optimized Machine Learning Frameworks

Nadir Omer[1,*], Ahmed H. Samak[2], Ahmed I. Taloba[3,4] and Rasha M. Abd El-Aziz[3,5]

[1]Department of Information Systems, College of Computing and Information Technology, University of Bisha, P. O. Box 551, Bisha, 61922, Saudi Arabia

[2]Department of Computer Science, College of Computing and Information Technology, University of Bisha, P. O. Box 551, Bisha, 61922, Saudi Arabia

[3]Department of Computer Science, College of Science and Arts in Qurayyat, Jouf University, Sakaka, 75911, Saudi Arabia

[4]Department of Information Systems, Faculty of Computer and Information, Assiut University, Assiut, 71515, Egypt

[5]Department of Computer Science, Faculty of Computer and Information, Assiut University, Assiut, 71515, Egypt

*Corresponding Author: Nadir Omer. Email: nhamed@ub.edu.sa

Received: 19 January 2023 Accepted: 13 April 2023 Published: 26 January 2024

ABSTRACT

Today's world depends on the Internet to meet all its daily needs. The usage of the Internet is growing rapidly. The world is using the Internet more frequently than ever. The hazards of harmful attacks have also increased due to the growing reliance on the Internet. Hazards to cyber security are actions taken by someone with malicious intent to steal data, destroy computer systems, or disrupt them. Due to rising cyber security concerns, cyber security has emerged as the key component in the fight against all online threats, forgeries, and assaults. A device capable of identifying network irregularities and cyber-attacks is intrusion detection. Several techniques have been created for Intrusion Detection Systems (IDS). There are elements in their effectiveness. Nevertheless, that provides room for more study. Finding an automatic method for detecting cyber-attacks is one of the biggest problems in cyber security. The recent trend is that the Machine Learning (ML) method has been demonstrated to be superior to conventional methods for IDS. Utilizing machine learning approaches, an effective intrusion prevention system will be designed. This research assessed different intrusion detection classification systems with particular applications. Before using ML classifiers for the classification process, the matrix factorization step of the Particle Swarm Optimization (PSO) technique was carried out. The categorization methods used in this study to classify network abnormalities were taken into consideration. Particle Swarm Optimization and Support Vector Machine classifiers (PSO + SVM) will be utilized in the proposed approach. The KDD-CUP 99 dataset will be used to confirm the results of the recognition algorithms. Due to the implementation, several performance metrics will be evaluated for various cyber-attack types, including specificity, recall, F1-score, accuracy, precision, and reliability.

KEYWORDS

Cybersecurity; machine learning; particle swarm optimization; support vector machine

1 Introduction

Cybersecurity technologies have been essential since the dawn of computer networks. Similar incidents have persisted to this day, and as a result, both known and unknown attack routes presently pose a severe cyber danger. However, security breaches started far earlier. By facilitating user authentication, assuring secure access, and guarding against data loss, IDS were used to defend computer systems from invasions. IDS collects and analyses data first, then employs a detector to produce alarms sent to a human network. Intrusion Detection Systems fall into two categories: Signature-based IDS (Sig-IDS) and Anomaly-based IDS (Anom-IDS) [1]. Only known incursions may be identified by matching specific data from the devices to the signatures, which prohibits signature-based approaches from identifying zero-day assaults. Contrarily, anomaly-based algorithms create a model by analyzing the properties of the behavior of the normal samples, and any deviation may be recognized as suspicious activity on the device [2]. The following will concentrate on ML-based security and Detection techniques because the volume of data being sent to the intermittent is increasing, leading to the adoption of cutting-edge networking concepts and complex inference models [3]. Despite the tiny sample size and the fact that some of the supplied data sets are no longer accessible, it nonetheless provides helpful instructions on how to spot and extract cutting-edge IP-based attack patterns and describes various ML-based strategies for the detection of network intrusions while concentrating on the characteristics of the various types of incursions. This highlights the importance of each criterion used to assess these aspects as well as how statistical parameters that are currently available may be used and modified for widespread detecting assaults [4].

The present interest in and developments in data and digital innovations over the past 10 years have made network security a significant research issue. It uses techniques like virus protection and intrusion detection systems to protect the network and all its associated assets within cyberspace. The PSO-SVM attack detection system is an attack detection technology that provides the necessary security by constantly checking Internet traffic for adverse and abnormal behavior [5]. The rapid advancements in the information and digital fields have led to a huge growth in network size and accompanying data. It has become challenging for network security to consistently identify breaches due to the proliferation of creative attacks that have resulted.

Additionally, because the invaders intend to launch various attacks inside the network, their existence cannot be ignored. One such technology is an intrusion detection system that keeps track of network traffic to preserve the network's secrecy, security, and reliability and protect against potential intrusions. Although the researchers' valiant efforts, IDS still struggles to identify new attacks, improve detection precision, and reduce false alarm rates [6]. The IMUTA attack is unique because upgrades gradually introduce a harmful function to a good program. Attackers use user trust to get past malware detection technologies [7]. DDoS attacks are now thought to be the most damaging online assaults. Attackers utilizing DDoS attempt to prevent authorized users from using facilities. Because of the potential for simultaneous assault from several sources, these attacks are risky [8].

But in the last ten years, technology has advanced so quickly that the size of networks and the number of applications they can accommodate have significantly increased. The growth of countless new attacks, either as variants of more well-established attacks or as completely original attacks, made it challenging to safeguard these networks and data hubs [9]. Any compromise of the node's data could significantly harm that firm's market value and cause financial losses. Cyberattack systems are now in use but are poor at spotting threats, including zero-day assaults and reducing false alarm rates. Security controls make up a system's second line of protection. Due to their complexities and high computational cost, IDS have received more attention [10]. IDSs can be combined with other security

precautions, such as network access, and the process is controlled from cyber-attacks [11]. Finding a large enough dataset is a significant difficulty in and of itself. Researchers find it difficult to get thorough and trustworthy statistics to validate and evaluate their suggested methodologies. To assess these strategies' efficacy, reliable datasets are needed.

Network traffic flows are analyzed using abuse detection, anomaly detection, and property protocol analysis. To discover the assaults, the detector employs filters and preset signatures. Human input is used to keep updating the database schema. The unexplained assaults cannot be found using this method, but the recognized attacks can be accurately located. Anomaly detection uses algorithmic techniques to find a hostile activity that has not yet been discovered [12]. Most of the time, anomaly detection produces many false positives. Most firms include abuse and outlier detection in their expert solutions to tackle this issue. This uses the designated vendor specification settings to discover pertinent protocol and application variants. Even though the techniques examined in this study are being explored more recently to increase the sophistication of such detection techniques, there is a lack of research to evaluate these machine learning with publicly available datasets.

Due to increased malicious programs, designing intrusion detection systems is increasingly challenging [13]. The fundamental challenge in detecting different and obscured malware is that its authors employ various information-hiding techniques to avoid detection by an IDS. Therefore, the top priority list has been moved up to the detection of zero-day threats. Greater cybercrime instances have demonstrated how easily a small hack may impact a company's critical operations and how fast attacks may spread abroad [14]. To find novel, sophisticated malware, an efficient IDS must be developed. A normal firewall is unable to immediately recognize different malware types. Hence an IDS's objective is to do so. The increase in computer malware has made the development of increasingly potent IDSs essential [15].

Even more, intrusion protection detection struggles to protect against contemporary threat attacks. Since a very long time ago, standard feature-based systems for attack detection have been in use. The breadth and dynamical range of the set of predefined signatures restricts the capacity to identify all types of attacks, especially innovative attack versions [16]. Researchers concentrated hard on developing new intrusion detection algorithms to solve this problem. One strategy is to use machine learning techniques. However, as is well known, there are no free meals, and every algorithm has benefits and drawbacks. Some systems might defend against one type of assault well while failing miserably in another. Many studies just focus on the total detection accuracy because the impact of identification for small-size data is typically relatively weak. We must focus on the capacity to recognize false strike information with small proportions. To achieve the best results, the current research proposes a model that may combine the advantages of each technique for different kinds of data recognition. Ensemble learning offers the benefit of boosting generalization ability and resilience compared to utilizing only one estimate by integrating the predictions of several base estimators [17].

Systems for intrusion detection can be categorized in several ways based on their intended function. For instance, the most common types of detection methods, venue and network-based, can be applied to both small networks and large numbers of devices [18]. Host-based intrusion detection programs rely on one system and monitor important operating system files for unusual or malicious activity to find unknown harmful code. In contrast, a detection system checks and monitors network links for suspicious activity. Similarly, attack or cryptography and unusual case detection are very well detection techniques that have undergone substantial investigation by the web security community for a long time. For instance, a very well virus design, sequence, or pattern of bytes in network activity could be considered a signature. Antivirus software uses these design types as a characteristic to match

patterns and discover attacks [19]. This cryptography detection method can easily spot known attacks, but it can be challenging to spot brand-new, unexpected attacks that use well-known signatures but lack any pattern. On the contrary, an unusual case identification system studies the network's activity and identifies patterns. Once the regular behavior has been profiled, it automatically creates a data-driven framework to identify differences in the presence of any abnormalities [20].

Any attacker who tries to send fraudulent communications or obstruct the system's functionality can be stopped [21]. The safety precautions protect against a variety of attacks. However, thorough packet analysis is impossible with these security technologies [22]. As a result, they cannot detect attacks to the appropriate degree. Solutions for detecting attacks have been developed to close the vulnerabilities left by these security measures. These systems can analyze data more extensively than conventional security systems because of their algorithms, which include ML, deep learning, and AI. Machine learning methods are often employed for smaller datasets, but deep learning models work with massive amounts of data. Applying sophisticated DL models on modest, straightforward datasets leads to excessive variance and misleading findings. Numerous privacy-preserving and security assault strategies have been suggested to meet these privacy and security needs [23]. Machine learning techniques have developed and are now widely used in various contexts to reliably identify abnormalities [24]. New test settings are frequently created to raise the precision of network analysis and threat identification. As a result, current data sets are created. Growing Internet usage has led to various security issues. The early intrusion detection system is now achievable through machine learning and network behavior analysis. Due to this, cyber security systems have become the most recent research subjects in both the research and enterprises engaged with cyber security [25].

The key contribution of this paper is listed below:

Machine learning could reduce cyber threats, which can also strengthen security systems. Machine learning is developing to counter new threats, even while cyber-attacks are becoming more numerous and complicated.

- To collect the data from KDD99 databases and it passes through the pre-processing phase.
- The min-max normalization was done in the pre-processing stage, a distinct kernel mapping method that simplifies computations.
- Then, feature extraction is carried out using PSO.
- Followed classification is done by utilizing SVM.

The article's remaining sections are organized as follows: Section 2 discusses the literature survey, while Section 3 provides the methodology. The study and results of the proposed PSO-SVM techniques are shown in Section 4, while the conclusion is presented in Section 5.

2 Related Works

The literature related to this work surveyed by various authors is discussed below:

A key component of ensuring information security is invasion identification, the capacity to find malware and other threats. And so is the capacity to recognize these various threats. A recognized and tested technique for precise categorization is to employ artificial neural networks and other machine learning bio-inspired techniques. Artificial Neural Networks (ANNs) are incredibly adaptable; a variety of configurations can produce categorization outcomes that are noticeably different. This article's major goal and contribution are to assess how the model variables may affect the overall categorization outcome. Several different ANN configurations are compared in this research. The standard databases, NSL-KDD and CICIDS2017, served as the basis for the investigations. The much

more efficient structure yields a multi-class classification performance of 99.909% on a benchmark dataset that has been generated. This study recommends that specific strategies be used to address the issue of data imbalances, as databases usually contain far fewer harmful network patterns than benign ones. The reality that only the more complex and quite well methods are highly optimized at installation and their effectiveness degrades over time is another continuous issue with the quality of information category. This problem is made substantially worse by the rapid advancement of communications networks. A lifetime training strategy might combat this negative situation [26].

Artificial Intelligence (AI) technology, which enables machines to replicate social actions, is among the newest innovations. The Attack Prevention System is crucial in identifying intrusions or other unwanted activity. AI technology is often regarded as the superior strategy for modifying and creating intrusion detection systems and performs a crucial part in attack detection. Neural net techniques are a novel artificial intelligence method that can be used to solve difficulties in the present day. The suggested technology is designed to identify a specific type of botnet intrusion that constitutes a severe risk to banks and economic areas. The suggested approach was developed using artificial intelligence on the most recent intrusion-detecting database (CSE-CIC-IDS2018), generated in 2018 by the Canadian Institute for Cybersecurity. The artificial neural network system that has been suggested has exceptional precision efficiency. The suggested artificial intelligence-based botnet assault detection system is highly efficient and reliable. The newly presented method could be used in machines for real-time network traffic monitoring and traditional network traffic and information from malware systems [27].

Due to people and businesses' growing reliance on the Internet and their worries regarding the safety and confidentiality of their web activity, cyber-security has many emphases. To guard against hostile Internet usage, many prior machine learning-based detection of network intrusions systems was developed. The ML-based NIDS architecture proposed in this study is a new multi-stage optimization method that detects effectiveness while lowering computing effort. The minimum appropriate training example amount is determined by analyzing the effects of oversampling strategies on the training sample size of the algorithms. Additionally, it contrasts the effects of information acquisition and correlation-based feature selection strategies on the rate and intricacy of recognition. Several ML hyperparameter optimization algorithms are also being researched to improve detection systems' effectiveness. The CICIDS 2017 and UNSW-NB 2015 datasets, 2 current intrusion prevention data sources, are used to assess the effectiveness of the suggested system. According to empirical results, the suggested model greatly decreases the character set number and needs training examples.

Additionally, hyper-parameter tweaking improves the algorithm's effectiveness, with identification accuracy and reliability for both databases above 99%. The usage of older databases like NLS KDD99 also places restrictions on them. Only a few research papers also examined the suggested framework's time complexity, an often-overlooked statistic [28].

The widespread use of Internet resources and apps across networked computers has increased intrusions and illicit application utilization, endangering the service's reliability and users' security. Networks Detection System seeks to identify unusual traffic patterns that virus protection is unable to identify. It is been demonstrated that applying the characteristic evaluation technique for dimensional reductions in IDSs increases efficiency. Several bio-inspired algorithms have been used to enhance the effectiveness of IDS by lowering the size of the data set and removing unimportant and chaotic information. The GWO, a customized bio-inspired method, is discussed in this research because it increases the effectiveness of the IDS in identifying both regular and abnormal traffic within the network. The primary enhancements are the clever startup step, which combines the wrappers

with the filtering methods to guarantee that the informative characteristics will be present in the beginning repetitions. Additionally, researchers utilized the improved GWO to fine-tune the Extreme Learning Machine (ELM) settings, a fast categorization technique. Using the UNSWNB-15 dataset, the suggested method was evaluated against different meta-heuristic methods. This paper's main objective was to identify generalized assaults in network activity since they are the most prevalent attack type in the database. Furthermore, The technique would be developed for IDS applications to cope with multi-classification issues and identify many assaults with effective behavior [29].

Deep Learning (DL), which derives from an ANN, is one of the advanced components used in today's modern, smarter cybersecurity methods or regulations. CNN or ConvNet, RNN or LSTM, SOM, AE, RBM, DBN, (DTL or Deep TL, DRL or Deep RL, or their ensembles and hybrid approaches are common deep learning methodologies. In this article, they seek to offer a complete description of these neural networks and deep learning methods in light of the many demands of today's society. They also describe how these methods can be used for cyber security jobs like phishing, malware detection, botnet recognition, intrusion detection, and assault prediction. Furthermore, discuss several research questions and potential approaches that fall under the purview of the experimental investigation. The final goal of this research is to serve as a resource and a source of references for academics and practitioners in the cyberspace sectors, particularly from the perspective of deep learning. Email databases are tough to collect because of concerns about confidentiality and are very challenging to access [30].

Because numerous protocols are used, there are a lot of zero-day assaults happening constantly. The majority of these assaults are merely truncated versions of earlier intrusions. This demonstrates that even the most advanced techniques, like traditional machine learning algorithms, have difficulty detecting these minute changes in assaults over time. To reduce the false alarm rate and increase the recognition accuracy for intrusion detection systems, an efficient semisupervised strategy is suggested in this study by taking into account a few issues with the current intrusion detection systems (IDSs). The suggested method suggests IDS using fivefold cross-validation on semisupervised learning and k-nearest neighbor hyperparameter tuning. The learning set's k-nearest neighbors are first found for every unprocessed data point. The new information is then categorized as regular or assault class based on statistical information obtained via hyperparameter tuning of this nearby information, including the amount of neighboring data points belonging to every potential class, distance measurement, and distance weighting. The NSL-KDD dataset, which is extensively utilized, is used to assess the model's resilience. The simulation results show that the suggested technique works better than IDS-based KNN algorithms. Due to the system's challenging implementation, this approach is ineffective [31]. Table 1 depicts the comparison of various existing approaches.

Table 1: Comparison of various existing approaches

Reference	Dataset	Method	Advantage	Disadvantage
[26]	CSE-CIC-IDS2018	ANN	It detects all types of attacks	Overfitting occurs on the training data
[29]	UNSWNB-15	GWO	Less crossover error rate	It cannot detect different kinds of attacks

(Continued)

Table 1 (continued)

Reference	Dataset	Method	Advantage	Disadvantage
[31]	NSL-KDD	KNN	The parameter does not need to be adjusted	Implementation is challenging
[32]	CICIDS2017	Decision tree	High computation speed	Foretelling the results of a continuous scale
[26]	NSL-KDD	ANN	Multi-class classification accuracy is high	The training process takes a long time

A unique Intrusion Detection System (IDS) that integrates many classifier methodologies, including REP Tree, JRip algorithm, and Forest PA, depends on decision trees and rules-based principles. The initial and second methods use attributes from the data set as inputs and categorize network activity as Attack/Benign. The outcomes of the first and second classifiers are used as inputs for the third classifier, together with characteristics from the initial data set. The empirical findings demonstrate the suggested IDS' superior over existing framework methods in terms of accuracy, detection rate, false alarm rate, and time overhead utilizing the CICIDS2017 dataset. This approach provides the highest DR with accuracy overall, and its quick computing speed makes it simple to include in a soft real-time system. This method is less successful in foretelling the results of a continuous scale [32].

Organizations may best understand their IT vulnerability with the use of threat detection and threat assessment, and they can proactively strengthen their security posture by using the appropriate threat response. Entrepreneurs are better able to anticipate future assaults and events and prevent them. It aids in defending the organization's data, devices, and networks against harmful assaults and online dangers. Employees that participate in training on security awareness learn the value of cyber security and how to recognize possible risks and take proper action.

3 Methodology

The study strategy focuses on developing an intrusion prevention system that can handle the most recent intrusions without using any conventional rule-based techniques. Since current assaults change daily, a system that can be pruned or adjusted to accommodate new attacks is essential. A KDD-CUP 99 database with 41 characteristics was used in the experiment. PSO used a statistical feature selection procedure on the dataset to identify the characteristics that have the greatest potential to improve the performance of the classifiers. Subsequently, the classification was completed using an SVM classifier that efficiently detects suspicious activities.

The KDD-CUP 99 database was utilized for this research to put the suggested ML algorithms into practice [33]. A standard collection of auditable data, including a wide range of simulated intrusions into a defense network environment, is contained in this database. The database is initially inserted into the system, and then pre-processing and characteristic extracting tasks, like cleaning up, and normalizing, are carried out on the database utilizing the particle swarm optimization, which ultimately results in the stages of training and testing. For the learning and evaluation in this investigation, a 60:40 proportion was adopted. The suggested algorithms were SVM and PSO

optimization techniques [34], which categorize the database into anomalous and regular conditions after receiving the 30% validation set. The confusion matrix parameters will be used in the effectiveness assessment of the classifier to calculate the characteristics of the key efficiency factors, such as accuracy, predictive accuracy, precision, false positive rate, and others. Fig. 1 depicts the suggested system's workflow.

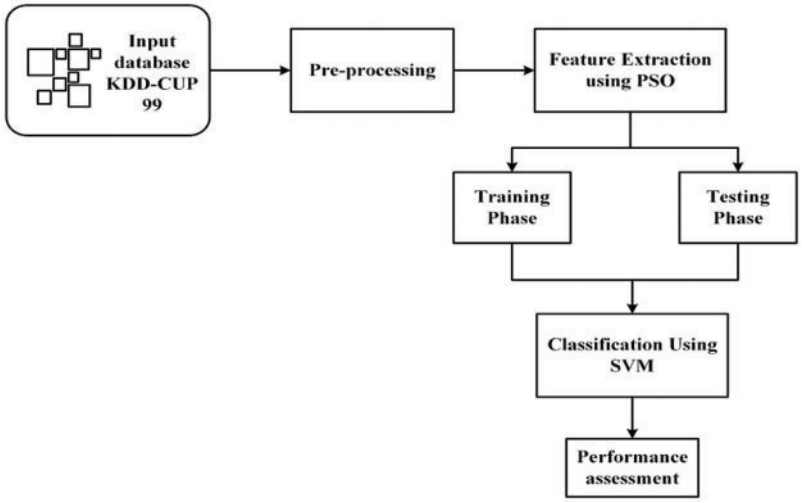

Figure 1: Workflow of the proposed system (PSO-SVM)

The feature-extracted data from Fig. 1 were divided into training and testing phases and categorized using SVM. After that, the performance of the system will be examined.

3.1 Data Collection

The initial standard database for intrusion detection systems was KDD99 from Defense advanced research projects agency. Threats of various types were modeled and categorized under the heading of anomalous. This study used regular traffic and anomalous activity as two different types of network traffic. This database includes flaws while being widely used, as mentioned in utilizing the link: https://datahub.io/machine-learning/kddcup99; you could get the information from the Kaggle repository databases. The database is comprised of 41 characteristics as well as a tag that designates whether an assault is regular or an attacker of a particular sort. 41 characteristics and 494,023 examples are used in the learning phase, whereas 41 characteristics and 148,206 examples are used in the validation process. Cases for the training and testing phases were 345,817 and 148,206, respectively.

3.2 Preprocessing

For categorization, the normalization kind of preprocessing is essential. The input data should be normalized to speed up the learning process. Additionally, some sort of data normalization may be necessary to avoid numerical problems like accuracy loss due to mathematical errors. After initially outnumbering features with originally lower ranges, traits with apparently big ranges would lead to a learning algorithm. Feature space normalization could be considered a kernel impression of preprocessing rather than, strictly speaking, a type of preprocessing because it is not supplied externally to the input vectors. The greatest and minimum values in a normal and assault are different by nine to ten times, for example, in several aspects of intrusion detection metrics. In other terms, by

transforming the data onto a usable plane, normalization is a distinct kernel mapping method that simplifies computations. Due to the enormous amount of data points, the complex normalization algorithm will take much time to execute. The Min-Max normalization technique that was selected is quick and effective.

Using Min-Max Normalization, the real data m is translated linearly into the required interval (max_{new}, min_{new}).

$$n = min_{new} + (max_{new} - min_{new}) * \left(\frac{n - min_y}{max_y - min_y} \right) \tag{1}$$

The method's advantage is that it precisely maintains all interconnections between the data points. There is no chance that it will bias the information in any manner.

3.3 Feature Extraction

An SVM classifier excels at multiclass classification conditions regarding resolution speed and identification rate. The PSO has improved the classifier, which selects the most instructional attributes as classification inputs. A potent thematic optimization technique, Particle Swarm Optimization (PSO), is motivated by swarm behavior seen in nature. PSO simulates a streamlined social structure. The PSO algorithm's original goal was to graphically imitate a flock of birds doing an elegant but unpredictable ballet. Any bird's viewable reach is constrained in nature to a certain area. However, having multiple birds in a swarm enables all birds to be conscious of the greater surface of a fitness function.

PSO is a randomized optimization method used in computational approaches for selecting features. A robust recognition system improves or maintains its classification performance by continuously choosing the most pertinent and practical collection of characteristics to achieve this. Instead of concentrating on one specific class of birds, the fundamental idea underlying this method is the coevolution of many categories of birds.

This algorithm contributes to effective search abilities.

The PSO algorithm is given below:

Make an evenly dispersed "population" of particles over X. Consider the optimal solution while evaluating the positions of each particle using Eq. (2).

$$Z = f(x, y) = \sin x^2 + \sin y^2 + \sin x \sin y \tag{2}$$

Update a particle's position and identify the best particle location. Update the speeds of the particles using Eq. (3).

$$V_j^{t+1} = U.\ V_j^t + b_1 W_1^t \left(P_{c1}^t - P_j^t \right) + b_2 W_2^t (g_c^t - g_j^t) \tag{3}$$

Transporting particles to the new location using Eq. (4).

$$P_j^{t+1} = P_j^t + v_j^{t+1} \tag{4}$$

3.4 Classification Using SVM

An SVM aims to identify the best separating maximum margin of the training data and minimize complexity. It is suitable for analyzing very big datasets and needs a short training dataset to be implemented. Once the ideal categorization hyperplane has been built, an SVM may complete the

classification procedure in a very short amount of time. A trained machine learning model known as a Support Vector Machine (SVM) uses classification methods to complete a binary class job. After being provided sets of labeled training data for each class, an SVM model may categorize new text. Fig. 2 depicts the flowchart of the proposed PSO-SVM system.

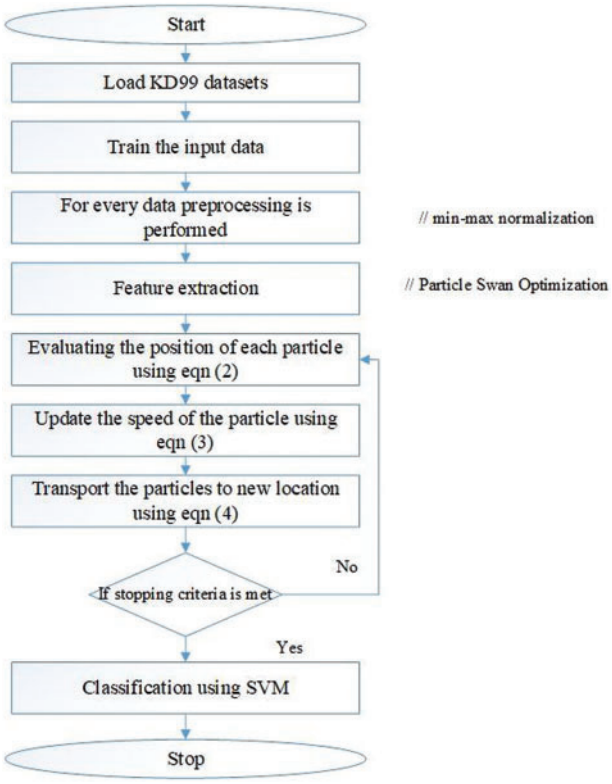

Figure 2: Flowchart of proposed PSO-SVM method

Compared to more current algorithms like learning algorithms, they have two major advantages: increased efficiency and improved results with fewer. This makes the method ideal for categorization jobs. SVM classifiers perform effectively in high-dimensional spaces and have excellent accuracy. SVM classifiers primarily employ a portion of training points, which uses extremely little memory.

The suggested system's entire procedure is shown in Fig. 2. The KDD99 dataset is first loaded and has been pre-processed with min-max normalization. PSO performs the extraction of features, and SVM fulfills the classification process.

4 Results and Discussion

The performance of the proposed system is validated by comparing it with other existing approaches such as KNN, Decision tree, and ANN. Having a single statistic is extremely helpful for assessing the success of a model in machine learning, whether during training, cross-validation, or monitoring after deployment. RMS error is one of the most popular measures for this. It is an adequate scoring system that is easy to understand and in line with some of the most common statistical presumptions. It displays the Euclidean distance between predictions and actual real values. To calculate the root-mean-square error, calculate the residual (difference between forecast and truth) for

each data point and its norm, mean, and square root (RMSE). It is widely used in supervised learning situations since it needs and uses actual measurements at every predicted data point. Fig. 3 represents the performance analysis of the proposed systems, and Table 2 shows the erroneous forecasting of the suggested PSO-SVM system concerning other existing models such as K-nearest neighbors, Decision tree, and ANN. Root mean square error is expressed as follows:

$$RMSE = \sqrt{\frac{\sum_{k=1}^{n} |Z(k) - Z'(k)|^2}{n}} \tag{5}$$

where Z(k) is the kth measurement and $Z'(k)$ is its corresponding prediction, and n is the number of data points.

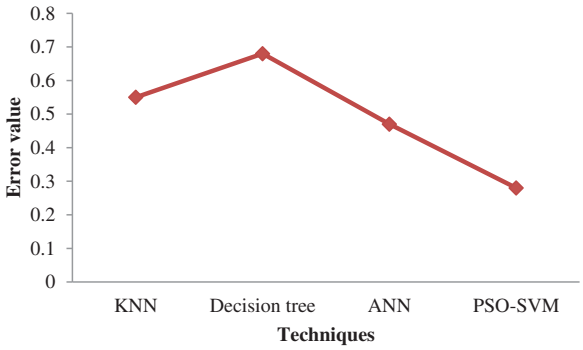

Figure 3: Analyzing error performance of the proposed systems

Table 2: Error prediction of the proposed system with other existing systems

Technique	RMSE
KNN	0.55
Decision tree	0.68
ANN	0.47
PSO-SVM	0.28

Fig. 3 shows that the suggested PSO-SVM's error performance is low, followed by the ANN's minimal value. The decision tree's error value is high when compared to all others.

4.1 Accuracy

The effectiveness of the system model across all categories is assessed using accuracy. It is, generally speaking, the notion that every observation will be correctly predicted. Table 3 depicts the accuracy of the proposed system, and its performance analysis is represented in Fig. 4. Accuracy is given in Eq. (6).

$$Accuracy = \frac{T_{Pos} + T_{Neg}}{T_{Pos} + T_{Neg} + F_{Pos} + F_{Neg}} \tag{6}$$

According to Fig. 4, the suggested PSO-SVM is highly accurate and is followed by the KNN. ANN is less accurate than other techniques.

Table 3: Comparison of accuracy

Methods	Accuracy (%)
KNN	96.33
Decision tree	95.23
ANN	93.99
PSO-SVM	97.09

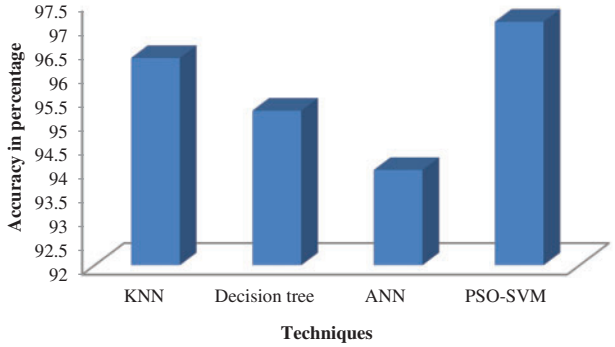

Figure 4: Comparison of accuracy

4.2 Precision

Table 4 depicts the precision of the proposed system, and its performance analysis is represented in Fig. 5. Precision is determined by counting the precise positive evaluations that differ from the total positive evaluations by utilizing Eq. (7).

$$P = \frac{T_{Pos}}{T_{Pos} + F_{Pos}} \tag{7}$$

According to Fig. 5, the suggested PSO-SVM is high precision and is followed by the KNN. ANN is less precise than other techniques.

Table 4: Comparison of precision

Methods	Precision (%)
KNN	89.01
Decision tree	85.69
ANN	81.99
PSO-SVM	91.29

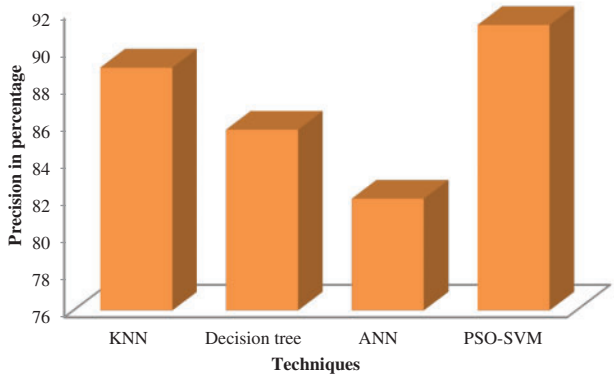

Figure 5: Comparison of precision

4.3 Recall

Table 5 depicts the recall of the proposed system, and its performance analysis is represented in Fig. 6. The recall is the ratio of the total number of positive data to the number of real positives accurately classified as positives. It gives the percentage of the Eq. (8) estimates that were accurate.

$$R = \frac{T_{Pos}}{T_{Pos} + F_{Neg}} \tag{8}$$

According to Fig. 6, the suggested PSO-SVM is a high recall value followed by the KNN. ANN is less recall value than other techniques.

Table 5: Comparison of recall

Methods	Recall (%)
KNN	89.01
Decision tree	85.69
ANN	81.99
PSO-SVM	91.29

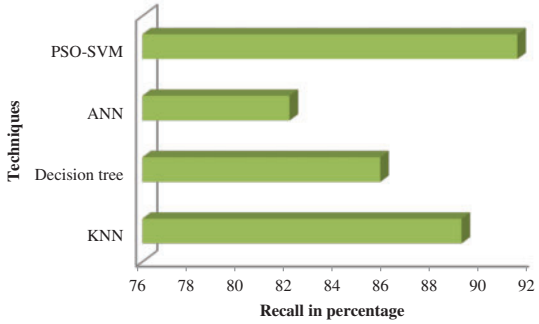

Figure 6: Comparison of recall

4.4 F1-Score

Table 6 depicts the F1-score of the proposed system, and its performance analysis is represented in Fig. 7. The F1-score formula combines precision and recall. Precision and recall construct the F1-score stated in Eq. (9).

$$F1 - score = \frac{2 \times precision \times recall}{precision + recall} \tag{9}$$

According to Fig. 7, the suggested PSO-SVM is a high F1-score value followed by the KNN. ANN is less F1-score value than other techniques.

Table 6: Comparison of F1-score

Methods	F1-score (%)
KNN	89.01
Decision tree	85.69
ANN	81.99
PSO-SVM	91.29

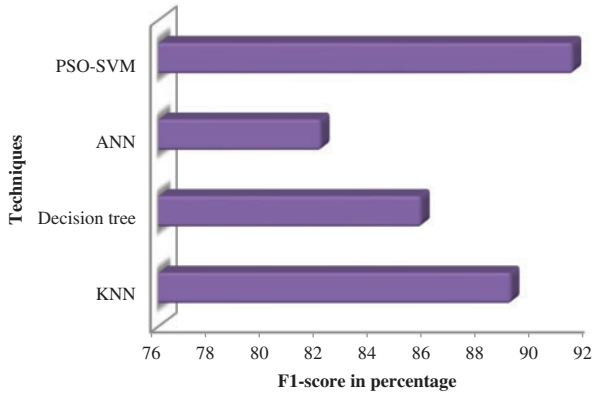

Figure 7: Comparison of F1-score

4.5 Sensitivity

The percentage of true positives the model properly predicted is known as sensitivity. Table 7 depicts the sensitivity of the proposed system, and its performance analysis is represented in Fig. 8. The sensitivity is expressed in Eq. (10).

$$Sensitivity = \frac{T_{Pos}}{T_{Pos} + F_{Neg}} \tag{10}$$

According to Fig. 8, the suggested PSO-SVM is highly sensitive and is followed by the KNN. ANN is less sensitive than other techniques.

Table 7: Comparison of sensitivity

Methods	Sensitivity (%)
KNN	89.01
Decision tree	85.69
ANN	81.99
PSO-SVM	91.29

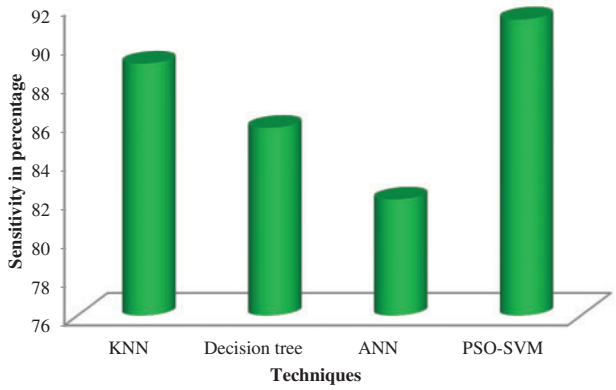

Figure 8: Comparison of sensitivity

4.6 Specificity

Specificity is the percentage of true negatives that the developed model predicted. Table 8 depicts the specificity of the proposed system, and its performance analysis is represented in Fig. 9. The specificity can be calculated by Eq. (11).

$$Specificity = \frac{T_{Neg}}{T_{Neg} + F_{Pos}} \tag{11}$$

According to Fig. 9, the suggested PSO-SVM is high specificity and is followed by the KNN. ANN is less specificity than other techniques.

Table 8: Comparison of specificity

Methods	Specificity (%)
KNN	97.80
Decision tree	97.13
ANN	96.39
PSO-SVM	98.25

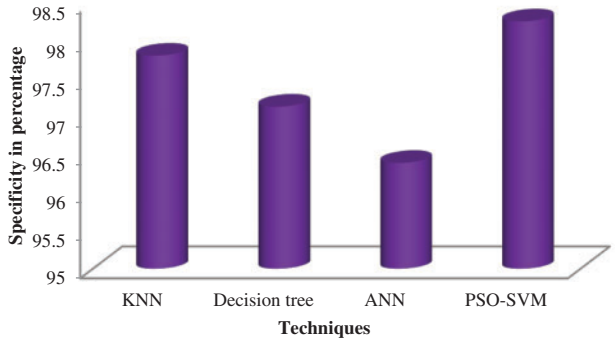

Figure 9: Comparison of specificity

The performance of the suggested system using PSO-SVM is compared to other currently used approaches. The accuracy of the PSO-SVM findings is compared with those of ANN and RNN. The results show that, compared to previous methods, the suggested PSO-SVM has good accuracy. Fig. 10 shows the performance assessment of the proposed system, and Table 9 compares its accuracy to that of other current systems.

According to Fig. 10, the suggested PSO-SVM is highly accurate and is followed by the ANN. RNN is less accurate than other techniques.

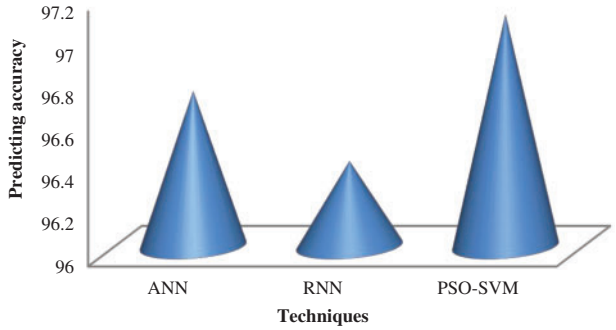

Figure 10: Performance analysis of proposed system accuracy with other existing systems

Table 9: Comparison of proposed system accuracy with the existing system

Reference	Methods	Accuracy (%)
[35]	ANN	96.73
[36]	RNN	96.4
	PSO-SVM	97.09

5 Conclusion

Network intrusion still happens despite the multiple attack detection technologies and strategies available. Attackers and intrusions riskily damage the network's permitted systems using modern

methods. The PSO-SVM model, a novel framework for more accurately spotting intrusions and tracking hacker behavior, is introduced in this study. Various techniques, such as feature extraction, pre-processing, and detection, are used in the experiment. The min-max normalization method is employed during the pre-processing to improve the efficiency of attack identification. The PCO method selects the best characteristics from the dataset during the feature extraction stage. The desirable qualities are enhanced by developing better fitness measurements. The invader and normal traits are classified more accurately using the SVM algorithm. The selected attributes are put through a training and testing process to deliver more precise features. RMSE, recall, specificity, accuracy, and F1-score are a few of the performance indicators analyzed. The comparative evaluation was done with the proposed PSO-SVM with other approaches such as KNN, Decision tree, and ANN. Better performance metrics are produced by the suggested PCO-SVM method. The experimental results support the claim that the suggested technology outperforms currently used methods. Even though the effectiveness of the suggested method has been good, it can be enhanced by additional classifier optimization. The research will be expanded to accurately categorize the huge datasets. Future generations of the techniques will integrate sophisticated optimization with classification algorithms to successfully identify more risks.

Acknowledgement: The authors extend their appreciation to the Deanship of Scientific Research at University of Bisha for funding this research through the general research project under Grant Number (UB-GRP-51-1444).

Funding Statement: The authors extend their appreciation to the Deanship of Scientific Research at the University of Bisha for funding this research through the general research project under Grant Number (UB-GRP-51-1444).

Author Contributions: Nadir Omer conceptualized the study, developed the methodology, and contributed to the original draft. Ahmed H. Samak handled data curation, investigation, and formal analysis. Ahmed I. Taloba conducted data collection, software implementation, and validation. Rasha M. Abd El-Aziz reviewed and edited the manuscript and managed project administration.

Availability of Data and Materials: The data and materials used in this research are available upon request from the corresponding author for academic and research purposes.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] S. Gautam, A. Henry, M. Zuhair, M. Rashid, A. R. Javed *et al.,* "A composite approach of intrusion detection systems: Hybrid rnn and correlation-based feature optimization," *Electronics*, vol. 11, no. 21, pp. 3529, 2022. https://doi.org/10.3390/electronics11213529

[2] S. A. Rahman, H. Tout, C. Talhi and A. Mourad, "Internet of Things intrusion detection: Centralized, on-device, or federated learning?" *IEEE Network*, vol. 34, no. 6, pp. 310–317, 2020. https://doi.org/10.1109/MNET.011.2000286

[3] I. F. Kilincer, F. Ertam and A. Sengur, "Machine learning methods for cyber security intrusion detection: Datasets and comparative study," *Computer Networks*, vol. 188, pp. 107840, 2021. https://doi.org/10.1016/j.comnet.2021.107840

[4] I. H. Sarker, Y. B. Abushark, F. Alsolami and A. I. Khan, "Intrudtree: A machine learning based cyber security intrusion detection model," *Symmetry*, vol. 12, no. 5, pp. 754, 2020. https://doi.org/10.3390/sym12050754

[5] X. Gao, C. Shan, C. Hu, Z. Niu and Z. Liu, "An adaptive ensemble machine learning model for intrusion detection," *IEEE Access*, vol. 7, pp. 82512–82521, 2019. https://doi.org/10.1109/ACCESS.2019.2923640
[6] A. Khraisat, I. Gondal, P. Vamplew and J. Kamruzzaman, "Survey of intrusion detection systems: Techniques, datasets and challenges," *Cybersecurity*, vol. 2, no. 1, pp. 1–22, 2019. https://doi.org/10.1186/s42400-019-0038-7
[7] Z. Muhammad, F. Amjad, Z. Iqbal, A. R. Javed and T. R. Gadekallu, "Circumventing google play vetting policies: A stealthy cyberattack that uses incremental updates to breach privacy," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–10, 2023. https://doi.org/10.1007/s12652-023-04535-7
[8] M. A. Al-Shareeda, S. Manickam and M. A. Saare, "DDoS attacks detection using machine learning and deep learning techniques: Analysis and comparison," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 2, pp. 930–939, 2023. https://doi.org/10.11591/eei.v12i2.4466
[9] R. Vinayakumar, M. Alazab, K. Soman, P. Poornachandran, A. Al-Nemrat *et al.,* "Deep learning approach for intelligent intrusion detection system," *IEEE Access*, vol. 7, pp. 41525–41550, 2019. https://doi.org/10.1109/ACCESS.2019.2895334
[10] A. Mourad, H. Tout, O. A. Wahab, H. Otrok and T. Dbouk, "Ad hoc vehicular fog enabling cooperative low-latency intrusion detection," *IEEE Internet of Things Journal*, vol. 8, no. 2, pp. 829–843, 2020. https://doi.org/10.1109/JIOT.2020.3008488
[11] M. A. Ferrag, L. Maglaras, S. Moschoyiannis and H. Janicke, "Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study," *Journal of Information Security and Applications*, vol. 50, pp. 102419, 2020. https://doi.org/10.1016/j.jisa.2019.102419
[12] Z. Ahmad, A. Shahid Khan, C. Wai Shiang, J. Abdullah and F. Ahmad, "Network intrusion detection system: A systematic study of machine learning and deep learning approaches," *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 1, pp. e4150, 2021. https://doi.org/10.1002/ett.4150
[13] D. Gümüşbaş, T. Yıldırım, A. Genovese and F. Scotti, "A comprehensive survey of databases and deep learning methods for cybersecurity and intrusion detection systems," *IEEE Systems Journal*, vol. 15, no. 2, pp. 1717–1731, 2020. https://doi.org/10.1109/JSYST.2020.2992966
[14] G. Andresini, A. Appice, N. Di Mauro, C. Loglisci and D. Malerba, "Multi-channel deep feature learning for intrusion detection," *IEEE Access*, vol. 8, pp. 53346–53359, 2020. https://doi.org/10.1109/ACCESS.2020.2980937
[15] S. N. Mighan and M. Kahani, "A novel scalable intrusion detection system based on deep learning," *International Journal of Information Security*, vol. 20, pp. 387–403, 2021. https://doi.org/10.1007/s10207-020-00508-5
[16] J. Shu, L. Zhou, W. Zhang, X. Du and M. Guizani, "Collaborative intrusion detection for VANETs: A deep learning-based distributed SDN approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4519–4530, 2020. https://doi.org/10.1109/TITS.2020.3027390
[17] O. R. Shahin, H. H. Alshammari, A. I. Taloba and R. M. Abd El-Aziz, "Machine learning approach for autonomous detection and classification of COVID-19 virus," *Computers and Electrical Engineering*, vol. 101, pp. 108055, 2022. https://doi.org/10.1016/j.compeleceng.2022.108055
[18] A. Kim, M. Park and D. H. Lee, "AI-IDS: Application of deep learning to real-time web intrusion detection," *IEEE Access*, vol. 8, pp. 70245–70261, 2020. https://doi.org/10.1109/ACCESS.2020.2986882
[19] A. I. Taloba, A. Elhadad, R. M. A. El-Aziz and O. R. Shahin, "Prediction of data threats over web medium using advanced blockchain based information security with crypto strategies," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–17, 2021. https://doi.org/10.1007/s12652-021-03109-9
[20] M. Alruily, O. R. Shahin, H. Al-Mahdi and A. I. Taloba, "Asymmetric DNA encryption and decryption technique for Arabic plaintext," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–17, 2021. https://doi.org/10.1007/s12652-021-03108-w
[21] M. A. Al-Shareeda and S. Manickam, "COVID-19 vehicle based on an efficient mutual authentication scheme for 5G-enabled vehicular fog computing," *International Journal of Environmental Research and Public Health*, vol. 19, no. 23, pp. 15618, 2022. https://doi.org/10.3390/ijerph192315618

[22] S. Singh, S. Agrawal, M. Rizvi and R. S. Thakur, "Improved support vector machine for cyber attack detection," in *Proc. of the World Cong. on Engineering and Computer Science*, vol. 1, pp. 394–399, 2011.
[23] M. A. Al-Shareeda and S. Manickam, "MSR-DoS: Modular square root-based scheme to resist denial of service (DoS) attacks in 5G-enabled vehicular networks," *IEEE Access*, vol. 10, pp. 120606–120615, 2022. https://doi.org/10.1109/ACCESS.2022.3222488
[24] F. Shahzad, A. Mannan, A. Javed, A. S. Almadhor, T. Baker *et al.,* "Cloud-based multiclass anomaly detection and categorization using ensemble learning," *Journal of Cloud Computing*, vol. 11, no. 1, pp. 1–12, 2022. https://doi.org/10.1186/s13677-022-00329-y
[25] K. Ghanem, F. J. Aparicio-Navarro, K. G. Kyriakopoulos, S. Lambotharan and J. A. Chambers, "Support vector machine for network intrusion and cyber-attack detection," in *2017 Sensor Signal Processing for Defence Conf. (SSPD)*, London, UK, pp. 1–5, 2017. https://doi.org/10.1109/SSPD.2017.8233268
[26] M. Choraś and M. Pawlicki, "Intrusion detection approach based on optimised artificial neural network," *Neurocomputing*, vol. 452, pp. 705–715, 2021. https://doi.org/10.1016/j.neucom.2020.07.138
[27] V. Kanimozhi and T. P. Jacob, "Artificial intelligence based network intrusion detection with hyper-parameter optimization tuning on the realistic cyber dataset CSE-CIC-IDS2018 using cloud computing," *ICT Express*, vol. 5, no. 3, pp. 211–214, 2019. https://doi.org/10.1016/j.icte.2019.03.003
[28] M. Injadat, A. Moubayed, A. B. Nassif and A. Shami, "Multi-stage optimized machine learning framework for network intrusion detection," *IEEE Transactions on Network and Service Management*, vol. 18, no. 2, pp. 1803–1816, 2020. https://doi.org/10.1109/TNSM.2020.3014929
[29] A. Alzaqebah, I. Aljarah, O. Al-Kadi and R. Damaševičius, "A modified grey wolf optimization algorithm for an intrusion detection system," *Mathematics*, vol. 10, no. 6, pp. 999, 2022. https://doi.org/10.3390/math10060999
[30] I. H. Sarker, "Deep cybersecurity: A comprehensive overview from neural network and deep learning perspective," *SN Computer Science*, vol. 2, no. 3, pp. 154, 2021. https://doi.org/10.1007/s42979-021-00535-6
[31] R. Wazirali, "An improved intrusion detection system based on KNN hyperparameter tuning and cross-validation," *Arabian Journal for Science and Engineering*, vol. 45, no. 12, pp. 10859–10873, 2020. https://doi.org/10.1007/s13369-020-04907-7
[32] A. Ahmim, L. Maglaras, M. A. Ferrag, M. Derdour and H. Janicke, "A novel hierarchical intrusion detection system based on decision tree and rules-based models," in *2019 15th Int. Conf. on Distributed Computing in Sensor Systems (DCOSS)*, Santorini, Greece, pp. 228–233, 2019. https://doi.org/10.1109/DCOSS.2019.00059
[33] M. Elloumi, M. A. Ahmad, A. H. Samak, A. M. Al-Sharafi, D. Kihara *et al.,* "Error correction algorithms in non-null aspheric testing next generation sequencing data," *Alexandria Engineering Journal*, vol. 61, no. 12, pp. 9819–9829, 2022. https://doi.org/10.1016/j.aej.2022.03.041
[34] S. Budilaksono, A. A. Riyadi, L. Azhari, D. D. Saputra, M. A. Suwarno *et al.,* "Comparison of data mining algorithm: PSO-KNN, PSO-RF, and PSO-DT to measure attack detection accuracy levels on intrusion detection system," *Journal of Physics: Conference Series*, vol. 1471, no. 1, pp. 012019, 2020. https://doi.org/10.1088/1742-6596/1471/1/012019
[35] A. Subroto and A. Apriyana, "Cyber risk prediction through social media big data analytics and statistical machine learning," *Journal of Big Data*, vol. 6, no. 1, pp. 50, 2019. https://doi.org/10.1186/s40537-019-0216-1
[36] T. Teoh, G. Chiew, Y. Jaddoo, H. Michael, A. Karunakaran *et al.,* "Applying RNN and J48 deep learning in android cyber security space for threat analysis," in *2018 Int. Conf. on Smart Computing and Electronic Enterprise (ICSCEE)*, Shah Alam, Malaysia, pp. 1–5, 2018. https://doi.org/10.1109/ICSCEE.2018.8538405