International Journal of Communication Networks and Information Security
ISSN: 2073-607X, 2076-0930
Volume 14 Issue 3 Year 2022 Page 358:371

Performance Evaluation of an Intelligent and Optimized Machine Learning Framework for Attack Detection

[1] Ghayth ALMahadin, [2] Mohammad O. Hiari, [3] Abdelrahman H. Hussein, [4] Nidal Mahmoud Mustafa Turab, [5] Ashraf Alkhresheh, [6] Mutaz A. B. Al-Tarawneh
[1]Assistant Professor, Department of Networks and Cybersecurity, Faculty of Information Technology / Al Ahliyya Amman University, Jordan.
[2]Lecturer, Department of Networks and Cybersecurity, Faculty of Information Technology / Al Ahliyya Amman University, Jordan.
[3]Associate Professor, Department of Networks and Cybersecurity, Faculty of Information Technology / Al-Ahliyya Amman University, Jordan
[4]Professor, Department of Networks & Cyber Security, Faculty of Information Technology/AL-Ahliyya Amman University, Jordan.
[5]Assistant Professor, Department of Computer Science, ITC / Tafila Technical University, Jordan.
[6]Professor, Department of Computer Engineering Department, Engineering / Mutah University, Jordan.
[1] g.mahadin@ammanu.edu.jo, [2] m.hyari@ammanu.edu.jo, [3] a.husein@ammanu.edu.jo, [4]N.turab@ammanu.edu.jo, [5] khashraf@ttu.edu.jo, [6] mutaz.altarawneh@mutah.edu.jo

Article History

Received: 13 July 2022
Revised: 20 September 2022
Accepted: 26 October 2022

Abstract

In current decades, the size and complexity of network traffic data have risen significantly, which increases the likelihood of network penetration. One of today's largest advanced security concerns is the botnet. They are the mechanisms behind several online assaults, including Distribute Denial of Service (DDoS), spams, rebate fraudulence, phishing as well as malware attacks. Several methodologies have been created over time to address these issues. Existing intrusion detection techniques have trouble in processing data from speedy networks and are unable to identify recently launched assaults. Ineffective network traffic categorization has been slowed down by repetitive and pointless characteristics. By identifying the critical attributes and removing the unimportant ones using a feature selection approach could indeed reduce the feature space dimensionality and resolve the problem.Therefore, this articledevelops aninnovative network attack recognitionmodel combining an optimization strategy with machine learning framework namely, Grey Wolf with Artificial Bee Colony optimization-based Support Vector Machine (GWABC-SVM) model. The efficient selection of attributes is accomplished using a novel Grey wolf with artificial bee colony optimization approach and finally the Botnet DDoS attack detection is accomplished through Support Vector machine.This articleconducted an experimental assessment of the machine learning approachesfor UNBS-NB 15 and KDD99 databases for Botnet DDoS attack identification. The proposed optimized machine learning (ML) based network attack detection framework is evaluated in the last phase for its effectiveness in detecting the possible threats. The main advantage of employing SVM is that it

CC License CC-BY-NC-SA 4.0	offers a wide range of possibilities for intrusion detection program development for difficult complicated situations like cloud computing. In comparison to conventional ML-based models, the suggested technique has a better detection rate of 99.62% and is less time-consuming and robust. ***Keywords: Attack detection, Machine Learning Grey wolf optimization, Artificial Bee Colony Optimization, Support Vector Machine***

1. Introduction

Network technology have advanced significantly over the past 20 years, but at the same time, risks towards securing the networks have grown tremendously. Some of the destructive cybercrimes include denial of service (DoS), web-based security assaults, and insider threats [1]. Such malevolent acts have the potential to seriously damage a network. Since several users are submitting requests to a central server during a distributed denial of service (DDoS) assault, the server is unable to provide the users with the necessary solutions as a result of the server's high consumption of resources [2].Botnets causes a number of security problems, including DDoS assaults, spam distribution, setting up snap-bending traps, stealing customer information, and misusing powerful computing resources [3]. A botnet is a collection of computers linked to the Internet which have been remotely manipulated and programmed to do harm by an invader known as a bot-master.
As shown in Fig. 1, the attacker accesses the controlling server to establish commands and takes the entire control of the network. A Control Server is a strong system having plenty of resources, including memory, processing power, and bandwidth. The operators, also known as Agents, are in charge of accepting orders from an intruder and monitoring Botnets. They communicate with the Botnets by sending commands for setup and updates. The real user of the infected computing device does not know if their machine is a member of the Botnet or has malwares set up on it. Attacks on the victims are launched by the attackers using the agents as a desk jump [4].

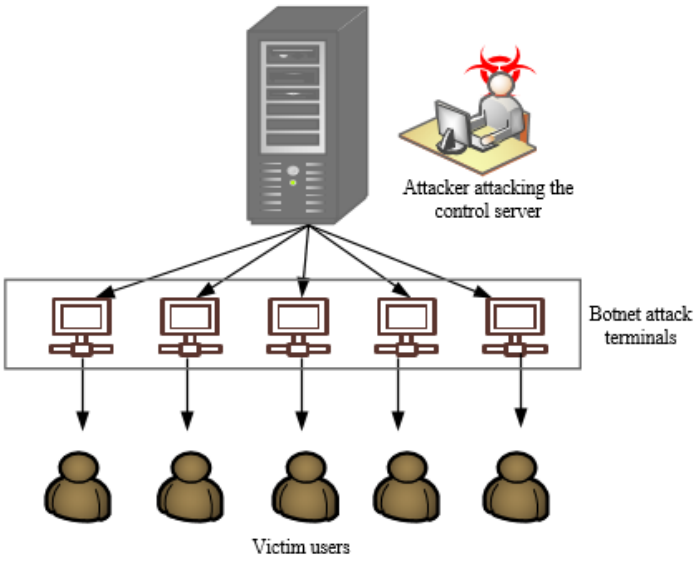

Figure 1: *Botnet DDoS Attack*

Numerous countermeasures against network assaults were already put out in the literary works to deal with this problem. The network security issue has not been fully resolved despite the numerous efforts made by researchers during the past 20 years. Therefore, a strong security system is required to protect from numerous assaults like denial-of-service (DoS) assaults, viruses, malwares as well as bugs [5]. User authentication, firewalls, and data encryption are among the initial lines of security

protection, but they are insufficient to protect the security demands of the whole system along with a persistent intrusion techniques [6].Consequently, a more advanced security strategy, like an attack detection framework has been implemented to assure security of the network. Most commonly utilized network attack detection models continually scan network data for such suspicious or illegal activity in order to identify intruders [7]. The network administrators are alerted by the intrusion detection system (IDS) concerning invasive efforts. Detecting various network vulnerabilities like distributed denial-of-service (DDoS) assaults, viruses & worms, are also helpful [8].

An effective defence system's central component is an intrusion detection system that enables accurate attack detection prior to any response. An intrusion detection model seeks to identify attacks before they cause significant network harm [9]. Currently, flexibility should be taken into account as a crucial necessity in both methods due to the rapid expansion of technology and the expanded availability of attack vectors. An automated or at least semi-automated detection phase is essential in order to respond to network attacks as quickly as feasible [9]. The harm to genuine users would be lessened as a result since an early identification system provides enough time for a suitable response. Numerous methods, including knowledge-based, statistical, and artificial intelligence-based methods, have been used to detect the intrusion in a network [10].

In this study, a Botnet DDoS assault detection system based on wrapped attributes such as Source byte, Smean, Destination byte, Dmean, duration flag, etc. is utilized to identify anomaly-based assaults. This mechanism employs support vector machine technique along with the optimization strategies. Machine learning-based IDS have the ability to monitor massive amounts of data in such a manner that it manages itself and identifies assault behaviours considerably more effectively. The intrusion detection databases accessible at public sourcesincorporating DDoS assaults are most commonly taken for assessment while constructing the machine learning mechanisms for the optimal machine learning-driven identification of DDoS assaults.The entire study is plotted on a suggested deliberated approach, which results in a thorough plan for eliminating data-related inherent issues and significantly lowering processing overhead.

The key contributions of thenovel proposed optimized feature selection-based attack detection system are

- Creating a feature-based attack detection framework that uses machine learning techniques to filter down the important attributes from the broad aspects and to simplify the process.
- Initially, both the KDD99 and UNBS-NB 15 databases are trained in the novel GWABC-SVM model
- Pre-processing is accomplished using Max-min normalization for maintaining the relationship among the source data values.
- Followed by this, feature selection is accomplished using a hybrid optimization technique called Grey Wolf with Artificial Bee Colony (GWABC) model
- Then, the optimization algorithms are integratedwith Support Vector Machine (SVM) model, thereby creatinga new GWABC-SVM model for better detection of attacks in network.
- At last, the evaluation process is accomplished to show the efficiency of the developedmodel compared to the existing mechanism in attack detection in a network.

The residual part of this study is described as mentioned below: Section 2 of this paper discusses the recent literature on network intrusion detection system. The problem description is presented in Section 3. The suggested optimized ML framework's approach and description are then covered in detail in Section 4. Sect. 5 includes a brief explanation with the comparison of the proposed work with current approaches, and the experimental findings. In Section 6, a summary of the report and recommendations for further research are given, along with a conclusion.

2. Related Works

Lima Filho et al.[11]developed a DoS recognizer that uses machine learning (ML) model. The method relies on signatures that were previously retrieved from examples of packet headers to draw conclusions. Utilizing four contemporary benchmark datasets, the tests were conducted. Using the samples obtained by the sFlow standard straightforwardly from connected systems, the application employs the RFT method to categorise network activity. The results of the developed strategy shows

that the Smart Identification technique offers enhanced detection accuracy, false alarm rate and precision rate. However, the approach requires several enhancements, such as an increase in the hit rate for assault classes and an automated parameter calibrating technique which maximises attack recognition performance.

In order to identify Botnet traffic, Homayoun et al. [12]employed the Botnet Traffic Shark, which is a Botnet traffic analyzer that uses deep learning methodology. The analyzer's operation relies on network transactions rather than the deep packet inspection approach. The metrics like False Positive Rate and True Positive Rate are utilized for evaluating the systemperformance. The research demonstrates that autoencoders outperform convolution neural networks owing to its less false positives produced. For an effective detection using deep learning, a significant quantity of information is needed. The necessary processing power is likewise very high.

The Random Forest (RF) and K-nearest Neighbours (KNN) classifiers were used to examine the outcomes of various ML hyper-parameter adjustment strategies on the accuracy of the NIDS in [13]. The minimum adequate training sample size was calculated by analysing the effects of resampling strategies on the development study population of the algorithms. To boost the performance of the NIDS, numerous ML hyper parameterization strategies are being researched. In order to build on the earlier work, this research utilized a multi-stage optimised Machine Learning-based intrusion detection system architecture that increased detection accuracy while lowering powerful cryptography. However, only non-linear and high-dimensional information enable these models to deliver excellent results.

The authors in [14]employed a distributed deep learning technique for detecting the malware in a IoT or Fog networks. The research sought to identify dangers in the social internet of things through deep learning, a novel cyber defence technique. This indicates that even sophisticated methods, like those used in conventional machine learning methods have difficulty in identifying vulnerabilities that have gradually changed. It was also shown that, while examined on previously untested test data, the deep approach surpasses more conventional machine learning methods like Softmax in categorizing network activity into legitimate and malignant. This methodology offers a more accurate distinction between regular and malicious network traffic. However, power consumption increases if a layer is placed in between the cloud and the server. Therefore, this approach is ineffective for the usage at hand.

A unique hybrid attack recognition approach comprising 2 stages namely, the feature selection stage and an attack recognition stage is described by Hosseini et al. in their study [15]. A wrapper approach called MGA-SVM is applied during the feature selection stage. The characteristics of Genetic algorithm and SVM such as multi-parent mutation and multi-parent crossover are combined in this method. An artificial neural network (ANN) is utilized for identifying assaults during the attack recognition stage. The classifier is trained using particle swarm optimization and hybrid gravitational search approach to enhance its effectiveness. It has a maximum detection performance of 99.3%, and it reduces the NSL-KDD dimension from 42 to four attributes, and it requires only three seconds for training.

To identify distributed denial-of-service (DDoS) threats, Aamir [16]provides a strategic-level architecture that combines the essential components of machine learning and feature engineering with a specified pipeline of experiment. Utilizing feature selection techniques like backward elimination, chi-squared, and information gain scores, feature engineering focuses on obtaining datasets of various dimensions with meaningful attributes. To illustrate the flexibility of databases for machine learning under the best parameter tuning in provided sets of values, various supervised machine learning algorithms are used on the attribute-based databases. It offers thorough solutions that could be relied upon to prevent collinearity issues and data overfitting problem. Moreover, it offers minimal processing overhead during identifying DDoS assaults. However, utilizing this strategy results in a less efficient feature selection procedure.

3. Problem Statement

There is obviously a need for a strategic approach to implement it in a structured way in order to minimize normal inherent problems with machine-mined data, like multicollinearity, collinearity, and duplication. This is true even after numerous study endeavours for identifying

DDoS assaults with machine learning mechanisms, some of which also take feature engineering into account. Additionally, when deploying machine learning algorithm, it is necessary to take into account every single significant need of data science-driven techniques.Simply implementing a model with default settings could not accomplish what is needed and instead bring overfitting-related issues. Integrating feature engineering, optimizations, and machine learning techniques into a coherent system, would be a solution to overcome all those issues. Thus, an efficient optimized machine learning framework is created in this work. Following is a thorough explanation of the proposed work.

4. Materials and Methods

4.1 Datasets description

This research uses the reference UNBS-NB-15 [19]database, among the newest and most popular databases, to verify the effectiveness of techniques. Hence, it efficientlyshows both conventional network traffic and a number of network attacks launched by botnets, which is represented in Table1.

Table.1: *UNBS-NB-15 database*

Features selection	Ranking	Feature description
Dbytes	0.491	Server-client transactions
Sload	0.464	Client (bits-per-second)
Sttl	0.444	Client-server (time)
Rate	0.429	Rate
Dmean	0.406	Mean packet size transferred by server
Sbytes	0.642	Client-server transactions
Smean	0.477	Mean packet size transferred by client
ct-state-ttl	0.454	Count
Dttl	0.439	Server-client (time)
Dur	0.409	Countdown time period

The database was developed by utilising the IXIA Perfect Storm tool that causes both acceptable client and attacker traffic [17], [18] that was then sub-divided into nine sections: Backdoor, DoS, Exploits, Reconnaissance, Generic, Shellcode, Fuzzers, Worms, and Analysis. Additionally, the KDD99 database [19]is used as a different testing source, which is depicted in Table2 (The CAIDA UCSD Dataset 2008-11-21).

Table.2: *KDD99 database*

Features	Descriptions
Protocol-type	Protocols that are utilized in a connectivity
Flag	Connection's status flag
Dst-bytes	Sending data (in bytes) from destination-source
Urgent	Urgent packets number (source-destination), etc.
Duration	Connection time (seconds) source-destination
Service	Services at the destination
Src-bytes	Sending data (in bytes) from source-destination
Wrong-fragment	Wrong-fragment lists (sender-receiver)
Land	1-connection from same source; otherwise-0

4.2 Proposed GWABC-SVM model

The currently available methods have a number of drawbacks in terms of scalability, huge databases, complex data, reliability, slow findings, etc. Therefore, it is necessary to think about a method that can handle all of these problems in a far more practical way. One such strategy combines a number of methods, and it is called machine learning (ML). The real-time detection of Botnet DDoS-attack is accomplished using the Grey Wolf with Artificial Bee Colony optimization-based Support Vector Machine (GWABC-SVM) model. The benefits of these techniques and their suitability for the type of data researchers are working with for effectiveness evaluation are the basis for the study's consideration of this SVM technique.This strategy addresses the problem by taking into account the hybrid optimization using ML algorithms that are very scalable. The majority of the strategies proposed in the past struggle with robustness and scalability. The block diagram of the presented model is shown in fig.2 and the components are explained as follows.

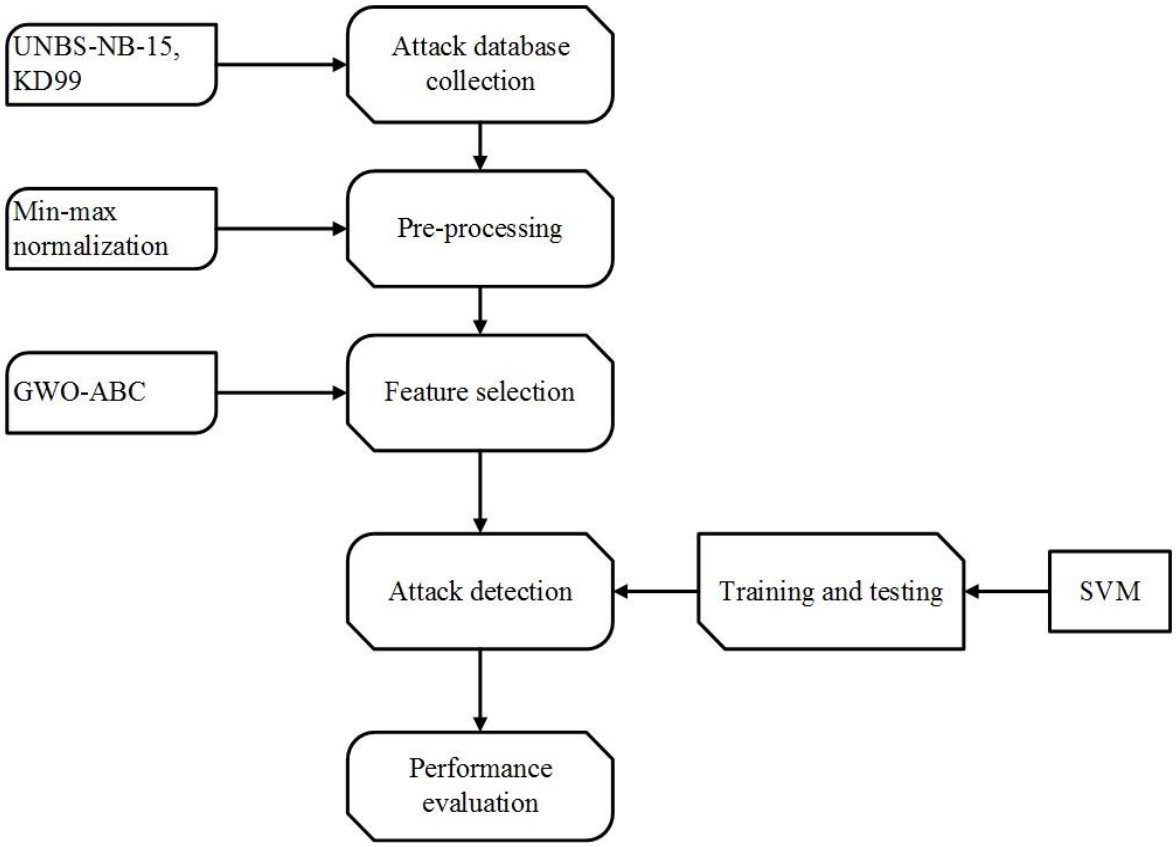

Figure 2: *Proposed GWABC-SVM model*

4.2.1 Min-max normalization-based pre-processing

In a designed system, the dataset has been pre-processed using min-max normalisation. Knowledge stream-mining can make use of knowledge normalisation as a knowledge preparation method. By increasing the numbers until they fall inside the desired points, like within 1.0, the database's correlate in nursing number is changed. Min-Max normalisation is used to do linear transformation starting from the initial step. A value p of d is transformed via min-max normalisation into p' within the range $[new_{\min(d)}, new_{\max(d)}]$. Prior applying the normalisation process to the database, the module computes tuples with incomplete data by recommending one of the numerous options, such as the majority, minimum, mean, constant, and variance. The min-max normalisation is computed by Eqn. (1):

$$p' = \left[p - \min(d)\right] \times \left[new_{\max(d)} - new_{\min(d)}\right] \Big/ \left[\max(d) - \min(d)\right] + new_{\min(d)} \quad (1)$$

Where, $\min(d)$= minimum attributevalue, and $\max(d)$= maximum attributevalue. Eqn. (2) displays the formula reduced for determining the social control.

$$p' = {p - \min(d)}\Big/{\max(d) - \min(d)} \quad (2)$$

A minimum and maximum normalisation (min-max normalisation) maintains the relationship between the values of the source data.

4.2.2 Feature selection by hybrid GWOABC mechanism

GWO seems to be a technique for bionic optimization. It imitates grey wolf hunting behaviour, with a definite division of work and mutual collaboration. Grey wolves normally live in a pack of 5-12 individuals and have a strict dominating hierarchy predicated on wolf leadership abilities. The GW pack's predatory procedure can be separated into three stages: hunting, encircling, and attacking. The group was usually led by the most notable wolf, which is referred to as wolf $\propto$. In GWO, the 2nd and 3rd levels of leadership wolves have been referred to as β and γ wolves, respectively. These subsidiary wolves in the 2nd and 3rd ranks help the wolf in determining hunting decisions. All other accompanying wolves have been labelled as wolves, and they chase and attack the prey with these high-ranking wolves[20].

For hunting, the enclosing of prey approach is used. For iteration i, the mathematical framework for this method is shown in below Eqn. (3) and (4).

$$\vec{F} = \left|\vec{Y} \times \overrightarrow{Q_p}(i) - \vec{Q}(i)\right| \quad (3)$$

$$\vec{Q}(i+1) = \overrightarrow{Q_p}(i) - \vec{D}.\vec{F} \quad (4)$$

Here, $\vec{D}$ and $\vec{Y}$ are coefficient vectors, which is described as $\vec{D} = 2\vec{d}.\overrightarrow{v_1} - \vec{d}$ and $\overrightarrow{Y} = 2.\overrightarrow{v_2}$. Where, the random vectors $\overrightarrow{v_1}, \overrightarrow{v_2}\ \varepsilon\ (0,1)$ and $\overrightarrow{d} = d_1\left(1 - {i}/{maxi}\right)$, linearly decreases from d_1 to zero; d_1 value was set as 2 in actual GWO. Moreover, $maxi$ represents maximum number of iterations. The GWO's hunting process has been headed by three finest solutions i.e., $\propto, \beta$ and γ wolves. Thus, these 3 leading solution's positions have been saved in the pack and the remaining ω wolves update their positions predicated on them. This position updating technique's mathematical model is represented in Eqn. (5).

$$\vec{Q}(i+1) = {\left(\overrightarrow{Q_1} + \overrightarrow{Q_2} + \overrightarrow{Q_3}\right)}\Big/{3} \quad (5)$$

Where, $\overrightarrow{Q_1}, \overrightarrow{Q_2}, and\ \overrightarrow{Q_3}$ is computed by Eqn. (6)

$$\begin{aligned} \overrightarrow{Q_1} &= \overrightarrow{Q_\propto}(i) - \overrightarrow{D_1}.\overrightarrow{F_\propto} \\ \overrightarrow{Q_2} &= \overrightarrow{Q_\beta}(i) - \overrightarrow{D_1}.\overrightarrow{F_\beta} \\ \overrightarrow{Q_3} &= \overrightarrow{Q_\gamma}(i) - \overrightarrow{D_1}.\overrightarrow{F_\gamma} \end{aligned} \quad (6)$$

Here, $\overrightarrow{F_\propto}$, $\overrightarrow{F_\beta}$, and $\overrightarrow{F_\gamma}$are computed by Eqn. (7)

$$\begin{aligned} \overrightarrow{F_\propto} &= \left|\overrightarrow{Y_1} \times \overrightarrow{Q_\propto}(i) - \vec{Q}\right| \\ \overrightarrow{F_\beta} &= \left|\overrightarrow{Y_2} \times \overrightarrow{Q_\beta}(i) - \vec{Q}\right| \\ \overrightarrow{F_\gamma} &= \left|\overrightarrow{Y_3} \times \overrightarrow{Q_\gamma}(i) - \vec{Q}\right| \end{aligned} \quad (7)$$

The artificial bee colony (ABC) seems to be probably of the newest algorithms, inspired by honey bees' clever foraging behaviour[21]. In ABC technique, colony of artificial bee is made up of three types of bees: bystanders, employed bees, and scouts. A food sources is a potential solution to the

issue that has to be solved. The quantity of nectar in a food resource correlates to the excellence of the solutions the food resource represents. There was only one employed bee for each source of food. In other terms, the amount of engaged bees was proportional to the number of sources of food in the immediate vicinity of the hive. The employed bee becoming a scout once the bees depart their food source. Onlooker bees investigate their new surroundings depending on the data. The search technique for both engaged and observer bees in ABC are driven by upgrading a random component in the solutions vector with another solution vector as shown in Eqn. (8).

$$b_{k,l} = \vec{Q}(i+1)_{k,l} + \varphi_{k,l}\left(\vec{Q}(i+1)_{k,l} - q_{i,j}\right) \quad (8)$$

Where, b_{kl} represents the new solution attained by mutating l^{th} dimension parameter of two distinct solutions in a pack and φ_{kl} represents the random number, which varies between -1 and 1.
Though ABC's updating technique improves exploration, it falls short of leveraging the optimal solution's knowledge. It has been recognized that the ABC technique functions differently from other population-based techniques such as GWO since it doesn't use the best answers to guide the search operation. The techniques convergence rate may suffer as an outcome of this. The optimum solution data has been found to play a vital influence in enhancing convergence efficiency. Because GWO makes use of the finest solutions in the organizational hierarchy, combining it with ABC will result in a powerful algorithm that combines the benefits of both. GWO might be a target-hunting wrappers quality reduction system that takes advantage of detection performance of wrapper-based approaches and the effectiveness of filter-based methods. So, this GWO operates in two stages: The mutual data is maximised by any attribute combination (δ), according to GWO, which is represented in Eqn. (9).

$$\delta = V - d \quad (9)$$

Where, the data, which is available in a mutual state has been referred to as optional (V), and P seems to be the middle value of that mutual data amongst chosen attributes like Dst-bytes, Src-bytes, Dmean, and Smean from UNBS-NB-15 database and Dbytes, Sbytes, Duration, and Flag from KDD99 database.
The values of the qualities chosen are determined by how well the class label is classified, and they can be autonomous, as seen by the fitness function in Eqn. (9). The GWO convergence relies on the fitness function (Mutual information computation) utilised and the exploring ability. After the specified number of iterations, this operation is halted.

4.2.3 DDoS attack detection using SVM

Support Vector Machines (SVMs) have been learning tools that exhibit the training variables in a higher-dimensional feature space and annotate each vector with a class. In order to categorise data, SVMs determine a series of support vectors, which have been training input set's components and delineate a hyperplane in the subspace. Furthermore, in SVMs the classes are in the form of hyperplane and that isrepresented in Eqn. (10).

$$S.G + b = 0 \quad (10)$$

Where, S is the weight of the vector, G is the input vector, and $b$represents the bias.
SVMs don't really need mitigation in the quantity of features in attempt to avoid over-fitting, which is an evident benefit in applications like intrusion detection. Instead, the number of free variables employed in SVMs varies based on the limit that partitions the data-points rather than on the amount of input characteristics. The low predicted frequency of generalisation errors seems to be another key benefit of SVMs[22]. The input features for SVMs for DDoS attack detection have been taken from raw Ttcpdump, which could also collect information as normal in attacks directed at intranet system resources. The result seems to be a single value, which denotes whether the sequence is indeed a DDOS attack or not. SVM has been trained on several attack types as well as regular data. It is divided into two classes: normal (-1) and DDOS attack data (+1). Performance testing is done on the learned SVM to make sure it has developed the necessary detection skills.

5. Result and Discussion

The UNBS-NB 15 and KDD 99, which are popular public databases used in the performance assessment part for Botnet DDoS attack identification. On the basis of performance parameters like Accuracy, False Alarm Rate (FAR), Recall, as well as Specificity, the performance of a novel suggested GWABC-SVM model is assessed. The metrics shown above are determined based on the factors listed below.

True Positive (tp') - The model appropriately identified the class attribute where the attack was identified

True Negative (tn'): The class attribute's value is negative, or "normal traffic"

False Positive (fp'): When the model misidentifies regular traffic as an assault

False Negative (fn'): The model misclassifies attack records as regular traffic

5.1 Accuracy

The likelihood that a record which might be either an attack or regular traffic is correctly recognized may serve as a proxy for accuracy. Eqn. (11) is used to assess the assault detection accuracy.

$$Accuracy = \frac{tp'+tn'}{tp'+fp'+tn'+fn'} \quad (11)$$

5.2 False Alarm Rate

False Alarm Rate, often known as FAR, is the likelihood that a record was being wrongly categorized. Eqn. (12) is used to estimate it.

$$FAR = \frac{fp'+fn'}{tp'+fp'+tn'+fn'} \quad (12)$$

5.3 Recall

Another name for recall is sensitivity. It calculates the percentage of "attacks" that are accurately identified as such in relation to all "attacks." Eqn. (13) is used to determine it.

$$Recall = \frac{tp'}{tp'+fn'} \quad (13)$$

5.4 Specificity

The likelihood of test assaults without producing false positive findings is represented by specificity. Eqn. (14) is used to compute it

$$Specificity = \frac{tn'}{tn'+fp'} \quad (14)$$

Table 3: *Comparison of Performance measures with and without optimization (GWABC)*

	Avg. Detection Accuracy	**Avg. False Alarm rate**	**Avg. Recall**	**Avg. Specificity**
Without optimization (SVM)	91.5%	5.3%	91.2%	94.6%
With Optimization (GWABC-SVM)	99.62%	0.85%	97.9%	97.02%

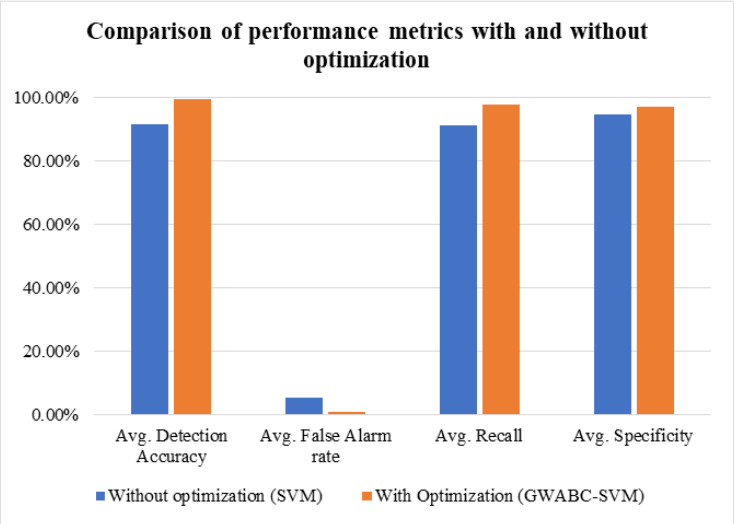

Figure 3: *Graphical representation of comparison of performance measures with and without GWABC*

The Table 3 illustrates the comparison of proposed ML mechanism with and without including optimization mechanism. The outcomes shows that the proposed SVM model with hybrid GWABC mechanism (GWABC-SVM) achieved very high detection accuracy (99.62%) and very low FAR (0.85%) compared to traditional machine learning model.

Table 4:*Performance evaluation for UNBS-NB 15 database for proposed and existing methods*

Method	Detection accuracy	False Alarm rate	Recall	Specificity
ANN[23]	97.44%	2.56%	84.89%	15.11%
IRIDS [24]	90.32%	2.01%	5%	50.37%
GWABC-SVM (Proposed)	99.74%	0.5%	98.38%	97.72%

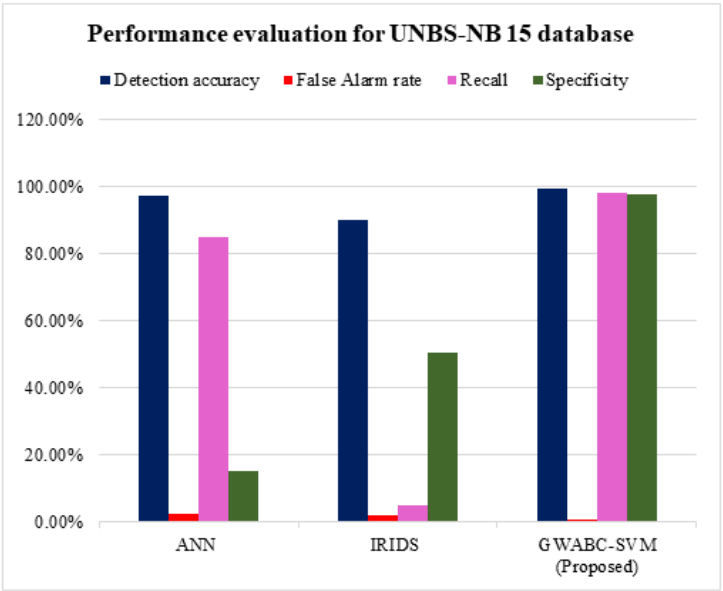

Figure 4: *Comparison of performance measures for existing and proposed method for UNBS-NB 15 database*

The Table 4 depicts the performance evaluation for UNBS-NB15 dataset. It is obvious from the outcomes of performance assessment that the detection accuracy for Proposed (GWABC-SVM) is higher (99.74%) when compared to all other existing techniques like Integrated-rule based intrusion detection system (IRIDS) and Artificial Neural Network (ANN). It is also observed that the proposed method achieves very low FAR value of 0.5% whilst IRIDS and ANN achieves 2.01% and 2.56% respectively. Moreover, the proposed model shows goodoutcomes in terms of recall and specificity while comparison with the existing techniques. The figure 4 shows the shows the pictorial description of comparison of prevailing and proposed techniquesfor UNBS-NB 15 database.

Table 5: *Performance evaluation for KDD99 database for proposed and existing methods*

Method	**Detection accuracy**	**False Alarm rate**	**Recall**	**Specificity**
HGWCSO-ETSVM [25]	99.2%	2.42%	79.5%	94.3%
RNN [26]	97.84%	2.87%	90.46%	94.62%
GWABC-SVM (Proposed)	99.5%	1.2%	97.42%	96.31%

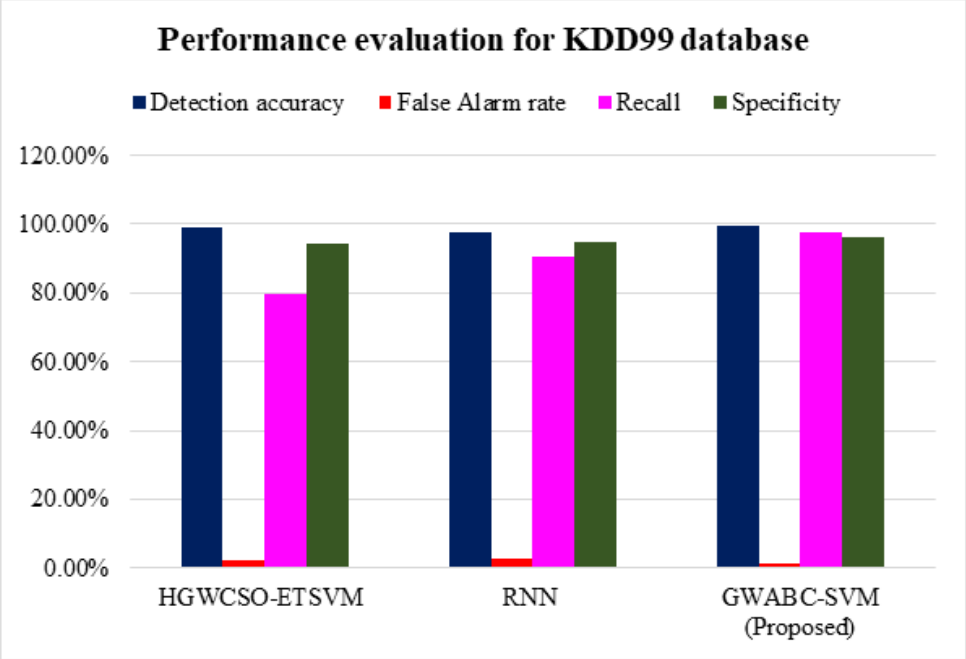

Figure 5: *Comparison of performance measures for existing and proposed method for KDD99 database*

The Table 5 depicts the performance evaluation for KDD99 dataset. It is obvious from the outcomes of performance assessment that the detection accuracy for Proposed (GWABC-SVM) is higher (99.5%)when compared to all other existing techniques likeHybrid Grey Wolf optimizer Cuckoo Search Optimization along with Enhanced Transductive Support VectorMachine(HGWCSO-ETSVM) and Recurrent Neural Network (RNN). It is also observed that the proposed method achieves very low FAR value of 1.2% whilst HGWCSO-ETSVM and RNN achieves 2.42% and 2.87% respectively. Moreover, the proposed model shows goodoutcomesin terms of recall and specificity while comparison with the existing techniques. The figure 5 shows the pictorialdescription of comparison of prevailing and proposed techniques for KDD99 database.
It is also observed that the UNBS-NB15 database provides better classification accuracy and FAR compared to the KDD99 database. From this explanation, it can be seen that the KDD99 database does not accurately represent the today's low tracing attack condition or the network traffic situation

of today. However, UNBS-NB15 database proven to be the best by taking all of these factors into account. In the future, a number of other databases may be considered in order to confirm the reliability of machine learning techniques for threat detection and prevention.

6. Conclusion

Regardless of the existence of several attack detection technologies and methodologies, the network intrusion continues to be unstoppable. Modern methods are being used by the intruders and attackers to dangerously affect the network's authorized systems. In this article, a novel framework termed the GWABC-SVM model is introduced to better identify intrusions and track the behaviours of attackers. Functionalities including pre-processing, feature selection, and detection are all part of the experiment. The min-max normalization approach is utilized during the pre-processing for improving the efficiency of attack detection. The GWABC method is employed in the feature selection step for choosing the best attributes from the dataset. By creating improved fitness values, the optimal attributes are upgraded.The intruder and normal traits are recognized more successfully when utilizing the SVM algorithm. The training and testing approach is applied to the selected attributes to generate more precise characteristics. Performance criteria including False Alarm Rate, recall, specificity, and accuracy are assessed. Higher performance metrics are provided by the suggested GWABC-SVM method. The conclusion is drawn from the experimental data that the suggested method performs better than current methods. Future versions of the approaches will be created to successfully find different more threats by combining sophisticated hybrid optimization with classification algorithms.

References

[1] H. Tabrizchi and M. Kuchaki Rafsanjani, “A survey on security challenges in cloud computing: issues, threats, and solutions,” J. Supercomput., vol. 76, no. 12, pp. 9493–9532, Dec. 2020, doi: 10.1007/s11227-020-03213-1.

[2] K. Doshi, Y. Yilmaz, and S. Uludag, “Timely Detection and Mitigation of Stealthy DDoS Attacks via IoT Networks,” IEEE Trans. Dependable Secure Comput., pp. 1–1, 2021, doi: 10.1109/TDSC.2021.3049942.

[3] A. R. Vishwakarma, “Network Traffic Based Botnet Detection Using Machine Learning,” Master of Science, San Jose State University, San Jose, CA, USA, 2020. doi: 10.31979/etd.4nd6-m6hp.

[4] T. Guarda, S. Bustos, W. Torres, and F. Villao, “Botnets the Cat-Mouse Hunting,” in Digital Science, vol. 850, T. Antipova and A. Rocha, Eds. Cham: Springer International Publishing, 2019, pp. 408–416. doi: 10.1007/978-3-030-02351-5_46.

[5] Y. Xu, G. Deng, T. Zhang, H. Qiu, and Y. Bao, “Novel denial-of-service attacks against cloud-based multi-robot systems,” Inf. Sci., vol. 576, pp. 329–344, Oct. 2021, doi: 10.1016/j.ins.2021.06.063.

[6] A. Mosteiro-Sanchez, M. Barcelo, J. Astorga, and A. Urbieta, “Securing IIoT using Defence-in-Depth: Towards an End-to-End secure Industry 4.0,” J. Manuf. Syst., vol. 57, pp. 367–378, Oct. 2020, doi: 10.1016/j.jmsy.2020.10.011.

[7] D. Gonzalez-Cuautle et al., “Synthetic Minority Oversampling Technique for Optimizing Classification Tasks in Botnet and Intrusion-Detection-System Datasets,” Appl. Sci., vol. 10, no. 3, p. 794, Jan. 2020, doi: 10.3390/app10030794.

[8] N. Chaabouni, M. Mosbah, A. Zemmari, C. Sauvignac, and P. Faruki, “Network Intrusion Detection for IoT Security Based on Learning Techniques,” IEEE Commun. Surv. Tutor., vol. 21, no. 3, pp. 2671–2701, 2019, doi: 10.1109/COMST.2019.2896380.

[9] P. Schneider and K. Böttinger, “High-Performance Unsupervised Anomaly Detection for Cyber-Physical System Networks,” in Proceedings of the 2018 Workshop on Cyber-Physical Systems Security and PrivaCy, New York, NY, USA, Jan. 2018, pp. 1–12. doi: 10.1145/3264888.3264890.

[10] A. Mishra and P. Yadav, "Anomaly-based IDS to Detect Attack Using Various Artificial Intelligence & Machine Learning Algorithms: A Review," in 2nd International Conference on Data, Engineering and Applications (IDEA), Bhopal, India, Feb. 2020, pp. 1–7. doi: 10.1109/IDEA49133.2020.9170674.

[11] F. S. de Lima Filho, F. A. F. Silveira, A. de Medeiros Brito Junior, G. Vargas-Solar, and L. F. Silveira, "Smart Detection: An Online Approach for DoS/DDoS Attack Detection Using Machine Learning," Secur. Commun. Netw., vol. 2019, pp. 1–15, Oct. 2019, doi: 10.1155/2019/1574749.

[12] S. Homayoun, M. Ahmadzadeh, S. Hashemi, A. Dehghantanha, and R. Khayami, "BoTShark: A Deep Learning Approach for Botnet Traffic Detection," in Cyber Threat Intelligence, vol. 70, A. Dehghantanha, M. Conti, and T. Dargahi, Eds. Cham: Springer International Publishing, 2018, pp. 137–153. doi: 10.1007/978-3-319-73951-9_7.

[13] M. Injadat, A. Moubayed, A. B. Nassif, and A. Shami, "Multi-Stage Optimized Machine Learning Framework for Network Intrusion Detection," IEEE Trans. Netw. Serv. Manag., vol. 18, no. 2, pp. 1803–1816, Jun. 2021, doi: 10.1109/TNSM.2020.3014929.

[14] A. A. Diro and N. Chilamkurti, "Distributed attack detection scheme using deep learning approach for Internet of Things," Future Gener. Comput. Syst., vol. 82, pp. 761–768, May 2018, doi: 10.1016/j.future.2017.08.043.

[15] S. Hosseini and B. M. H. Zade, "New Hybrid Method for Attack Detection Using Combination of Evolutionary Algorithms, SVM, and ANN," Comput. Netw., vol. 173, p. 40, 2020, doi: https://doi.org/10.1016/j.comnet.2020.107168.

[16] M. Aamir and S. M. A. Zaidi, "DDoS attack detection with feature engineering and machine learning: the framework and performance evaluation," Int. J. Inf. Secur., vol. 18, no. 6, pp. 761–785, 2019.

[17] N. Koroniotis, N. Moustafa, E. Sitnikova, and J. Slay, "Towards Developing Network Forensic Mechanism for Botnet Activities in the IoT Based on Machine Learning Techniques," in Mobile Networks and Management, vol. 235, J. Hu, I. Khalil, Z. Tari, and S. Wen, Eds. Cham: Springer International Publishing, 2018, pp. 30–44. doi: 10.1007/978-3-319-90775-8_3.

[18] N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in 2015 Military Communications and Information Systems Conference (MilCIS), Canberra, Australia, Nov. 2015, pp. 1–6. doi: 10.1109/MilCIS.2015.7348942.

[19] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, Ottawa, ON, Canada, Jul. 2009, pp. 1–6. doi: 10.1109/CISDA.2009.5356528.

[20] H. Rezaei, O. Bozorg-Haddad, and X. Chu, "Grey Wolf Optimization (GWO) Algorithm," in Advanced Optimization by Nature-Inspired Algorithms, vol. 720, O. Bozorg-Haddad, Ed. Singapore: Springer Singapore, 2018, pp. 81–91. doi: 10.1007/978-981-10-5221-7_9.

[21] M. Schiezaro and H. Pedrini, "Data feature selection based on Artificial Bee Colony algorithm," EURASIP J. Image Video Process., vol. 2013, no. 1, p. 47, Dec. 2013, doi: 10.1186/1687-5281-2013-47.

[22] K. S. Sahoo et al., "An Evolutionary SVM Model for DDOS Attack Detection in Software Defined Networks," IEEE Access, vol. 8, pp. 132502–132513, 2020, doi: 10.1109/ACCESS.2020.3009733.

[23] T. A. Tuan, H. V. Long, L. H. Son, R. Kumar, I. Priyadarshini, and N. T. K. Son, "Performance evaluation of Botnet DDoS attack detection using machine learning," Evol. Intell., vol. 13, no. 2, pp. 283–294, Jun. 2020, doi: 10.1007/s12065-019-00310-w.

[24] V. Kumar, D. Sinha, A. K. Das, S. C. Pandey, and R. T. Goswami, "An integrated rule based intrusion detection system: analysis on UNSW-NB15 data set and the real time online dataset," Clust. Comput., vol. 23, no. 2, pp. 1397–1418, Jun. 2020, doi: 10.1007/s10586-019-03008-x.

[25] E. M. Roopa Devi and R. C. Suganthe, “Enhanced transductive support vector machine classification with grey wolf optimizer cuckoo search optimization for intrusion detection system,” Concurr. Comput. Pract. Exp., vol. 32, no. 4, Feb. 2020, doi: 10.1002/cpe.4999.

[26] P. R. Kshirsagar, R. K. Yadav, N. N. Patil, and others, “Intrusion Detection System Attack Detection and Classification Model with Feed-Forward LSTM Gate in Conventional Dataset,” Mach. Learn. Appl. Eng. Educ. Manag., vol. 2, no. 1, pp. 20–29, 2022.