International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

Binary Arithmetic Optimization Algorithm with Machine Learning based Intrusion Detection System

[1,*]S. P. Senthilkumar, [2]Dr. Aranga Arivarasn
[1]Research Scholar, Department of Computer & Information Science,
Annamalai University, Annamalai Nagar - 608 002
E-Mail: senthil.sp74@gmail.com
[2]Assistant Professor/Programmer, Department of Computer & Information Science,
Annamalai University, Annamalai Nagar - 608 002
E-Mail: profarivarasan@yahoo.com

Abstract—Intrusion Detection Systems (IDS) are significant for preventing and identifying malicious actions in computer networks. Machine Learning (ML) approaches are extremely executed for recognizing intrusion since it is investigating huge volumes of network traffic data and recognize designs indicative of intrusions. But, the performance of these ML approaches is greatly dependent upon the choice of relevant features which efficiently represent the network traffic data. Feature Selection (FS) is the procedure of recognizing the most informative and discriminative aspects in a given database. As part of Intrusion Detection (ID) utilizing ML, FS purposes for identifying the subset of features that are efficiently differentiated between normal network behaviour and malicious activities. This article proposes a Binary Arithmetic Optimization Algorithm with Machine Learning based Intrusion Detection System (BAOA-MLIDS) technique. The BAOA-MLIDS technique employs FS with an optimal ML classifier for the ID process. To accomplish this, the BAOA-MLIDS technique performs data preprocessing to scale the input data. Besides, the BAOA-MLIDS technique comprises BAOA based FS approach to choose optimal features. Moreover, Extreme Learning Machine (ELM) approach is utilized for the identification of the intrusions. Furthermore, Hunger Games Search Optimization (HGSO) approach was employed for the hyperparameter optimization of the ELM approach. The performance assessment of the BAOA-MLIDS model was examined on a standard dataset and the outputs outperformed the advancement of the BAOA-MLIDS model in the ID process.

Keywords- Network security; Parameter tuning; Intrusion detection system; Feature selection; Machine learning.

I. INTRODUCTION

The network safety system has become a serious worldwide problem which can affect governments, enterprises, and individuals. The attacks rate against network systems has increased significantly and the attackers are continuing their strategies, which are used for development. ID is one solution to the problems against these outbreaks [1]. The IDS is efficient for identifying potential cyber-attacks. It applies the techniques for the classification and detection of the attacks [2]. There are two classes of IDS, such as (i) Anomaly (ii) Signature based IDS, represented as Anom-IDS and Sig-IDS. The sig-IDS method is detected outbreaks dependent upon formerly identified sequences, patterns, or a group of principles determined for the attack [3]. In the meantime, the Anom-IDS method can identify something changed than normal traffic, for instance, anomalies. The development of Anom-IDS over Sig-IDS; could be able to identify new attacks in the network system. Furthermore, due to the data source, Host and Network based IDS, represented as HIDS and NIDS, are two categories of IDS [4]. The algorithm of HIDS can identify the attacks across the system by analyzing the data from audits on apps or database logs, firewall logs, and the operating system [5]. The method of NIDS can identify outside attacks before it arrives in the computer networks. NIDS is monitoring the traffic data extracted from various network data sources in the network for detecting some threats. A general and effectual technique to design the IDS is ML.

IDS researchers have used different approaches for ID [6]. One of these methods is based on ML. This technique can detect and predict threats before they outcome in the main security cases [7]. Classification of instances into 2 categories is known as binary classification. On the other way, multiple-class classification is referred to categorizing samples into more than three categories. ANN is a self-adaptable computational and mathematical process which is made with a connected group of Artificial Neurons [8]. There are many kinds ANNs namely Auto-Encoder Neural Networks (AENNs), Deep Convolution Neural Networks (DCNNs), and Recurrent Neural Networks (RNNs), which occur with their degree of complexity and individual specific applications. In ML difficulties, the higher dimension features lead to extend classification process. While lower dimensional features can decrease the classification process [9]. Besides, the classification of network traffic data with unbalanced class allocations takes modelled as an important disadvantage on the performance achievable by most famous classifiers that accept comparatively balanced class allocations and equal mis-classification expenses [10]. The regular incidence and problems related to unbalanced class allocations specify the requirement for additional research works.

This article proposes a Binary Arithmetic Optimization Algorithm with Machine Learning based Intrusion Detection System (BAOA-MLIDS) technique. The BAOA-MLIDS technique employs FS with an optimal ML classifier for the ID process. To accomplish this, the BAOA-MLIDS technique performs data preprocessing to scale the input data. Besides, the BAOA-MLIDS technique comprises BAOA based FS approach to choose optimal features. Moreover, Extreme Learning Machine (ELM) model is utilized for the identification of the intrusions. Furthermore, Hunger Games Search Optimization (HGSO) approach was employed for the hyperparameter optimization of the ELM approach. The performance assessment of the BAOA-MLIDS methodology was examined on a benchmark dataset.

II. RELATED WORKS

Wang et al. [11] developed a novel Ensemble FS-based DNN (EFS-DNN) for attack detection in the network with large-volume traffic data. Especially, the LightGBM was leveraged as a base selector in the EFS model to improve the effectiveness of the optimum subset. In addition, a DNN with embedding and batch normalization algorithm is used as a classifier for improving expressiveness. Mhawi et al. [12] introduced a new Ensemble Learning (EL) model-based network IDS algorithm. The effective FS can be obtained by the hybrid of Correlation FS combined with Forest Panelized Attributes (CFS–FPA). The modified IDS includes AdaBoosting and bagging ensemble learning models for modifying the four dissimilar classifiers: KNN, RF, SVM, and NB. The authors [13] develop a robust DL technique namely AE-IDS based on the RF method. This technique creates the training subset with feature grouping and FS. Afterwards training, the model could forecast the performances with AE which efficiently enhances the prediction performance and considerably decreases the recognition time.

In [14], the authors introduced a Binary form of the Farmland Fertility Algorithm (BFFA) to FS during the IDS categorization. During this work, the V-shaped function can be applied for moving the FFA process under the binary space and for the unremitting location of performances from the FFA model to binary mode. Vijayanand and Devaraj [15] presented a wrapper-based technique using the WOA. One disadvantage of WOA is that early convergence leads to local optimum solutions. To resolve these limitations, we developed a technique where the GA operator was fused with the WOA. The presented technique chooses the relevant feature in the network dataset that assists in precisely detecting the intrusion. We recognized the type of intrusion based on the features selected using an SVM model.

Shakya [16] developed a fusion of ML and modified GWO (MLGWO) technique for enhanced IDS. The optimum amount of wolves is created by conducting examination with various wolves. In the WSN platform, the false alarm rate was decreased together with the decrease in processing time but increasing the accuracy of ID and the detection rates with the decline in the count of resulting features. Upadhyay et al. [17] designed a complex system for smart grids that incorporates feature engineering-based preprocessing with an ML classifier for IDS. The ML method finetunes the hyperparameter for improving the detection rate, the study mainly focuses on selecting the relevant features of datasets utilizing Gradient Boosting FS (GBFS) before using the classifier technique, a group which increases the execution speed and rate of detection.

III. THE PROPOSED MODEL

This article has presented a novel BAOA-MLIDS methodology to accomplish network security. The BAOA-MLIDS technique employs FS with an optimal ML classifier for the ID process. To achieve this, the BAOA-MLIDS technique comprises preprocessing, BAOA-based feature subset selection, ELM classification, and HGSO-based tuning procedure. Fig. 1 exemplifies the complete workflow of the BAOA-MLIDS technique.

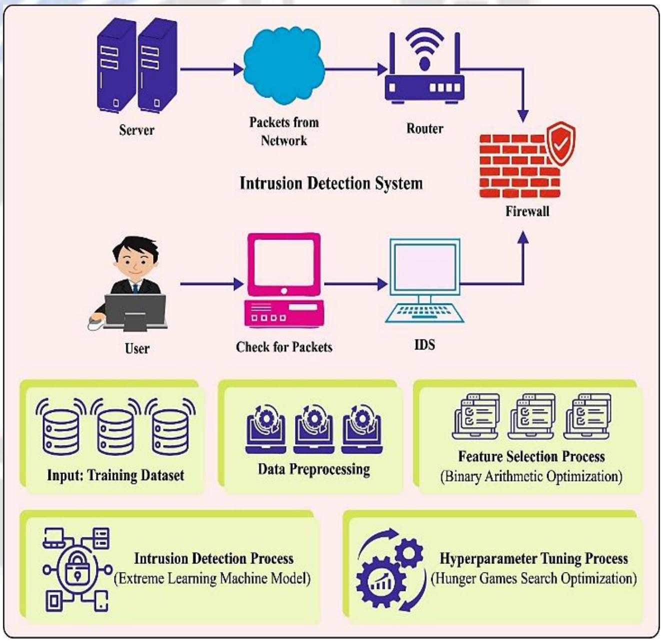
Server
Packets from Network
Router
Firewall
Intrusion Detection System
User
Check for Packets
IDS
Input: Training Dataset
Data Preprocessing
Feature Selection Process (Binary Arithmetic Optimization)
Intrusion Detection Process (Extreme Learning Machine Model)
Hyperparameter Tuning Process (Hunger Games Search Optimization)

Figure 1. Overall flow of BAOA-MLIDS methodology

A. *Data Preprocessing*

Diverse data normalization approaches were accessible such that z-score normalization, min-max normalization, and normalization by scaling decimal, and so on [18]. This effort utilizes a min-max normalized approach for normalizing all individual data features that cause biases in its values. This

Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

approach was employed for scaling normalized feature values among zero and one; for achieving standardization in the values of data features. The subsequent Eq. (1) was utilized for normalizing data feature values in any particular range.

$$\text{Min Max} = \frac{X_i - \min(A)}{\max(A) - \min(A)} \quad (1)$$

B. Feature Selection using BAOA

For selecting optimal features, the BAOA is used. As the arithmetic optimization algorithm only performs well in dealing with continuous optimization problems, BAOA is proposed to perform FS [19]. The steps of BAOA are given in the following.

Step1: Initialized

The generation of an arbitrary set of solutions or particles comprises the count of buses of load associated, as follows.

$$x_i = randi()$$
$$den_i = rand()$$
$$vol_i = rand()$$
$$acc_i = lb_i + rand() \times (ub_i - lb) \quad (2)$$

Where $randi()$ indicates the integer number produced as zero or one for all the i particles, den, and vol shows the density and volume, correspondingly; and $rand()$ denotes random number within [0,1]. acc shows the particle acceleration, whereas lb and ub signify the lower as well as upper bounds of searching space, correspondingly. The volume, acceleration, and density particles are produced from the vector dimensional of d and N implies the size of populations.

Step2: Upgrading volume and density

The volume and density of all the particles i for t iteration was upgraded as:

$$den_i^{t+1} = den_i^t + rand \times (den_{best} - den_i^t) \quad (3)$$
$$vol_i^{t+1} = vol_i^t + rand \times (vol_{best} - vol_i^t) \quad (4)$$

Where, acc_i, n_i, and vol_i denote the acceleration, density, and volume of the particle i during iteration t and vol_{best}, den_{best} indicate the volume and density of the better particle, correspondingly.

Step3: Upgrading acceleration

Based on two operators, the acceleration is updated as transfer function TF, and the density operator d, but the transfer function changeover the searching in exploration to exploitation, while the density operator helps the global to local searches.

$$TF = exp\left(\frac{t - T}{T}\right) \quad (5)$$
$$d = exp\left(\frac{t - T}{T}\right) - \left(\frac{t}{T}\right) \quad (6)$$

when $TF > 0.5$, it is upgraded dependent upon the exploitation phase, when $TF \leq 0.5$, the acceleration was upgraded dependent upon the exploration phase as follows.

$$acc_i^{t+1} = \frac{den_r + vol_r \times acc_r}{den_i^{t+1} \times vol_i^{t+1}}; TF \leq 0.5 \quad (7)$$

$$acc_i^{t+1} = \frac{den_{best} + vol_{best} \times acc_{best}}{den_i^{t+1} \times vol_i^{t+1}}; TP > 0.5 \quad (8)$$

$$acc_{i-norm}^{t+1} = u \times \frac{acc_i^{t+1} - \min(acc)}{\max(acc) - \min(acc)} + l \quad (9)$$

Where, acc_r, n_r, and vol_r represent the acceleration, density, and volume of chosen arbitrary particle correspondingly; acc_{best} signifies the optimum particle's acceleration; and u, l represents the normalization limit that is fixed to [0.9, 0.1], correspondingly.

Step4: Upgrading position

A novel position of particles from the population can be upgraded as written in (10) and (11).

$$x_i^{t+1} = \begin{cases} x_i^t + C1 \times rand \times \\ acc_{i-norm}^{t+1} \times (x_{rand} - x_i^t) \times d & if\ TF \leq 0.5 \\ x_{best}^t + f \times C2 \times rand \times \\ acc_{i-norm}^{t+1} \times (T \times x_{best} - x_i^t) \times d & if\ TF > 0.5 \end{cases} \quad (10)$$

$$f = \begin{cases} +1 \text{ if } P \leq 0.5 \\ -1 \text{ if } P > 0.5 \end{cases} P = 2 \times rand - C4 \quad (11)$$

In which, $C1$ and $C2$ represent the constants with values 2 and 6, correspondingly; $T = C3 \times TF$ and f denote the flag parameter determined in Eq. (11); and $C3$ and $C4$ indicate the constants with values [2, 0.5], correspondingly. But, to upgrade the position of particles from a separate searching space, a sigmoidal transfer function can be executed in BAOA, as illustrated in (12). Thus, the upgrade particle position from the BAOA is x', between the limit [0, 1], as depicted in Eq. (13).

$$sig(x_i^{t+1}) = \frac{1}{1 + e^{-(x_i^{t+1})}} \quad (12)$$

$$x_i'^{t+1} = \begin{cases} 0\ if\ rand \geq sig\ (x_i^{t+1}) \\ 1\ if\ rand < sig\ (x_i^{t+1}) \end{cases} \quad (13)$$

C. ID using ELM Model

In this work, the ELM approach can be executed for the identification and classification of intrusions. The ELM is a single hidden layer Supervised Learning approach exhibiting the Feedforward Neural Network (SLFNN) [20]. The ELM exceeds at modelling non-linear data performance in difficult methods. Fig. 2 depicts the framework of ELM.

(1) While the mapping functions of hidden states are recognized, if the better weighted can be selected next the resultant weight is defined analytically, the ELM approach showcases an important estimate accuracy that creates the ELM a fast learner. (2) It takes an easy execution, there is no requirement for artificially fixed a huge count of trained parameters before the trained model. (3) It takes an optimum generalized, thus the challenge of creating local optimal solutions could not simply be created.

For estimating the PV power plant production, it can be considered as: trained instance (x_i, y_i), input variable $x_i = [x_{i1}, x_{i2}, \ldots, x_{iN}]^T \in R^n$, and the predictable outcome $y_i =$

Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

$[y_{i1}, y_{i2}, \dots, y_{im}]^T \in R^m$. The mathematical formula of the ELM method is as:

$$Y_j = \sum_{i=1}^{n} \beta_i \, G(\omega_i X_j + v_i),' \tag{14}$$

In which, ω_i, β_i denotes the input and output layer biases, correspondingly; v_i signifies the hidden state bias, and $G(x)$ denotes the activation function.

$$H_{\omega,v,X}\beta = T \tag{15}$$

whereas T stands for the preferred resultant vector; and $H_{\omega,v,X}$ denotes the outcome of the implicit layer matrix that is written as:

$$H_{\omega,v,X} = \begin{bmatrix} G(\omega_1 X_1 + v_1) & \dots & G(\omega_i X_n + v_i) \\ \vdots & \ddots & \vdots \\ G(\omega_1 X_n + v_1) & \dots & G(\omega_i X_1 + v_i) \end{bmatrix} \tag{16}$$

Execute the resulting Eq. (17) to resolve for resultant weights.

$$\beta^* = H^+ T \tag{17}$$

In which, $\beta^* = H^+ T$ represents the generalization form of Moore's inverse that is executed to matrix H (Moore-Penrose).

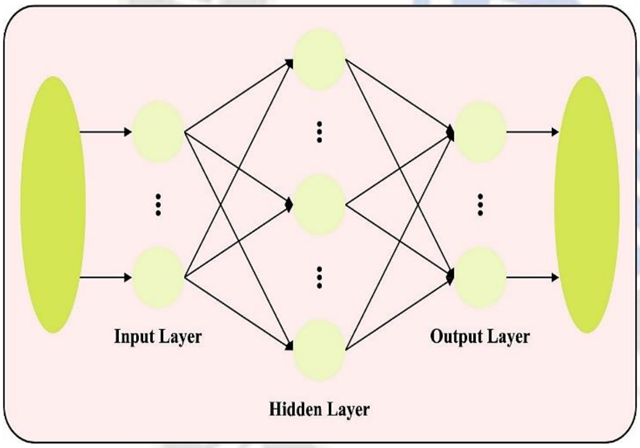

Figure 2. ELM structure

D. Parameter Tuning using HGSO Algorithm

Eventually, the HGSO algorithm was implemented for the parameter tuning of the ELM method. The HGSO inspires the hunger performance of chosen animals [21]. The process employed for tracking hunger performance as a critical homeostatic incentive defines HGSO fitness. Several behaviours that create performance and selection in the animals' lives are utilized for understanding and confining the optimizer procedure to novel uses. This method feature procedure is an adaptive weight depending on the hunger method utilized and creates replicates all the hunger effects searching step. The fundamental concept is that the presented system is more effective due to its higher solution, dynamic nature, and easy design concerning convergence and quality for suitable solutions.

The mathematical models expressed in the performance of the food method and the subsequent processes can be presented to inspire the contraction mode. The mathematical expression is represented by Eq. (18).

$$\overrightarrow{fobj\big(D,G,X(t+1)\big)} = \begin{cases} \overrightarrow{fobj\big(D,G,X(t)\big)} \cdot \big(1 + randn(1)\big), r_1 < l \\ \overrightarrow{W_1} \cdot \overrightarrow{X_b} + R \cdot \overrightarrow{W_r} \cdot \left| \overrightarrow{fobj(D,G,X_b)} - \overrightarrow{fobj\big(D,G,X(t)\big)} \right|, r_1 > l, r_r > E \\ \overrightarrow{W_1} \cdot \overrightarrow{fobj(D,G,X_b)} - R \cdot \overrightarrow{W_r} \cdot \left| \overrightarrow{fobj(D,G,X_b)} - \overrightarrow{fobj\big(D,G,X(t)\big)} \right|, r_1 > l, r_r < E \end{cases} \tag{18}$$

whereas R is within $[-a, a]$; r_1 and r_2 denote the arbitrary numbers between the limit [0, 1]; W_1 and W_2 imply the hunger weight; X_b refers to the arbitrary person from the population; and $X(t)$ denotes the individual.

$$E = sech(|F(i) - BF|) \tag{19}$$

whereas $i \in 1, r, \dots, n, F(i)$, and BF stands for the individual i, and optimum fitness value to present iteration. The hyperbolic function (Sech) is defined as Eq. (20):

$$\left(sech(x) = \frac{r}{e^x + e^{-x}} \right) \tag{20}$$

$$R = r \times a \times rand - a \tag{21}$$

Eqs. (22) and (23) demonstrate the mathematical equation of the role:

$$\overrightarrow{W_1(\iota)} = \begin{cases} hungry(i) \cdot \dfrac{N}{SHungry} \times r_4, r_3 < l \\ 1 r_3 > l \end{cases} \tag{22}$$

$$\overrightarrow{W_2(\iota)} = \big(1 - \exp(|hungry\ (i) - SHungry|)\big) \times r_5 \times 2 (23)$$

whereas the individual population is referred to by N, in which $SHungry$ demonstrates the sum(hungry). The randomized search was introduced utilizing arbitrary variables r_3, r_4 and r_5. Eq. (24) is an expressed for hungry(i):

$$hungry\ (i) = \begin{cases} \cdot & AllFitness\ (i) == BF \\ hungry(i) + H, & AllFitness\ (i)! = BF \end{cases} \tag{24}$$

in which, $AllFitness(i)$ preserves all the individual's fitness from the present iteration.

$$TH = \frac{F(i) - BF}{WF - BF} \times r_6 \times 2 \times (UB - LB) \tag{25}$$

$$H = \begin{cases} LH \times (1 + r), & TH < LH \\ TH, & TH \geq LH \end{cases} \tag{26}$$

whereas r_6 is another randomized variable and $F(i)$ refers the all the individual's fitness values. The worse and best fitness is represented by WF and BF, where the lower and upper searching bounds are LB and UB, correspondingly. While the hunger sensation H is a lower bound, LH, it gives the technique an optimal solution.

The fitness choice is a key feature of the HGSO method. The encoding performance was employed to develop a better solution for candidate outcomes. Presently, the value of accuracy is the major condition engaged for building a FF.

$$Fitness = \max(P) \tag{27}$$

$$P = \frac{TP}{TP + FP} \tag{28}$$

In which, FP and TP imply the false and true positive values.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

IV. EXPERIMENTAL VALIDATION

The ID outputs of the BAOA-MLIDS methodology are inspected on the IDS2017 dataset [22], comprising 8000 samples as displayed in Table 1. Among the available 77 features, the BAOA has chosen 46 features.

TABLE I. DATASET DESCRIPTION

Class	Sample Numbers
Benign	1000
DoS	1000
Port Scan	1000
DDoS	1000
FTP Patator	1000
SSH Patator	1000
BotNet	1000
Web Attack	1000
Total Samples	8000

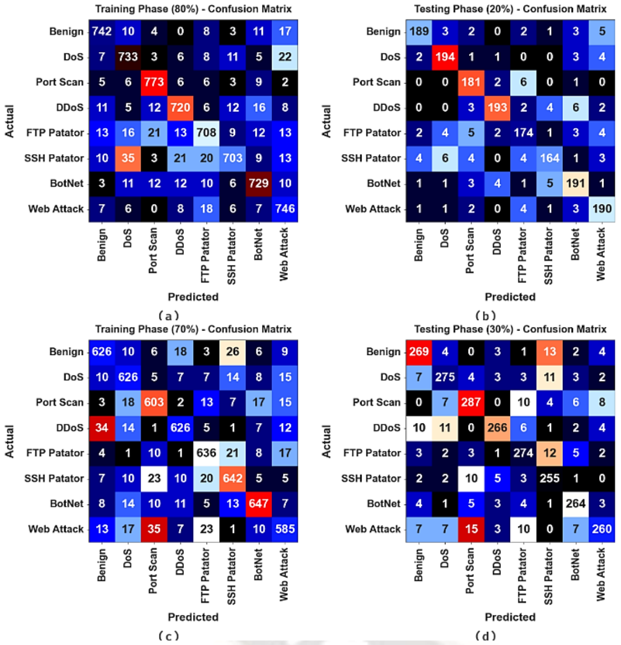

Figure 3. Confusion matrix of (a-b) 80:20 and (c-d) 70:30 of TR/TS set

Fig. 3, the confusion matrix of the BAOA-MLIDS methodology on ID are given. The outputs demonstrate that the BAOA-MLIDS methodology attains capable ID and classification process.

In Table 2 and Fig. 4, the overall ID result of the BAOA-MLIDS approach is inspected at 80:20 of the TR/TS set. The outputs indicate that the BAOA-MLIDS approach accomplishes effective outcomes under all measures. On 80% of the TR set, the BAOA-MLIDS model obtains an average $accu_y$ of 97.87%, $prec_n$ of 91.49%, $sens_y$ of 91.48%, $spec_y$ of 98.78%, and F_{score} of 91.46%. Also, on 20% of the TS set, the BAOA-MLIDS model gets an average $accu_y$ of 98.06%, $prec_n$ of 92.27%, $sens_y$ of 92.22%, $spec_y$ of 98.89%, and F_{score} of 92.21%.

TABLE II. ID OUTCOME OF BAOA-MLIDS METHOD ON 80:20 OF TR/TS SET

Class	$Accu_y$	$Prec_n$	$Sens_y$	$Spec_y$	F_{Score}
Training (80%)					
Benign	98.30	92.98	93.33	99.00	93.16
DoS	97.64	89.17	92.20	98.41	90.66
Port Scan	98.56	93.36	95.43	99.02	94.38
DDoS	97.88	91.60	91.14	98.82	91.37
FTP Patator	97.30	90.31	87.95	98.64	89.11
SSH Patator	97.48	93.36	86.36	99.10	89.73
BotNet	97.92	91.35	91.93	98.77	91.64
Web Attack	97.86	89.77	93.48	98.48	91.59
Average	**97.87**	**91.49**	**91.48**	**98.78**	**91.46**
Testing (20%)					
Benign	98.38	94.97	92.20	99.28	93.56
DoS	98.38	92.82	94.63	98.92	93.72
Port Scan	98.19	90.05	95.26	98.58	92.58
DDoS	98.38	95.54	91.90	99.35	93.69
FTP Patator	97.50	90.16	89.23	98.65	89.69
SSH Patator	97.88	93.18	88.17	99.15	90.61
BotNet	97.75	90.52	92.27	98.56	91.39
Web Attack	98.06	90.91	94.06	98.64	92.46
Average	**98.06**	**92.27**	**92.22**	**98.89**	**92.21**

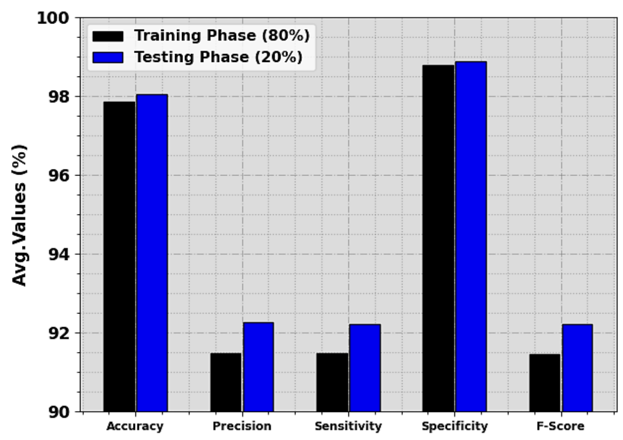

Figure 4. Average output of BAOA-MLIDS method on 80:20 of TR/TS set

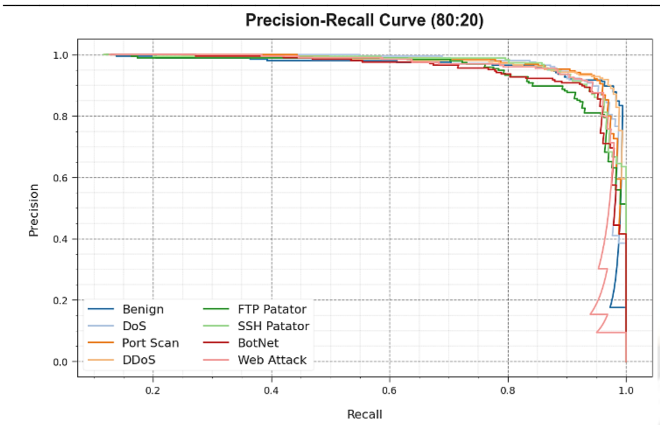

Figure 5. PR curve of BAOA-MLIDS method on 80:20 of TR/TS set

A detailed PR analysis of the BAOA-MLIDS model is shown at 80:20 of the TR/TS set in Fig. 5. The outcome stated that the BAOA-MLIDS model outcomes in increased values of PR. Moreover, the BAOA-MLIDS approach can achieve greater values of PR values on overall class labels.

In Fig. 6, a ROC analysis of the BAOA-MLIDS system is demonstrated on 80:20 of the TR/TS set. The outcome stated that the BAOA-MLIDS system outcomes in enhanced values of ROC. Also, the BAOA-MLIDS approach can achieve greater values of PR values on overall class labels.

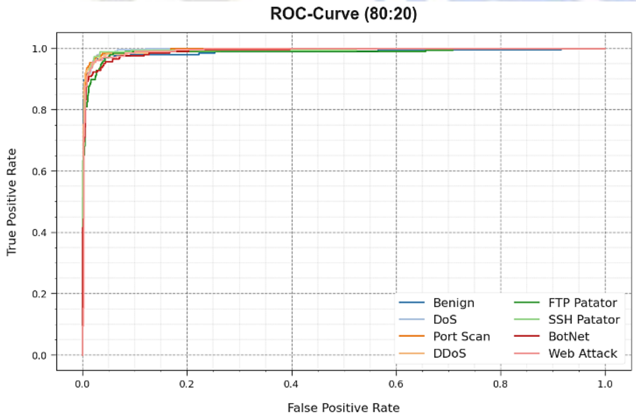

Figure 6. ROC curve of BAOA-MLIDS method on 80:20 of TR/TS set

In Table 3 and Fig. 7, the comprehensive ID analysis of the BAOA-MLIDS method is studied on 70:30 of the TR/TS set.

TABLE III. ID OUTCOME OF BAOA-MLIDS APPROACH ON 70:30 OF TR/TS SET

Class	$Accu_y$	$Prec_n$	$Sens_y$	$Spec_y$	F_{Score}
Training (70%)					
Benign	97.20	88.79	88.92	98.39	88.86
DoS	97.32	88.17	90.46	98.29	89.30
Port Scan	97.05	87.01	88.94	98.17	87.96
DDoS	97.68	91.79	89.43	98.86	90.59
FTP Patator	97.54	89.33	91.12	98.45	90.21
SSH Patator	97.09	88.55	88.92	98.30	88.74
BotNet	97.70	91.38	90.49	98.75	90.93
Web Attack	96.68	87.97	84.66	98.37	86.28
Average	**97.28**	**89.12**	**89.12**	**98.45**	**89.11**
Testing (30%)					
Benign	97.50	89.07	90.88	98.43	89.97
DoS	97.21	89.00	89.29	98.37	89.14
Port Scan	97.00	88.58	89.13	98.22	88.85
DDoS	97.83	93.66	88.67	99.14	91.10
FTP Patator	97.29	88.10	90.73	98.24	89.40
SSH Patator	97.29	85.86	91.73	98.02	88.70
BotNet	98.04	91.03	92.63	98.77	91.83
Web Attack	97.00	91.87	84.14	98.90	87.84
Average	**97.40**	**89.65**	**89.65**	**98.51**	**89.60**

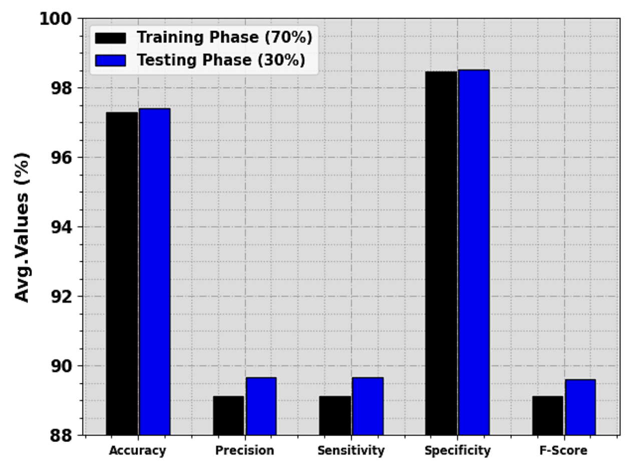

Figure 7. Average output of BAOA-MLIDS methodology on 70:30 of TR/TS set

The outcomes indicate that the BAOA-MLIDS method achieves effective results under all measures. On 70% of the TR set, the BAOA-MLIDS model gets average $accu_y$ of 97.28%, $prec_n$ of 89.12%, $sens_y$ of 89.12%, $spec_y$ of 98.45%, and F_{score} of 89.11%. Also, on 30% of the TS set, the BAOA-MLIDS model attains an average $accu_y$ of 97.40%, $prec_n$ of 89.65%, $sens_y$ of 89.65%, $spec_y$ of 98.51%, an F_{score} of 89.60%.

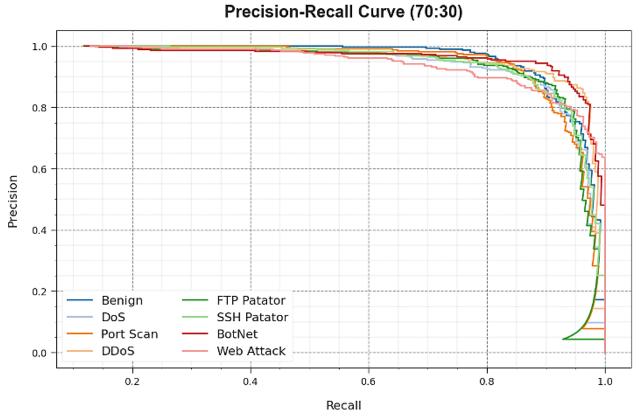

Figure 8. PR curve of BAOA-MLIDS methodology on 70:30 of TR/TS set

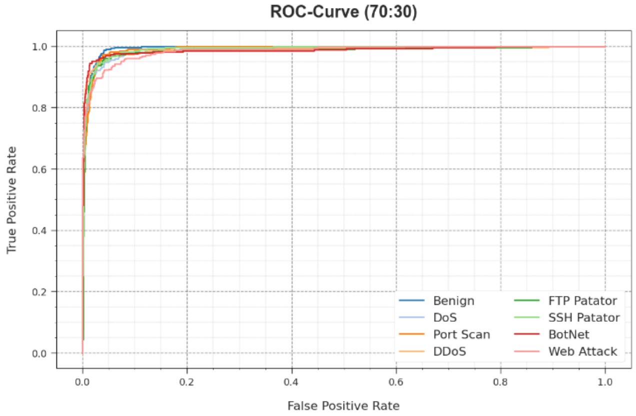

Figure 9. ROC curve of BAOA-MLIDS methodology on 70:30 of TR/TS set

A comprehensive PR evaluation of the BAOA-MLIDS method is revealed at 70:30 of the TR/TS set in Fig. 8. The output stated that the BAOA-MLIDS system outcomes in raising values of PR. Furthermore, the BAOA-MLIDS method can attain greater PR values on overall classes.

In Fig. 9, a ROC evaluation of the BAOA-MLIDS algorithm is demonstrated at 70:30 of the TR/TS set. The result inferred that the BAOA-MLIDS algorithm results in enhanced ROC values. Additionally, the BAOA-MLIDS system can extend greater ROC values on overall classes.

In Table 4 and Fig. 10, the BAOA-MLIDS approach is related with other ML models is made [23, 24].

TABLE IV. COMPARATIVE OUTCOME OF BAOA-MLIDS ALGORITHM WITH OTHER ML TECHNIQUES

Algorithms	$Accu_y$	$Prec_n$	$Reca_l$	F_{Score}
BAOA-MLIDS	98.06	92.27	92.22	92.21
Random Forest	93.6	62.04	76.97	66.19
SVM Model	92.1	59.14	67.25	65.89
Logistic Regression	91.8	59.91	67.61	65.76
Naive Bayes	91.5	72.35	60.4	59.74
MLP Algorithm	76.8	70.31	64.72	74.52

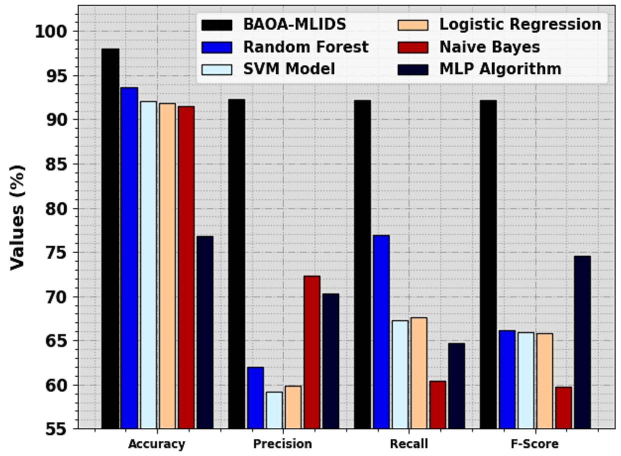

Figure 10. Comparative output of BAOA-MLIDS methodology with other ML techniques

The experimental outcomes inferred that the BAOA-MLIDS approach is greater than other models in terms of several measures. In addition, it is noticed that the BAOA-MLIDS technique reaches effectual performance with a maximum $accu_y$ of 98.06%, $prec_n$ of 92.27%, $reca_l$ of 92.22%, and F_{score} of 92.21%. These outcomes illustrated the enhanced performance of the BAOA-MLIDS methodology.

V. CONCLUSION

This study has presented a novel BAOA-MLIDS technique to accomplish network security. The BAOA-MLIDS technique employs FS with an optimal ML classifier for the ID process. To accomplish this, the BAOA-MLIDS technique performs data preprocessing to scale the input data. Besides, the BAOA-MLIDS technique comprises BAOA based FS approach to choose optimal features. Moreover, ELM model is utilized for the identification of the intrusions. Furthermore, Hunger Games Search Optimization (HGSO) approach was employed for the parameter tuning of the ELM approach. The simulation outcome of the BAOA-MLIDS model was inspected on a benchmark database and the outcome demonstrates the betterment of the BAOA-MLIDS model in the ID process.

REFERENCES

[1] Ngo, V.D., Vuong, T.C., Van Luong, T. and Tran, H., 2023. Machine learning-based intrusion detection: feature selection versus feature extraction. Cluster Computing, pp.1-15.

[2] Kasongo, S.M. and Sun, Y., 2020. Performance analysis of intrusion detection systems using a feature selection method on the UNSW-NB15 dataset. Journal of Big Data, 7, pp.1-20.

[3] Ayo, F.E., Folorunso, S.O., Abayomi-Alli, A.A., Adekunle, A.O. and Awotunde, J.B., 2020. Network intrusion detection based on deep learning model optimized with rule-based hybrid feature selection. Information Security Journal: A Global Perspective, 29(6), pp.267-283.

[4] Chkirbene, Z., Erbad, A., Hamila, R., Mohamed, A., Guizani, M. and Hamdi, M., 2020. TIDCS: A dynamic intrusion detection

and classification system based feature selection. IEEE Access, 8, pp.95864-95877.

[5] Fitni, Q.R.S. and Ramli, K., 2020, July. Implementation of ensemble learning and feature selection for performance improvements in anomaly-based intrusion detection systems. In 2020 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT) (pp. 118-124). IEEE.

[6] Illavarason, P. and Sundaram, B.K., 2019, December. A Study of Intrusion Detection System using Machine Learning Classification Algorithm based on different feature selection approach. In 2019 Third international conference on I-SMAC (IoT in social, mobile, analytics and cloud)(I-SMAC) (pp. 295-299). IEEE.

[7] Kalimuthan, C. and Renjit, J.A., 2020. Review on intrusion detection using feature selection with machine learning techniques. Materials Today: Proceedings, 33, pp.3794-3802.

[8] Disha, R.A. and Waheed, S., 2022. Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique. Cybersecurity, 5(1), p.1.

[9] Salih, A.A. and Abdulrazaq, M.B., 2019, April. Combining best features selection using three classifiers in intrusion detection system. In 2019 International Conference on Advanced Science and Engineering (ICOASE) (pp. 94-99). IEEE.

[10] Otoum, Y., Liu, D. and Nayak, A., 2022. DL-IDS: a deep learning–based intrusion detection framework for securing IoT. Transactions on Emerging Telecommunications Technologies, 33(3), p.e3803.

[11] Wang, Z., Liu, J. and Sun, L., 2022. EFS-DNN: an ensemble feature selection-based deep learning approach to network intrusion detection system. Security and Communication Networks, 2022.

[12] Mhawi, D.N., Aldallal, A. and Hassan, S., 2022. Advanced feature-selection-based hybrid ensemble learning algorithms for network intrusion detection systems. Symmetry, 14(7), p.1461.

[13] Li, X., Chen, W., Zhang, Q. and Wu, L., 2020. Building auto-encoder intrusion detection system based on random forest feature selection. Computers & Security, 95, p.101851.

[14] Naseri, T.S. and Gharehchopogh, F.S., 2022. A feature selection based on the farmland fertility algorithm for improved intrusion detection systems. Journal of Network and Systems Management, 30(3), p.40.

[15] Vijayanand, R. and Devaraj, D., 2020. A novel feature selection method using whale optimization algorithm and genetic operators for intrusion detection system in wireless mesh network. IEEE Access, 8, pp.56847-56854.

[16] Shakya, S., 2021. Modified gray wolf feature selection and machine learning classification for wireless sensor network intrusion detection. IRO Journal on Sustainable Wireless Systems, 3(2), pp.118-127.

[17] Upadhyay, D., Manero, J., Zaman, M. and Sampalli, S., 2020. Gradient boosting feature selection with machine learning classifiers for intrusion detection on power grids. IEEE Transactions on Network and Service Management, 18(1), pp.1104-1116.

[18] Iqbal, N., Jamil, F., Ahmad, S. and Kim, D., 2021. A novel blockchain-based integrity and reliable veterinary clinic information management system using predictive analytics for provisioning of quality health services. IEEE Access, 9, pp.8069-8098.

[19] Rosli, H.M., Mokhlis, H., Mansor, N.N., Sapari, N.M., Halim, S.A., Wang, L. and Sulaima, M.F., 2023. A Binary Archimedes Optimization Algorithm and Weighted Sum Method for UFLS in Islanded Distribution Systems Considering the Stability Index and Load Priority. Energies, 16(13), pp.1-21.

[20] Liu, L., Guo, K., Chen, J., Guo, L., Ke, C., Liang, J. and He, D., 2023. A Photovoltaic Power Prediction Approach Based on Data Decomposition and Stacked Deep Learning Model. Electronics, 12(13), p.2764.

[21] Houssein, E.H., Hosney, M.E., Mohamed, W.M., Ali, A.A. and Younis, E.M., 2023. Fuzzy-based hunger games search algorithm for global optimization and feature selection using medical data. Neural Computing and Applications, 35(7), pp.5251-5275.

[22] https://www.unb.ca/cic/datasets/ids-2017.html

[23] Gautam, S.; Henry, A.; Zuhair, M.; Rashid, M.; Javed, A.R.; Maddikunta, P.K.R. A Composite Approach of Intrusion Detection Systems: Hybrid RNN and Correlation-Based Feature Optimization. Electronics 2022, 11, 3529. https://doi.org/10.3390/ electronics11213529.

[24] Abdulhammed, R., Musafer, H., Alessa, A., Faezipour, M. and Abuzneid, A., 2019. Features dimensionality reduction approaches for machine learning based network intrusion detection. Electronics, 8(3), p.322.