Received: November 10, 2024. Revised: December 12, 2024.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Di-Strategy Based Beluga Whale Optimization Algorithm for Feature Selection in Cloud Network Traffic Prediction

Mala Kariyappa[1]* **Hulikal Somashekharaiah Annapurna[2]**

[1]*Department of Information Science and Engineering,*
Channabasaveshwara Institution of Technology, Tumkur, India
[2]*Department of Information Science and Engineering, Sri Siddhartha Institute of Technology, Tumkur, India*
* Corresponding author's Email: mala.k@cittumkur.org

Abstract: Through the rapid advancement of network technology, Network Intrusion Detection Systems (IDS) have become a vital constituent of network security. However, many existing traditional models struggle to identify key features due to the complexity of the feature dimensions, which diminishes classification accuracy. Hence, this research proposes the Di-strategy-based Beluga Whale Optimization (DS-BWO) approach for performing the feature selection in intrusion detections. The DS-BWO approach excels in selecting the most relevant features from high-dimensional datasets, effectively reducing dimensionality and improving model performance while maintaining important information. A Hybrid Recurrent Neural Network with Bidirectional Long Short-Term Memory (RNN-BiLSTM) approach is used for the classification of intrusions into binary classes like normal and malignant. By processing data in both forward and backward directions, BiLSTM captures dependencies in sequences more effectively. The proposed DS-BWO approach is implemented using three standard datasets: CIC-IDS2018, NSL-KDD, and UNSW-NB15. The implementation results demonstrate that the proposed DS-BWO approach reaches a superior accuracy of 0.999 on the CIC-IDS2018 dataset in comparison to the existing algorithms like Multi-Agent Feature Selection IDS namely (MAFSIDS) and Extreme Gradient Boosting.

Keywords: Beluga optimization algorithm, Bidirectional long short-term memory, Di-strategy, Intrusion detection system, Recurrent neural network.

1. Introduction

The cloud network is a major part of cloud computing which provides the infrastructure for the communication and transfer of data between various devices inside the cloud environment. Through the fast growth of Internet of Things ((IoT) advancements, an increasing number of devices have become more intelligent and are gradually being integrated into various applications [1]. The conventional approaches manage the intrusions through access control, firewalls, encryption and attack-resistant software. However, these techniques are appropriate only for small-scale attack determination and are impracticable with a large number of advanced intrusions [2]. Due to the enhanced security occurrences in IoT devices, various researchers put forward to develop effective and efficient Intrusion Detection Systems (IDS) [3]. IDS is a software and hardware system that supports to identification of the intrusive behaviour of the network traffic. Mostly, IDS plays an important role in defence against intrusions into any network infrastructure [4, 5]. The IDS verifies the network traffic to detect any suspicious attacks on the given network, preventing cyber security threats. The IDS is categorized into two types: one is an anomaly detection system which is commonly known as profile-based detection and the other one is a misuse detection system which is commonly known as signature-based detection [6, 7]. Misuse detection performs rule-based or feature-matching techniques to determine if data represents an intrusion, whereas anomaly detection identifies whether the data is abnormal. Nevertheless, both approaches have

challenging such as low detection rates and high false alarm rates [8].

The IDS has become more challenging in recent times due to the emergence of new network attacks and the continuous growth in network traffic [9]. Researchers are using Machine Learning (ML) and Deep Learning (DL) approaches to extend IDS results with intelligent systems [10]. The ML and DL approaches have better capability to learn and recognise patterns from challenging information by statistical approaches and advanced algorithms [11, 12]. The DL algorithm is used for the classification of network traffic data. The information related to the network traffic is gathered and utilized to train an approach to learning the features of network traffic [13]. The various IDS datasets are used to enhance the classification of network traffic. The data is pre-processed to solve imbalance class issues and then used to select the significant features from actual data. The network traffic is classified based on the selected features that enhance the performance of the model [14, 15]. However, various existing traditional models are challenging to identify the significant features due to the complex dimensionality of the features that reduced the accuracy of classification. Hence, this research proposes the Di-strategy-based Beluga Whale Optimization (DS-BWO) approach for feature selection in intrusion detection.

The primary highlights of this manuscript are provided in the following:

- The preprocessing techniques like label encoder and robust scaler are used to enhance the input feature data. This leads to more efficient and accurate data representation, which enhances the model's capability to identify subtle and large intrusion patterns.
- The DS-BWO approach is proposed for the feature selection to reduce the dimensionality reduction problem. By concentrating on the most relevant features, the model can make more accurate predictions, minimizing the chance of false positives and false negatives.
- The hybrid Recurrent Neural Network with Bidirectional Long Short-Term Memory (RNN-BiLSTM) approach is used for the classification of intrusions. By capturing both past and future dependencies in the data, the RNN-BiLSTM model can more accurately classify normal and malicious behaviour.

This research paper is organized as follows: Section 2 demonstrates the literature survey. Section 3 illustrates the proposed methodology. Section 4 gives the results and discussion, and Section 5 provides the conclusion.

2. Literature survey

Nowadays, various studies have been introduced to acquire unique patterns from attack intrusions and classify the attacks from normal traffic. Various authors introduced different feature selection approaches to reduce the high-dimensionality problem and to select the important features. These are discussed in this section, along with the advantages and limitations.

Yu [16] introduced the Hybrid network classification approach of enhanced residual network blocks and Bidirectional Gated Recurrent Unit (BiGRU). The greedy strategy was utilized to enhance a self-encoder and solve an issue of a number of network layers. An idea of the double pooled layer was introduced and it was initially applied to the BiGRU approach and then applied to the residual network blocks to enhance the capability of this approach for the extraction of the time-series features. However, the greedy strategy mitigated the number of layers, it does not fully optimize feature selection for various time-series data.

Vashishtha [17] presented a hybrid model for the detection of intrusions. Gradient Boosting (GB), RF, and Neural Network (NN) models were ensembled for the intrusion classification. The Information Gain Evaluation Algorithm (IGEA) technique was used to select the significant features of the intrusion data that improved the performance of the model's effectiveness. However, the ML approaches were trained through the majority class which neglected the minority class, resulting in misclassification of data and reduced the accuracy of classification of the intrusions.

Ren [18] introduced the Multi-Agent Feature Selection IDS namely MAFSIDS with Deep Q Learning (DQL) which comprised of MAFS feature selection approach. The MAFIDS involved two significant characteristics such as the feature selection approach and the DRL attack detection module. A feature selection approach solved a high-dimensionality problem through an enhancement of the multi-agent RL approach. The DRL approach utilized the easy policy network, which enabled the network to provide quick responses. However, the DRL approach with an easy policy network provided fast responses and resulted in suboptimal feature selection and limited generalization when applied to challenging network environments.

Songma [19] aimed to determine the intrusions through the most significant classifier with effective preprocessing and feature selection approaches. The supervised ML approaches were utilized for the classification of IDS into binary classes. The

preprocessing techniques like data cleaning, exploratory data analysis and data normalizations (min-max and Z-score) techniques were utilized. The Principal Component Analysis (PCA) and RF were utilized to eliminate irrelevant features. However, feature selection with ML approaches involved potential loss of critical information during feature elimination with PCA and RF, which resulted in minimized classification accuracy.

Ullah [20] presented a transformer-based transfer learning to classify the intrusions in network traffic. The LSTM model was used for the classification of intrusion. The Synthetic Minority Oversampling Technique (SMOTE) was used to balance abnormal traffic data. The convolutional Neural Network (CNN) approach was utilized to obtain significant details of an intrusion that enhanced the accuracy of the classification of normal and abnormal network traffic. However, the performance of the model was reduced due to the rise of noises during the training process that affected the classification of intrusion.

H. Kanakadurga Bella1and S. Vasundra [21] presented the Extra Tree Regression Classifier and Grid Search Optimized Long ShortTerm Memory (ETR-GSO-LSTM) for the identification and classification of the intrusions in IoT and cloud environmentsHowever, data augmentation technique was potentially introduced the bias or noise, which poorly caused the model's capability to generalize to unknown data.

Ruqaya Abdulhasan Abed [22] presented the different feature selection approaches named Principal Component Analysis (PCA) and Singular Value Decomposition (SVD) to minimize feature space's dimensionality and to extract the most significant features. Furthermore, Ridge Regression, Stochastic Gradient Descent (SGD), and CNN were the approaches utilized for the classification in IDS. However, PCA and SVD were sensitive to noise and outliers in the data, potentially resulted in suboptimal feature extraction process.

From this overall analysis, some of the limitations have been identified: lack of performing the feature selection, minimized classification accuracy, limited generalization capability and misclassification. Hence, to overcome these limitations, this research proposes the DS-BWO method for the feature selection process and the RNN-BiLSTM approach for the classification of intrusions.

3. Proposed methodology

This paper proposes the DS-BWO method for the feature selection of the network intrusion and the RNN-BiLSTM approach is used for the classification of the intrusion. This research is comprised of different important steps like data collection, preprocessing, feature selection and finally classification. Fig. 1 demonstrates the framework of the proposed DS-BWO approach for IDS.

3.1 Dataset

Data collection is the initial stage of this research to assess a significance of the proposed DS-BWO approach. Three standard datasets such as CIC-IDS2018 [23], NSL-KDD [24] and UNSW-NB15 [25] are considered in this research.:

The CIC-IDS2018 dataset is an advanced IDS dataset that involves various attack states like Botnet, Web assaults, Brute-force, DoS, Heartbleed, DDoS and network dissemination. This dataset involves various machines: 50 attackers, 420 victims and 30 servers over different regions. It involves the network traffic, and the system logs from every machine and 80 features obtained from network traffic acquired through CICFlowMeter-V3.

The NSL-KDD training set involves 126,620 network traffic data, with separately involving 41 features. This dataset involves various feature categories like important features, event traffic, content, host-based traffic and some other additional features.

Actual network packets of the UNSW-NB-15 dataset are designed through an ACCS Network Scope Lab utilising the IXI PrimeSturf tool and involve both real-time network action and the synthetic modern attacks. The dataset involves various types of attacks such as backdoors, generic attacks, penetration analysis, worms Denial of Service (DoS) attacks, obfuscation testing, stepping stones, shellcode, and exploits.

3.2 Preprocessing

The collected datasets are utilized as input to the pre-processing step.

Figure. 1 Workflow of the proposed DS-BWO approach for IDS

The data preprocessing step becomes a significant part of the proposed method as it considers performing a network assault dataset. In this research, the two important pre-processing techniques Label encoder and Robust scaler are used to enhance the input feature. The data encoding is required to convert categorical data into numerical data. This research utilizes the label encoder in this preprocessing stage, as it does not alter the dimensions of the data [26]. The robust scalar approach is considered which has the stronger parameter control for the data centralization and the data scaling robustness [27].

3.3 Feature selection

The pre-processed data are provided to the feature selection step and it is done by proposing a novel DS-BWO approach. BWO approach has demonstrated excellent handling of high-dimensional feature spaces, which is crucial for IDS tasks where feature sets are often large and complex. The algorithm is efficient in dimension reduction and selecting the most relevant features as compared to the other approaches like Carpet Weaver Optimization (CWO) [28], Sculptor Optimization Algorithm (SOA) [29], Apiary Organizational-Based Optimization Algorithm (AOBA) [30], Swarm Bipolar Algorithm (SBA) [31], and Swarm Space Hopping Algorithm (SSHA) [32]. The bubble net feeding technique used by beluga whales is mimicked in BWO to enhance search efficiency. Whales create a bubble net to corral fish, and BWO mimics this by refining candidate solutions in a controlled manner, allowing the algorithm to focus more effectively on promising regions of the solution space. BWO is a swarm-based approach for solving the optimization issues which is motivated by beluga's whale behavior like swimming, prey hunting and falls. To improve a convergence capability, an exploitation stage utilizes the Levy flight function. The beluga whales measured the search agents due to a population-based course of BWO and every beluga whale is the candidate solution which is updated at an optimization. A position of a matrix of the search agent is formulated in Eq. (1) as follows:

$$X = \begin{bmatrix} X_{11} & X_{11} & \cdots & X_{1D} \\ X_{21} & X_{22} & \cdots & X_{2D} \\ X_{31} & X_{32} & \cdots & X_{3D} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N1} & X_{N2} & \cdots & X_{ND} \end{bmatrix} \quad (1)$$

Where, N demonstrates a beluga whale population size; D denotes a dimension. A balancing factor B_f which is formulated in Eq. (2). This process identifies whether a BWO approach changes from exploration to exploitation.

$$B_f = B_f \left(1 - \frac{T}{2} * T_{max}\right) \quad (2)$$

Where, T denotes a current iteration and T_{max} depicts the maximum number of iterations.

3.3.1. Exploitation phase

The swimming behaviour of the beluga whale is considered when establishing an exploring stage. The Beluga whales are fetching in the social-sexual behaviors in different postures, as evidenced through the activities preserved in beluga whales reserved in huma care like a pair of nearer beluga whales swimming in a synchronized way. As an outcome, the positions of beluga whales are rationalized in Eq. (3) as follows:

$$\begin{pmatrix} X_{ij}^{T+1} = X_{i,p_j}^{T} + \begin{pmatrix} X_{r,p_1}^{T} \\ -X_{i,p_j}^{T} \end{pmatrix} \begin{pmatrix} 1+ \\ r_1 \end{pmatrix} * sin(2\pi r_2), j = even \\ X_{ij}^{T+1} = X_{i,p_j}^{T} + \begin{pmatrix} X_{r,p_1}^{T} \\ -X_{i,p_j}^{T} \end{pmatrix} \begin{pmatrix} 1+ \\ r_1 \end{pmatrix} * cos(2\pi r_2), j = odd \end{pmatrix} \quad (3)$$

Where, T demonstrates the current iteration; X_{ij}^{T+1} illustrates the next iteration; r_1 and r_2 depicts the random numbers in the range between 0 and 1 respectively. j illustrates a new position of ith beluga whale in jth dimension.

3.3.2. Exploration phase

Levy flight scheme is performed in the exploitation phase to enhance convergence. Through Levy flight approach as the supposition, a mathematical expression of this dataset is formulated in Eq. (4) as follows:

$$X_i^{T+1} = r_3 * X_{best}^{T} - r_4 * X_i^{T} + C_1 * L_F * \left(X_r^{T} - X_i^{T}\right) \quad (4)$$

Where, X_i^{T+1} denotes the position of ith whale in the solution space at next iteration $T+1$; r_3 and r_4 denotes the random coefficients; X_{best}^{T} depicts the best solution; X_i^{T} denotes the current position of ith whale in a solution space; C_1 denotes the control parameter and L_F depicts the c denotes the Levy FLight. C_1 is estimated in Eq. (5). The L_F function is estimated through Eqs. (6) and (7) as follows:

$$C_1 = 2r_4 * \left(\frac{1-T}{T_{max}}\right) \quad (5)$$

$$L_F = 0.05 * \left(\frac{u*\sigma}{|v|^{\frac{1}{\beta}}}\right) \quad (6)$$

$$\sigma = \left(\frac{(1+\beta)*sin\left(\frac{\pi*\beta}{2}\right)}{((1+\beta)/2)*\beta*2^{\left(\frac{\beta-1}{2}\right)}}\right) \quad (7)$$

Where, σ denotes the scaling factor; β denotes the parameter; the constant fixed to 1.5 and u and v denotes the random values through normal distribution estimated through Eqs. (8) and (9) as follows:

$$u = randn(1, \dim) * \sigma \quad (8)$$

$$v = randn(1, \dim) \quad (9)$$

3.3.3. Whale Falls phase

The behavior of whale fall is represented in each iteration by simulating minor modifications within the groups, based on the likelihood of individuals in the population. Assumed that the beluga whales are moved or terminated as well as fallen within an ocean. The positions of the beluga whales and a degree of the whale fall are utilized for determining an updated position to handle the constant population size, which is formulated in Eqs. (10-13) as follows:

$$X_i^{T+1} = r_5 * X_i^T - r_6 * X_r^T + r_7 * X_{step}^T \quad (10)$$

$$X_{step} = (u_b - l_b) * exp(-C_2 * T/T_{max}) \quad (11)$$

$$C_2 = 2 * W_f * n \quad (12)$$

$$W_f = 0.1 - 0.05 * T/T_{max} \quad (13)$$

Where, r_5 , r_6 and r_7 denotes the random coefficients; X_{step}^T denotes the movement direction which following a global best; W_f depicts the weighting factor; n denotes the number of individuals.

3.3.4. Dynamic Dual Elite Learning

Elite learning involves using high-quality individuals within a population to direct other individuals toward modelling a best position over time, with the goal of improving the effectiveness of the BWO approach. This research proposes a dynamic, elite learning plan that utilizes various positions like current, optimal and suboptimal to develop the candidate solutions of further groups. The dynamic factor is developed to change a learning weight of two elite individuals. A dynamic learning factor is integrated through the fitness value (Root Mean Squared Error (RMSE)) of the Beluga Whale. The generation of the candidate solution and the dynamic learning factor is formulated in Eqs. (14) and (15) as follows:

$$X_i^*(t+1) = r_1 \times F \times (X_{second}(t) - X_i(t)) - r_2 \times (1-F) \times X_{best}(t) \quad (14)$$

$$F = \frac{F_{best}(t)}{F_{best}(t)+F_{second}(t)} \quad (15)$$

Where, $X_{best}(t)$ and $X_{second}(t)$ denotes an individual position ordered first and second in fitness at tth iteration. F a dynamic learning factor; $F_{best}(t)$ and $F_{second}(t)$ demonstrates a fitness value of first and second individual positions ordered in t th iteration.

3.3.5. Sinusoidal mutation

To solve the local optima problem, this research proposes another strategy of utilizing the sine function to direct a present optimal beluga whale to make the mutations to pretend an immediate jump off the beluga whale, which is supportive for an approach to jump away of local optimal and enhances a likelihood of identifying the optimal position. A sine function value on a range of $[0,2\pi]$ is $[-1,1]$ to provide a bounded output that can be used for various optimization techniques, including guiding individuals toward more promising areas of the solution space during the search process. This flexibility allows the approach to explore a narrower range over a present global optimal position, enhancing a local exploitation capability of BWO algorithm and mitigating the issue of local optima. The sinusoidal mutation process of the BWO approach is formulated in Eqs. (16) and (17) as follows:

$$X'_{best} = \sin\left(2\pi \times rand(1, \dim)\right) \oplus X_{best}, if\ rand < c \quad (16)$$

$$X_{best} = \begin{cases} X'_{best}, f(X'_{best}) < T \\ X_{best}, otherwise \end{cases} \quad (17)$$

Where, X_{best} denotes a global optimal position; $\oplus$ depicts a dot product; c denotes the accepted mutation probability.

3.4 Classification using RNN-BiLSTM

The selected features from the DS-BWO approach are provided as input to the RNN-BiLSTM layer for the classification. The RNN approach involves the idea of memory which supports them to store the data of the prior inputs to develop a further sequential output. The LSTM network is the kind of RNN approach that solves the vanishing gradient problem by introducing the different kinds of gates like input and forgot. Moreover, this network is significant when it comes to handling long-range influences.

This research employs the BiLSTM approach, a specialized RNN technique that processes data in both directions. This means the BiLSTM considers the data from both past and future time steps in a sequence, enhancing its ability to capture context. On the other hand, a traditional LSTM processes the data only in the forward direction, using information from past time steps but not from future ones. By capturing the sequences in both directions, BiLSTM provides a better understanding of patterns, especially when contextual information from both ends of the sequence is important. The BiLSTM approach effectively performs through long-term dependencies of the time series data, since every token encoding involves the past and future context data.

In this research, RNN and BiLSTM approach are integrated for the binary class of intrusion detection. The network involves five layers such an input layer, RNN layer with 100 hidden units, BiLSTM layer through 200 hidden units, Fully Connected (FC), Softmax as well as classification output layers. A BiLSTM network is represented to be reliable as well as effective for simulating the sequence which is utilized for different purposes as well as large dependencies. After the RNN module, the BiLSTM layer is utilized. To solve the overfitting problem, the two dropout layers are utilized which are positioned under the ReLU activation function as well as the BiLSTM layer. Utilizing these dropout layers enables to solving an overfitting issue. Simultaneously, the dropout provides the objective of minimizing a generalization error, which is followed in integration with the development of the number of layers into the neural network. The parameter settings of the proposed RNN-BiLSTM approach are: Softmax activation function, Adam Optimizer, categorical cross entropy loss function, o.001 learning rate, 10 epochs and 32 batch size.

4. Experimental results and discussion

A significance of the proposed DS-BWOA approach is implemented on Python 3.10 with the system requirements of Intel i5 processor, 16GB RAM and Windows 10 OS. The proposed DS-BWO approach is estimated through different performance metrics named accuracy, precision, recall and F1-score. The mathematical expressions of these performance metrices are formulated in Eq. (18) to (21) as follows:

Accuracy: Proportion of the overall sample that the classifier correctly classifies the sample.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{18}$$

Precision: proportion of positive samples in the positive examples classified by the classifier.

$$Precision = \frac{TP}{TP+FP} \tag{19}$$

Recall: Proportion of the number of samples predicted as positive samples by the classifier to the total positive samples.

$$Recall = \frac{TP}{TP+FN} \tag{20}$$

F1-score: It is the weighted average of Precision and Recall, which is used to integrate the scores of precisions and recall.

$$F1 - score = 2 * \frac{Precsion*Recall}{Precision+Recall} \tag{21}$$

Where, TP, TN, FP and FN demonstrates the True Positive; True Negative; False Positive and False Negative.

4.1 Performance analysis

The significance of the proposed method is examined through numerous performance metrics based on CIC-IDS-2018, NSL-KDD and UNSW-NB15 datasets. Table 1 demonstrates the performance analysis of the feature selection results. The different optimization algorithms such as Grey Wolf Optimization (GWO), Tuna Swarm Optimization (TSO) and Golden Jackle Optimization (GJO) are estimated and compared with the proposed DS-BWO approach. In CIC-IDS2018 dataset, the proposed DS-BWO approach attains a better accuracy of 0.999, 0.999 and 0.946 on CIC-IDS2018, NSL-KDD and UNSE-NB15 datasets respectively.

Table 1. Performance Analysis of the feature selection

Dataset	Methods	Accuracy	Precision	Recall	F1-score
CIC-IDS2018	GWO	0.972	0.968	0.966	0.967
	TSO	0.968	0.963	0.961	0.962
	GJO	0.977	0.96.5	0.964	0.964
	DS-BWO	0.999	0.999	0.999	0.999
NSL-KDD	GWO	0.965	0.950	0.945	0.947
	TSO	0.959	0.945	0.940	0.942
	GJO	0.961	0.948	0.943	0.945
	DS-BWO	0.999	0.999	0.999	0.999
UNSW-NB15	GWO	0.918	0.914	0.911	0.912
	TSO	0.915	0.910	0.908	0.909
	GJO	0.916	0.912	0.909	0.910
	DS-BWO	0.946	0.948	0.945	0.946

Table 2. Performance analysis of different classifiers

Dataset	Methods	Accuracy	Precision	Recall	F1-score
CIC-IDS2018	CNN	0.975	0.976	0.975	0.975
	RNN	0.982	0.982	0.982	0.982
	LSTM	0.978	0.977	0.977	0.977
	RNN-BiLSTM	0.999	0.999	0.999	0.999
NSL-KDD	CNN	0.982	0.981	0.982	0.981
	RNN	0.989	0.988	0.988	0.988
	LSTM	0.985	0.986	0.985	0.985
	RNN-BiLSTM	0.999	0.999	0.999	0.999
UNSW-NB15	CNN	0.923	0.925	0.923	0.924
	RNN	0.937	0.938	0.937	0.937
	LSTM	0.930	0.930	0.930	0.930
	RNN-BiLSTM	0.946	0.948	0.945	0.946

Table 2 represents the performance analysis of the various classifiers based on CIC-IDS2018, NSL-KDD and UNSW-NB15 datasets. The different classifiers such as CNN, RNN and LSTM are estimated and compared with the proposed RNN-BiLSTM approach. By using both forward and backward directions, BiLSTM captures both past and future dependencies in the network traffic, improving the system's ability to detect complex intrusion patterns that might otherwise be missed in a unidirectional model.

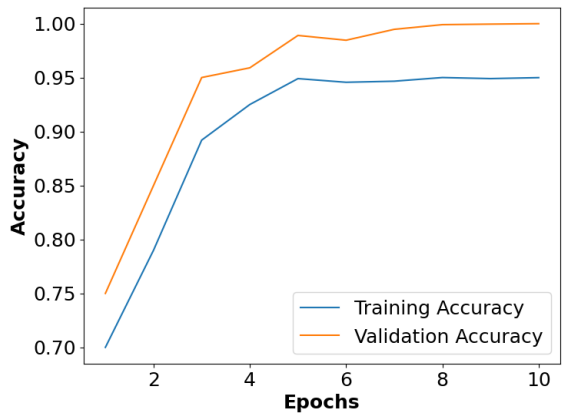

Figure. 2 Accuracy function of RNN-BiLSTM approach for CIC-IDS2018 dataset

4.1.1. Accuracy function

Fig. 2 illustrates the accuracy function for the RNN-BiLSTM on CIC-IDS2018 dataset. The maximum accuracy of the RNN-BiLSTM is reached on the 10th epoch.

4.1.2. Confusion matrix

Fig. 3 describes a confusion matrix for the CIC-IDS2018 dataset. Fig. 4 describes the confusion matrix for the NSL-KDD dataset. Fig. 5 describes a confusion matrix for the UNSW-NB15 dataset. A confusion matrix is estimated with the True class and predicted class.

4.1.3. ROC curve

Figure 6 demonstrates the ROC curve of the proposed method based on CICIDS-2018 dataset. Figure 7 illustrates the ROC curve for the NSL-KDD dataset. Figure 8 represents the ROC curve for UNSW-NB15 dataset.

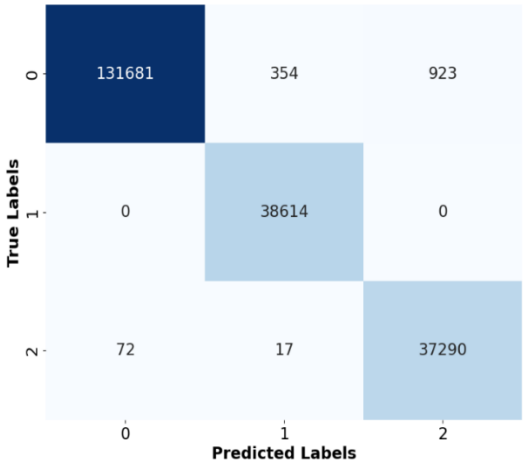

Figure. 3 Confusion Matrix for CIC-IDS 2018 dataset

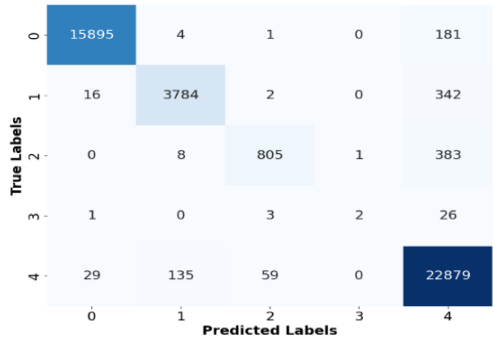

Figure. 4 Confusion Matrix for NSL-KDD dataset

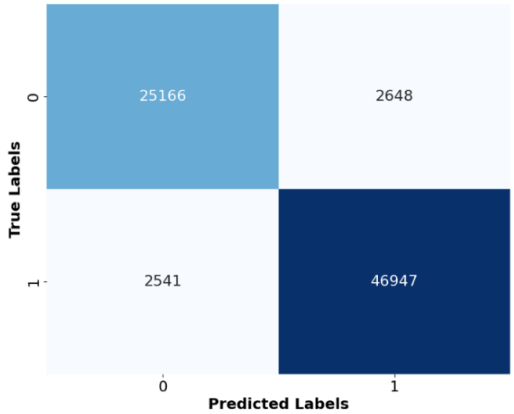

Figure. 5 Confusion Matrix for UNSW-NB15 database

4.2 Comparative analysis

This section discusses the comparison results of the proposed DS-BWO approach with the existing approaches based on the three standard datasets CIC-IDS2018, NSL-KDD and UNSE-NB15. The existing methods like [16-19] are considered in this research to estimate a significance of the proposed DS-BWO method. Table 3 represents the comparison results of the proposed DS-BWO approach with existing approaches.

4.3 Discussion

In this section, the limitations of the existing works and how the proposed DS-BWO with RNN-BiLSTM overcome these limitations are discussed along with their advantage. The limitations of the existing works are as follows: In the Hybrid network [16] approach, the greedy strategy reduces the number of layers, but it does not fully optimize feature selection for diverse time-series data. HIDM

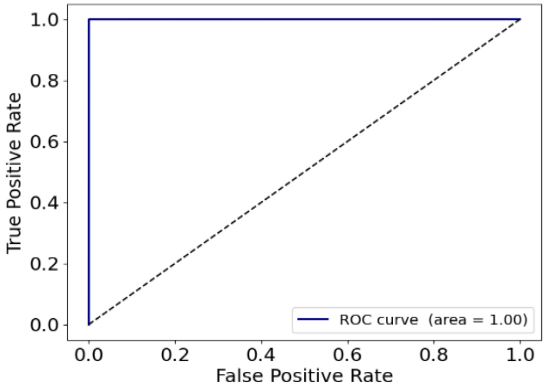

Figure. 6 ROC curve for CIC-IDS2018 dataset

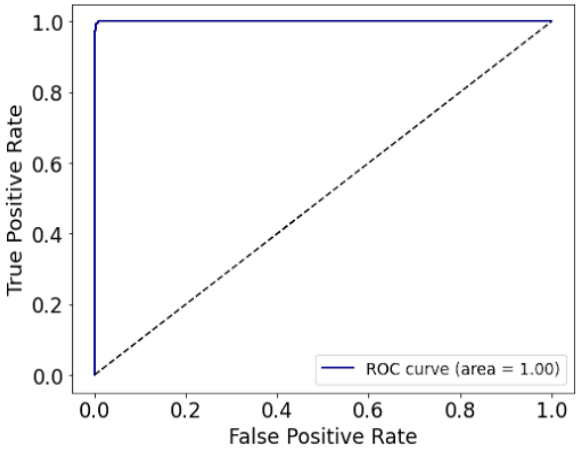

Figure. 7 ROC curve for NSL-KDD dataset

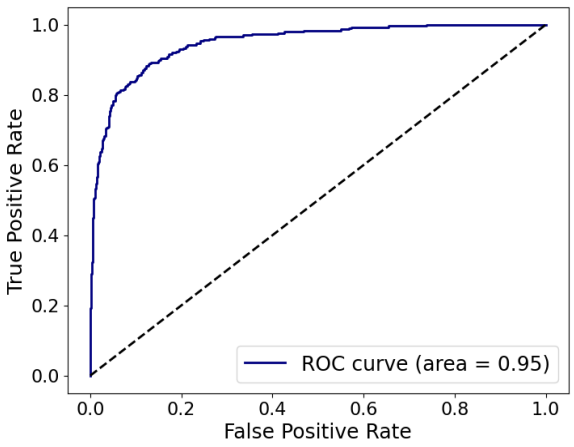

Figure. ROC curve for UNSW-NB15 dataset

Table 3. Comparison results of the proposed DS-BWO approach with existing methods (NA = Not Available)

Dataset	Methods	Accuracy	Precision	Recall	F1-score
CIC-IDS2018	MAFSIDS+DQL [18]	0.968	NA	NA	0.963
	XGBoost + Z-score normalization [19]	0.997	0.920	0.920	0.949
	Proposed DS-BWO	0.999	0.999	0.999	0.999
NSL-KDD	Hybrid network [16]	0.93	0.92	0.92	0.929
	HIDM [17]	0.998	0.998	0.998	0.998
	MAFSIDS+DQL [18]	0.991	-	-	0.991
	Proposed DS-BWO	0.999	0.999	0.999	0.999
UNSW-NB15	Hybrid network [16]	0.93	0.91	0.99	0.952
	HIDM [17]	0.927	0.951	0.927	0.935
	Proposed DS-BWO	0.946	0.948	0.945	0.946

[17] was trained through the majority class which neglected the minority class, resulted in misclassification of data and reduced the accuracy of classification of the intrusions. DRL [18] approach with an easy policy network provided fast responses, resulting in suboptimal feature selection and limited generalization when applied to challenging network environments. XGBoost + Z-score normalization [19] involved potential loss of critical information during feature elimination with PCA and RF, which resulted in minimized classification accuracy. To overcome these challenges, this research proposes the DS-BWO approach for the important feature selection. The DS-BWO integrated the principles of swarm intelligence with differential evolution strategies, enabling more efficient exploration and exploitation of the feature space. This results in a more effective search for optimal feature subsets. The RNN-BiLSTM architecture effectively adapts to different types of intrusion detection tasks, including anomaly detection and pattern recognition, making it adaptable to different network environments and attack types.

5. Conclusion

This research proposes the DS-BWO approach for the important feature selection for the IDS. The hybrid nature of DS-BWO resulted in faster convergence over optimal solutions compared to conventional approaches. By leveraging the strengths of both differential evolution and whale optimization algorithms, this method efficiently identifies important features. The RNN-BiLSTM approach is used for the classification of IDS into binary classes, which automatically learn relevant features from raw data, minimizing the requirements for extensive manual feature engineering and enabling a more efficient setup process. The three benchmark datasets such as CIC-IDS2018, NSL-KDD and UNSE-NB15 are utilized to estimate the effectiveness of the proposed DS-BWO approach. The experimental results demonstrate that the proposed DS-BWO achieve a better accuracy of 0.999, 0.999 and 0.949 on datasets CIC-IDS2018, NSL-KDD and UNSE-NB15 respectively when compared to the existing works like MAFSIDS+DQL and XGBoost + Z-score normalization. Future work will consider enhancing the optimization algorithm for the important feature selection to enhance the overall model performance.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, have been done by 1st author. The supervision and project administration, have been done by 2nd author.

Notation list

Variables	Descriptions
N	Beluga whale population size
D	Dimension
B_f	Balancing factor
T	Current iteration
T_{max}	Maximum number of iterations
X_{ij}^{T+1}	Next iteration
r_1 and r_2	Random numbers in the range between 0 and 1 respectively
j	New position of ith beluga whale in jth dimension
X_i^{T+1}	Position of ith whale in the solution space at next iteration $T+1$
r_3 and r_4	Random coefficients
X_{best}^T	Best solution
X_i^T	Current position of ith whale in a solution space
C_1	Control parameter
L_F	Levy Flight
σ	Scaling factor

β	Parameter
u and v	Random values through normal distribution
r_5, r_6 and r_7	Random coefficients
X_{step}^{T}	Movement direction which following a global best
W_f	Weighting factor
n	Number of individuals
$X_{best}(t)$ and $X_{second}(t)$	Individual position ordered first and second in fitness at tth iteration
F	Dynamic learning factor
$F_{best}(t)$ and $F_{second}(t)$	Fitness value of first and second individual positions ordered in tth iteration
X_{best}	Global optimal position
$\oplus$	Dot product
c	Accepted mutation probability

References

[1] M.Y. Aldarwbi, A.H. Lashkari, and A.A. Ghorbani, "The sound of intrusion: A novel network intrusion detection system", *Computers and Electrical Engineering*, Vol. 104, p. 108455, 2022.

[2] E.M. Onyema, S. Dalal, C.A.T. Romero, B. Seth, P. Young, and M.A. Wajid, "Design of intrusion detection system based on cyborg intelligence for security of cloud network traffic of smart cities", *Journal of Cloud Computing*, Vol. 11, p. 26, 2022.

[3] E.M. Maseno, and Z. Wang, "Hybrid wrapper feature selection method based on genetic algorithm and extreme learning machine for intrusion detection", *Journal of Big Data*, Vol. 11, p. 24, 2024.

[4] R. Setiawan, R.R. Ganga, P. Velayutham, K. Thangavel, D.K. Sharma, R. Rajan, S. Krishnamoorthy, and S. Sengan, "Encrypted network traffic classification and resource allocation with deep learning in software defined network", *Wireless Personal Communications*, Vol. 127, No. 1, pp. 749-765, 2022.

[5] S. Ali, O. Abusabha, F. Ali, M. Imran, and T. Abuhmed, "Effective multitask deep learning for iot malware detection and identification using behavioral traffic analysis", *IEEE Transactions on Network and Service Management*, Vol. 20, No. 2, pp. 1199-1209, 2023.

[6] M. Bakro, R.R. Kumar, M. Husain, Z. Ashraf, A. Ali, S.I. Yaqoob, M.N. Ahmed, and N. Parveen, "Building a cloud-IDS by hybrid bio-inspired feature selection algorithms along with random forest model", *IEEE Access, Vol.* 12, pp. 8846-8874, 2024.

[7] Y. Yin, J. Jang-Jaccard, W. Xu, A. Singh, J. Zhu, F. Sabrina, and J. Kwak, "IGRF-RFE: a hybrid feature selection method for MLP-based network intrusion detection on UNSW-NB15 dataset", *Journal of Big Data*, Vol. 10, p. 15, 2023.

[8] S. Sachdeva, and A. Ali, "Machine learning with digital forensics for attack classification in cloud network environment", *International Journal of System Assurance Engineering and Management*, Vol. 13, No. Suppl 1, pp. 156-165, 2022.

[9] J. Vitorino, M. Silva, E. Maia, and I. Praça, "Reliable feature selection for adversarially robust cyber-attack detection", *Annals of Telecommunications*, 2024.

[10] S.M. Kasongo, "A deep learning technique for intrusion detection system using a Recurrent Neural Networks based framework", *Computer Communications*, Vol. 199, pp. 113-125, 2023.

[11] A. Ayantayo, A. Kaur, A. Kour, X. Schmoor, F. Shah, I. Vickers, P. Kearney, and M.M. Abdelsamea, "Network intrusion detection using feature fusion with deep learning", *Journal of Big Data*, Vol. 10, p. 167, 2023.

[12] L.C.B. Guimaraes, G.A.F. Rebello, G.F. Camilo, L.A.C. de Souza, and O.C.M.B. Duarte, "A threat monitoring system for intelligent data analytics of network traffic", *Annals of Telecommunications*, Vol. 77, No. 7-8, pp. 539-554, 2022.

[13] F. Hu, S. Zhang, X. Lin, L. Wu, N. Liao, and Y. Song, "Network traffic classification model based on attention mechanism and spatiotemporal features", *EURASIP Journal on Information Security*, Vol. 2023, p. 6, 2023.

[14] A.A. Alsulami, Q.A. Al-Haija, A. Tayeb, and A. Alqahtani, "An intrusion detection and classification system for IoT traffic with improved data engineering", *Applied Sciences*, Vol. 12, No. 23, p. 12336, 2022.

[15] C. Yao, Y. Yang, K. Yin, and J. Yang, "Traffic anomaly detection in wireless sensor networks based on principal component analysis and deep convolution neural network", *IEEE Access*, Vol. 10, pp. 103136-103149, 2022.

[16] H. Yu, C. Kang, Y. Xiao, and Y. Yang, "Network intrusion detection method based on hybrid improved residual network blocks and bidirectional gated recurrent units", *IEEE Access*, Vol. 11, pp. 68961-68971, 2023.

[17] L.K. Vashishtha, A.P. Singh, and K. Chatterjee, "HIDM: A hybrid intrusion detection model for cloud based systems", *Wireless Personal Communications*, Vol. 128, No. 4, pp. 2637-2666, 2023.

[18] K. Ren, Y. Zeng, Y. Zhong, B. Sheng, and Y. Zhang, "MAFSIDS: a reinforcement learning-based intrusion detection model for multi-agent feature selection networks", *Journal of Big Data*, Vol. 10, p. 137, 2023.

[19] S. Songma, T. Sathuphan, and T. Pamutha, "Optimizing Intrusion Detection Systems in Three Phases on the CSE-CIC-IDS-2018 Dataset", *Computers*, Vol. 12, No. 12, p. 245, 2023.

[20] F. Ullah, S. Ullah, G. Srivastava, and J.C.W. Lin, "IDS-INT: Intrusion detection system using transformer-based transfer learning for imbalanced network traffic", *Digital Communications and Networks*, Vol. 10, No. 1, pp. 190-204, 2024.

[21] H.K. Bella, and S. Vasundra, "A Novel Framework based on Extra Tree Regression Classifier and Grid Search LSTM for Intrusion Detection in IoT and Cloud Environment", *International Journal of Intelligent Engineering & Systems*, Vol. 17, No. 4, pp. 504-517, 2024, doi: 10.22266/ijies2024.0831.39.

[22] R.A. Abed, E.K. Hamza, and A.J. Humaidi, "A modified CNN-IDS model for enhancing the efficacy of intrusion detection system", *Measurement: Sensors*, Vol. *35*, p. 101299, 2024.

[23] CIC-IDS2028 dataset link: https://www.kaggle.com/datasets/towhidultonmoy/cicids2018. (Accessed on October 2024)

[24] NSL-KDD dataset link: https://www.kaggle.com/datasets/hassan06/nslkdd (Accessed on October 2024)

[25] UNSW-NB15 dataset link: https://www.kaggle.com/datasets/mrwellsdavid/unsw-nb15 (Accessed on October 2024)

[26] M.A. Faizin, D.T. Kurniasari, N. Elqolby, M.A.R. Putra, and T. Ahmad, "Optimizing Feature Selection Method in Intrusion Detection System Using Thresholding", *International Journal of Intelligent Engineering & Systems*, Vol. 17, No. 3, pp. 214-226, 2024, doi: 10.22266/ijies2024.0630.18.

[27] D. Chen, Q. Song, Y. Zhang, L. Li, and Z. Yang, "Identification of network traffic intrusion using decision tree", *Journal of Sensors*, Vol. 2023, No. 1, p. 5997304, 2023.

[28] S. Alomari, K. Kaabneh, I. AbuFalahah, S. Gochhait, I. Leonova, Z. Montazeri, M. Dehghani, and K. Eguchi, "Carpet Weaver Optimization: A Novel Simple and Effective Human-Inspired Metaheuristic Algorithm", *International Journal of Intelligent Engineering & Systems*, Vol. 17, No. 4, pp. 230-242, 2024, doi: 10.22266/ijies2024.0831.18.

[29] T. Hamadneh, K. Kaabneh, O. AlSayed, G. Bektemyssova, Z. Montazeri, M. Dehghani, and K. Eguchi, "Sculptor Optimization Algorithm: A New Human-Inspired Metaheuristic Algorithm for Solving Optimization Problems", *International Journal of Intelligent Engineering and Systems*, Vol. 17, No. 4, pp. 564-575, 2024, doi: 10.22266/ijies2024.0831.43.

[30] M.A. Al-Sharqi, A.T.S. Al-Obaidi, and S.O. Al-mamory, "Apiary Organizational-Based Optimization Algorithm: A New Nature-Inspired Metaheuristic Algorithm", *International Journal of Intelligent Engineering & Systems*, vol. 17, No. 3, pp. 783-801, 2024, doi: 10.22266/ijies2024.0630.61.

[31] P.D. Kusuma, and A. Dinimaharawati, "Swarm Bipolar Algorithm: A Metaheuristic Based on Polarization of Two Equal Size Sub Swarms", *International Journal of Intelligent Engineering and Systems*, Vol. 17, No. 2, pp. 377-389, 2024, doi: 10.22266/ijies2024.0430.31.

[32] P.D. Kusuma, and M. Kallista, "Swarm Space Hopping Algorithm: A Swarm-based Stochastic Optimizer Enriched with Half Space Hopping Search", *International Journal of Intelligent Engineering and Systems*, Vol. 17, No. 2, pp. 670-682, 2024, doi: 10.22266/ijies2024.0430.54.