

Intrusion Detection Using Whale Optimization Based Weighted Extreme Learning Machine in Applied Nonlinear Analysis

P. Kaliraj¹ Subramani. B²

¹ Research Scholar, SNMV College of Arts and Science, Coimbatore, Tamil Nadu, India,
pkalirajmsc@gmail.com

²Principal and Research Supervisor, SNMV College of Arts and Science, Coimbatore, Tamil Nadu, India,
drbsubramani@gmail.com

Article History:

Received: 16-03-2024

Revised: 30-04-2024

Accepted: 19-05-2024

Abstract:

The development of computer networks and technology has led to a sharp rise in network assaults, making network intrusion a crucial area for research and the development of counter measures. The issue of network intrusion can be resolved with the advancement of artificial intelligence. We employ a Whale Optimization Algorithm (WOA) based Weighted Extreme Learning Machine (WELM) in this study to identify network intrusions. Prior to training the neural network and obtaining the output weight, ELM randomly allocates weight to the network. The ELM weights need to be optimized, thus we're utilising the whale optimisation technique to do this. To compare and evaluate the effectiveness of the model suggested in this study, NSL-KDD dataset is utilised. The experimental findings demonstrate that weighted ELM-based whale optimisation outperformed other approaches in detecting network intrusions and reducing false positive and false negative rates.

Keywords: Whale Optimization, Weighted Extreme Learning Machine, Intrusion Detection in Networking, False Alarm Reduction in Networking.

1. Introduction

The Internet has become a part of thousands of families thanks to the advancement of science and technology. It has made our lives more convenient, but it has also led to problems with network security. Scholars in network security have continually focused on Intrusion Detection System (IDS), network security systems important element. Industry and academia use a variety of technologies and measures to solve network security problems. IDS serves as the second line of defense in terms of security after the firewall. It gathers and examines network data, looks for network activity that defies security rules, and then reacts in concert with other devices. Information security events happen regularly despite the fact that people's understanding of information security is always rising. This is due to the complexity of network assaults and the wide variety of attack techniques. A significant obstacle to intrusion detection is the massive and complicated intrusion data with imbalanced labelling. In the area of network security, there is significant research value and a wide range of application possibilities for how to efficiently choose characteristics from intrusion data for multi-classification and enhance intrusion detection accuracy. Intrusion detection (ID) is an active defence technique that has steadily emerged as a crucial component of network system security. An Intrusion Detection System's (IDS) job is to spot suspicious activity on protected internal networks, such as assaults or unauthorised access. An important area of research for IDS is user behaviour modelling based on machine learning, which is used to differentiate between the system's normal and

aberrant behaviour by learning observational data such as network traffic and host audit logs. Zhu and co. [1] A meta-heuristic approach based on population known as the Whale optimisation algorithm (WOA) has been used for model prediction, parameter optimization, and other things. Wang and co. [2] WOA has a distinctive search technique, and its use The method has great overall stability, minimal parameters, and significant optimisation capabilities. The meta-heuristic algorithm has specific advantages in solving the job-shop scheduling problem. Li et al. [3] firstly applied WOA to the workshop scheduling problem, and solved JSP by using the improved WOA of quantum computing. Zhang Siqi et al. [4] proposed an improved WOA to optimize single-objective FJSP. The nonlinear convergence factor and the sine-cosine algorithm are used to enhance the WOA in this paper's FJSP solution. A mixed sine-cosine whale optimisation algorithm (SCWOA) is also proposed, and the SCWOA solution is used to minimise and maximise the FJSP of the time-to-make. Guo et al.'s [5] enhanced the Whale Optimization Algorithm (WOA), which successfully increased the Whale Algorithm's capacity for global search while assuring algorithm convergence. To address the flaw that the WOA algorithm is liable to local optimum, Huang et al. [6] combined the cosine factor and polynomial mutation approach. Although the swarm intelligence algorithm has made some progress in the field of hypersonic trajectory optimization, the swarm intelligence optimisation method still has issues with complicated restricted problem solving, including low solution efficiency, premature, and easy to slip into local optimum solutions.

To forecast how much power residents will use, Du et al. [7] used the grey seasonal variation index model and the modified whale optimisation method. The starting value and parameter optimisation of the GM (1,1) model is still an issue worth researching, according to the aforementioned literature evaluation. It would be worthwhile to do study in the area of intelligent algorithm optimisation of the GM (1,1) model's starting value and model parameters. This work suggests the GM (1,1) model optimized using the moth flame optimisation technique to minimize the fitting error of the GM (1,1) model based on the examination of the fitting error problem present in the GM (1,1) model. To solve the optimisation issue of the power exponent and initial condition of the GM (1,1) power model, Guo et al.'s collaborative initial condition and power exponent optimisation strategy was proposed. The suggested method's accuracy was better than the comparative model's. The traditional GM (1,1) model estimates parameters using the least squares criterion; additionally, there are the minimum and maximum criterion, the minimum and maximum relative error reaches the minimum criterion, the relative average error reaches the minimum criterion, and so forth.

Mei and co. [9] ELM, however, has the following drawbacks. 1) A lack of steadiness. The algorithm learns much more quickly when the calculated output weight directly using the least squares approach, but because the input weight and hidden layer threshold are determined at random, they cannot be changed to better match the real data. 2) There is weak robustness. The error's weight is not taken into account. The model's performance may be significantly impacted when the data set contains outliers or isolated points. 3. It's simple to overfit. Only take into account empirical risk; disregard structural risk. Shi [11] Aiming at the problem of asymmetric computing resource asymmetry in threat perception of intelligent vehicle networking, a low-latency threat perception mechanism based on vehicle edge computing is proposed, and the collaborative computing of surrounding trusted vehicles is used to make threat-aware intelligent decision-making, so as to solve the problem of insufficient local computing resources and cloud

computing delay and higher difficulty. Han and Li [12] In addition, the way of exchanging intrusion samples through point-to-point communication mode in the community is too slow, and the intrusion detection scheme centered on the cooperation between nodes of the Internet of Things is difficult to deal with the massive potential threats in the Internet of Things. Researchers have turned their attention to cloud computing technology with massive data processing capabilities. Using cloud computing, the samples in each IoT device can be uploaded to the cloud center, and the powerful computing power of the cloud center combined with machine learning technology is used to build a machine learning model.

Zhang et al [13] proposes a threat perception model that can perceive replay attacks and fuzzy attacks. This model uses Bloom filters to save storage resources required for threat perception. However, the acquisition of intrusion detection samples centered on IoT devices has strong limitations, and it is difficult to deal with the ubiquitous and complex threats of IoT. Zhang et al [14] proposed an in-vehicle CAN network based on a circular convolutional neural network. The intrusion detection mechanism of the bus improves the coordination security of the electronic control units in the car. Zhao [15] proposed the edge device is responsible for collecting intrusion samples from the Internet of Things devices, and configures different intrusion sample distribution strategies for different intrusion detection servers according to the pricing of intrusion detection servers. Secondly, in the proposed IoT distributed intrusion detection system model, the intrusion detection server is responsible for intrusion detection model aggregation, intrusion sample pricing, intrusion detection model management, and intrusion detection model distribution. Intrusion detection servers can measure the value of different machine learning datasets by pricing intrusion samples.

[16] Swarm intelligence optimisation techniques, which draw their inspiration from biological systems, have been in use since the 1940s. A random search optimisation technique based on group iteration is called the swarm intelligence approach. It features distributed search capabilities as well as the possibility for parallelism. It has a powerful capacity to search globally and can successfully avoid some local extremums. In recent years, it has developed into a hub for research. The bionic objects for swarm intelligence optimization mainly come from nature, such as early ant colonies, bird colonies, bee colonies, etc. Recently, different scholars have proposed many new approaches to optimisation. However, there are associated issues including limited convergence accuracy, weak local avoidance, and general optimisation for large-scale problems. Numerical tests demonstrate that these approaches are applicable to some situations. An enhanced approach based on segmented random inertia weights and an ideal feedback mechanism was presented by Liu et al. [17]. The bubble net predation technique is extended with the contraction encirclement strategy and spiral the segmented random inertia weight for the purpose of increasing the algorithm's optimisation accuracy and ability to depart the local extremum. The random walk strategy's present global optimum solution serves as the foundation for this feedback mechanism. By adding polynomial variation to the inertia weight of the cosine parameter and the ideal whale position, Liu. [18] suggested an improved approach based on the polynomial variation and cosine control factor. To initialize the population Huang [19] combined the reverse learning strategy, used the normal mutation operator to select the population, combined with the nonlinear convergence factor and the sinusoidal spiral update strategy, thus forming an improved threshold control method.

This paper consists of five sections. Section 1 gives the introduction, Section 2 deals with whale optimisation algorithm, Section 3 is analyzing the Weighted Extreme Learning Machine, Section 4 is imparting algorithm simulation and results, Section 5 draws the conclusion finally reference taken for this paper is discussed

2. Whale Optimization Algorithm

In order to mimic the hunting behaviour of humpback whales, Mirjalili of Griffith University in Australia created the whale optimisation algorithm (Whale optimisation algorithm, WOA) in 2016 [10]. The three essential elements of WOA are surrounding prey, bubble-net assaulting, and hunting for prey.

Surround Prey

When humpback whales hunt, they surround their prey. This behavior can be described by the following model

$$\begin{cases} D = |C \cdot X'(t) - X(t)| \\ X(t + 1) = X(t) - A \cdot D \end{cases} \quad (1)$$

With the help of the below given formulas coefficient vectors A and C are calculated. The iteration number that is present currently, t, is represented by the symbol t in the following formula: whale position vector of present individual X(t), X*(t) is the current ideal solution location, D is the distance between optimal position and current whale individual.

$$\begin{cases} A = 2a \cdot r - a \\ C = 2r \end{cases} \quad (2)$$

In the formula, r is a random vector between 0 and 1 and a declines linearly from 2 to 0 in the iterative phase.

2.1 Bubble Attack

The humpback whale swims in a circular motion to its prey when it is hunting, in accordance with its hunting behaviour. Here is the mathematical model:

$$\begin{cases} X(t + 1) = X^*(t) + D_p \cdot e^{bt} \cdot \cos(2\pi/) \\ D_p = |X'(t) - X(t)| \end{cases} \quad (3)$$

The distance between the prey and the whale is denoted by the symbol Dp in the equation. The shape parameter of the logarithmic spiral is denoted by the symbol b. As the whale swims in a spiral motion towards the prey, the circle of encirclement gets smaller. In order to update the whales position, it is thought that in this synchronous behavior model, a probability of 1-Pi will be used to select the spiral model and a probability of Pi will be used to select the contraction encirclement method. The mathematical Described as follows:

$$X(t + 1) = \begin{cases} X'(t) - A \cdot D & p < P_i \\ X'(t) + D_p \cdot e^{b \cdot} \cdot \cos(2\pi/) & p \geq P_i \end{cases} \quad (4)$$

2.2 Search for Prey

In addition to the bubble attack strategy of humpback whales, random predation is also an important means. If $|A| > 1$, the distance data D will be updated randomly, indicating that humpback whales conduct random searches according to each other's positions. The hunting model at this time is:

$$\begin{cases} D = |C \cdot X_{\text{rand}}(t) - X(t)| \\ X(t+1) = X_{\text{rand}}(t) - A \cdot D \end{cases} \quad (5)$$

In the formula, $X_{\text{rand}}(t)$ is the random position vector for selecting individuals from the current population.

2.3 Improvement of WOA Based on Weight Adaptation

When compared to other intelligent algorithms, WOA has the benefits of cheap execution, quick convergence, and simple computation, but it also has drawbacks including premature convergence and a tendency to easily enter local optimum. The weight in particular has a big effect on the algorithm: when the weight is high, the convergence speed is quick and the search space is broad; when the weight is low, it's difficult to miss the overall best solution, but the convergence time is sluggish. Consequently, it is vital to develop adaptive weights improvements.

The nonlinear weights S_1 and S_2 are introduced into WOA, and the calculation formula is as follows:

$$\begin{cases} S_1 = -\gamma \left[\cos \left(\pi \cdot \frac{t}{T_{\text{max}}} - \lambda \right) \right] \\ S_2 = \gamma \left[\cos \left(\pi \cdot \frac{t}{T_{\text{max}}} + \lambda \right) \right] \end{cases} \quad (6)$$

In the formula, γ is the value range of S_1 and S_2 ; λ is the value step of S_1 and S_2 . S_2 declines nonlinearly with the number of iterations and has a smaller step size in the late iteration to speed up the convergence pace. To enable the population to fully migrate to the ideal site, S_1 rises nonlinearly with the number of iterations. Add nonlinear weights S_1 and S_2 to formulae (4) and (5), which will enhance surrounding prey, bubble assault, and prey hunting, and you'll obtain:

$$X(t+1) = \begin{cases} X'(t) - S_2 \cdot A \cdot D & p < P_i \\ S_1 [X'(t) + D_p \cdot e^{bl} \cdot \cos(2\pi/)] & p \geq P_i \end{cases} \quad (7)$$

$$X(t+1) = X_{\text{rand}}(t) - S_2 \cdot A \cdot D$$

The WOA (WWOA) process based on weight adaptive improvement is shown in Figure 1

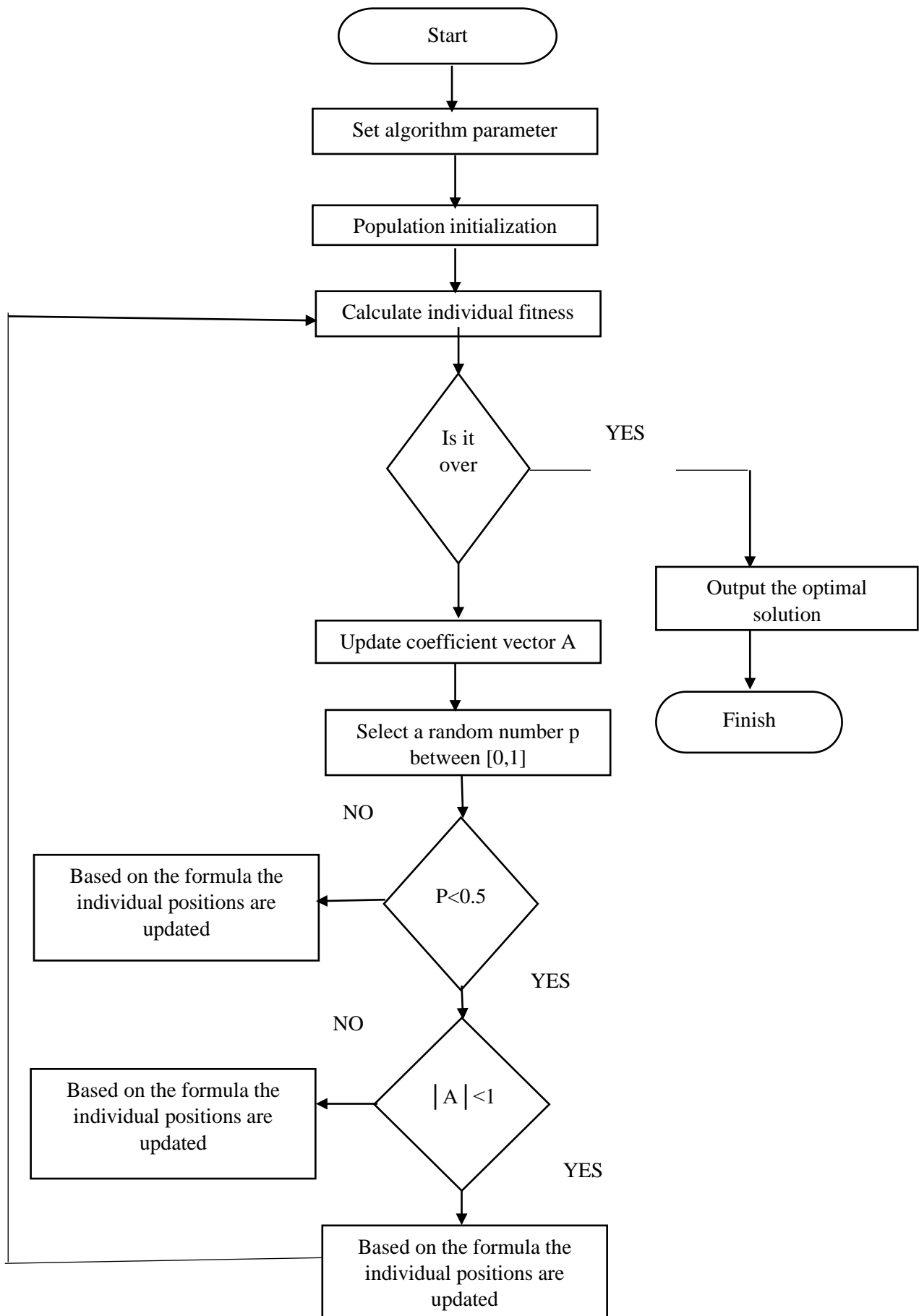


Figure. 1 Flow Chart of WWOA

2.4 Performance Analysis of Improved WOA Algorithm

In this part, traditional test functions are used for performance evaluation of the WWOA, conventional WOA, and genetic algorithm. (GA). The test functions are shown in Table 1, and each of their ideal values is zero. They include f_1 , a unimodal function that can evaluate the algorithm's speed of convergence and solution accuracy, and f_2 , a that the algorithm's ability to conduct global exploration is evaluated by the multimodal function.

Table. 1 Standard Test Function

Function	Dimension	Domain	The Optimal Value
$f_1(x) = \sum_{i=1}^N x_i + \prod_{i=1}^N x_i $	30	[-10, 10]	0
$f_2(x) = -20 \exp \left(-0.2 \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \right) - \exp \left[\frac{1}{N} \sum_{i=1}^N \cos(2\pi x_i) \right] + 20 + e$	30	[-32, 32]	0

The three methods were independently optimised 50 times for the standard test function using Matlab 2020a, and the average value and mean square error of each algorithm's optimisation skills were determined, as shown in Table 2. As can be shown, the accuracy of the WWOA algorithm while optimising f_1 is significantly greater than that of the normal WOA method and the GA algorithm; when optimizing f_2 , although the optimization accuracy is not much different, even slightly lower than standard WOA, However, the mean square error of optimization is smaller than that of WOA and GA, indicating that the algorithms stability is better.

Table. 2 The Optimized Mean Square Error and Mean Value of 3 Algorithms

Statistical Results	Algorithm	f_1	f_2
Average Value	WWOA	1.15×10^{-35}	2.64
	WOA	2.17×10^{-28}	1.86
	GA	6.07×10^{-13}	5.42E
Mean Square Error	WWOA	2.55×10^{-34}	1.14
	WOA	7.83×10^{-26}	9.57
	GA	4.99×10^{-11}	18.7

The convergence curve is a crucial metric for assessing the effectiveness of the algorithm. When solving functions, the convergence curve allows for analysis of the algorithm's convergence speed,

solution correctness, and global search capability. Figure 3 depicts the application of three methods in turn.

When the test function's convergence curve graph is solved, the following properties of WWOA are evident: The algorithm's high solution accuracy can be seen by the quicker convergence speed and low convergence value; also, the convergence curve's inflection point demonstrates a higher capacity for global search.

2.5 Energy Storage Optimal Configuration Process

The wind-light-fire-storage multi-energy system's configuration method for energy storage optimisation is shown in Figure 2. By using WWOA generation to solve the configuration algorithm, the best site and capacity for the energy storage power station are ultimately determined.

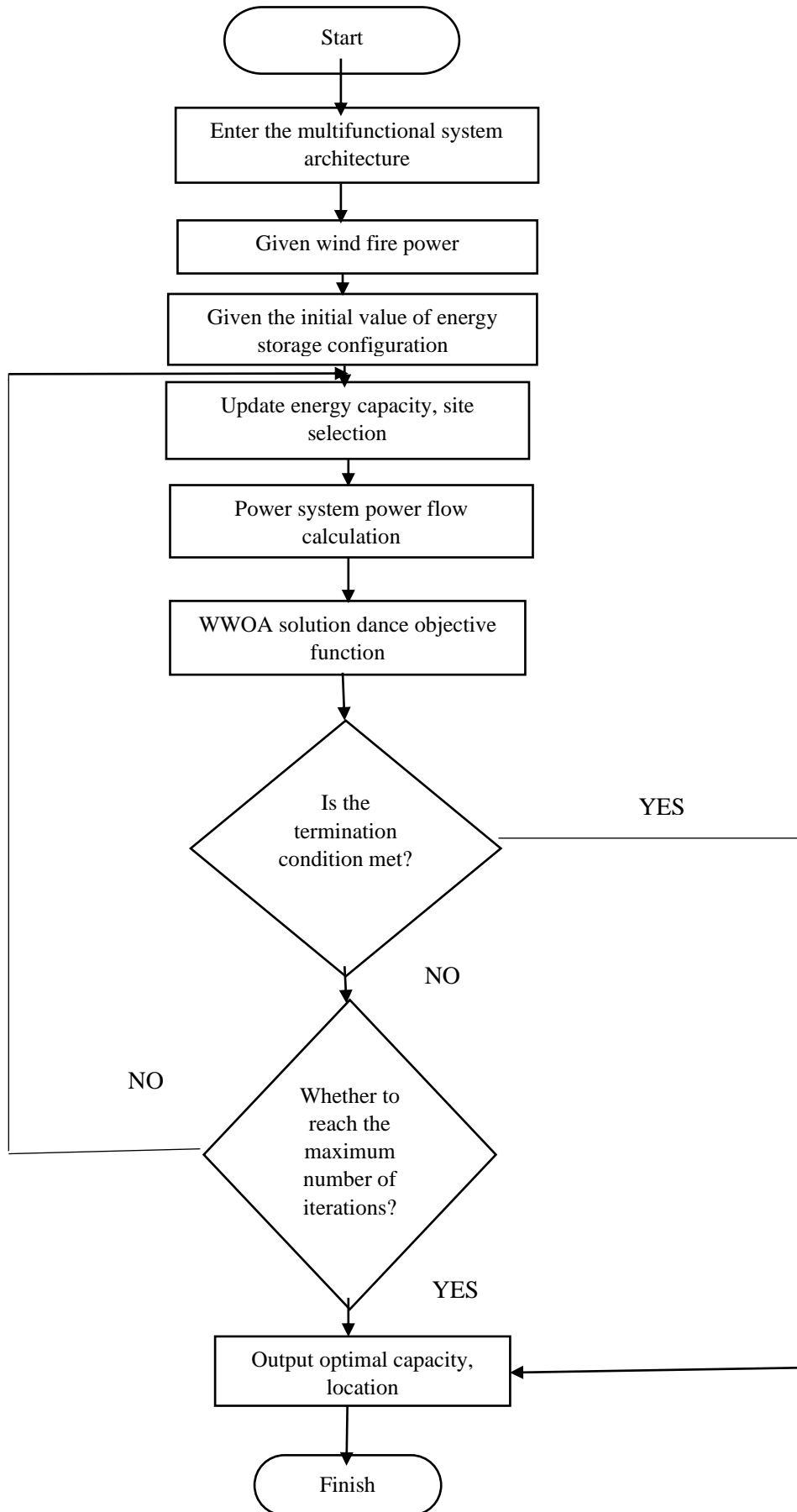


Figure. 2 Flow Chart of Configuration of Energy Storage

2.6. Weighted Extreme Learning Machine

Based on ELM, this paper establishes a weighted extreme learning machine classification model. The ELM model network structure is shown in Figure 1. In Figure 1, suppose given N training samples $\{x_i, t_i\}_{i=1}^N$, $x_i=[x_{i1},x_{i2},\dots,x_{in}]^T \in R^n$, $t_i=[t_{i1},t_{i2},\dots,t_{im}]^T \in R^m$, where n is the number of features of the sample, m is the number of categories of the sample. A feedforward neural network output model with L hidden layer expressed nodes as follows:

$$\sum_{h=1}^L \beta_h G(a_h, b_h, x_i) = t_{i,i=1,2,\dots,N} \quad (8)$$

where: h^{th} hidden layer neurons output weight β_h ; G is the neurons activation function of hidden layer; a_h, b_h respectively h individual Input weights and biases of hidden layer neurons; x is the input sample, o_i for the first The actual output value of i training samples; t_i for the first Expected output for i training samples.

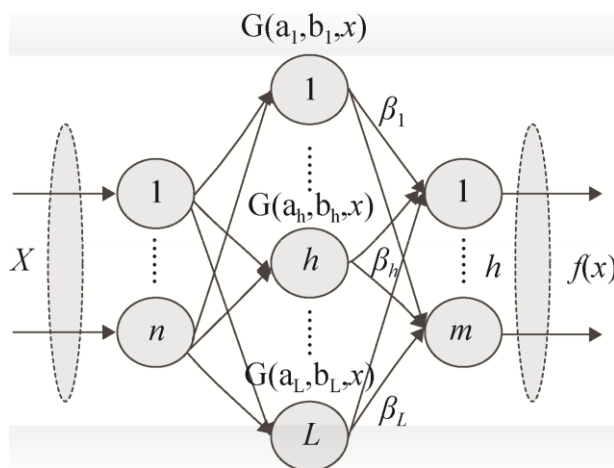


Figure 3. Basic Structure Diagram of Single Hidden Layer Feed forward Neural Network

For a quantity of N training samples, $\{x_i, t_i\}_{i=1}^N, x_i \in R^n$ There exists a (a_h, b_h) and β_h with $\sum_{i=1}^L \|o_i - t_i\| = 0$, so that the training set may be approximated by the feed forward single-hidden layer neural network model with zero error. $\{x_i, t_i\}_{i=1}^N, x_i \in R^n$ which is,

$$\sum_{h=1}^L \beta_h G(a_h, b_h, x_i) = t_{i,i=1,2,\dots,N} \quad (9)$$

The equation can be further simplified as: $H\beta=T$.

Where T is the anticipated output matrix corresponding to the training samples; H is the hidden layers output matrix; β is the hidden layer matrix output weight.

The input weights a_h of the hidden layer and the bias b_h of the hidden layer are produced at random when the network parameters are initialised during the training phase of ELM, and they remain constant throughout the training and testing processes. The whole training procedure is to acquire the output weight matrix hidden layer in the ELM model in order to create an entire classification model since the input training samples, biases of the hidden layer and the input weights, and the predicted output are all known.

From the Moore-Penrose generalized inverse matrix H^+ of the hidden layer output matrix H , we can get,

$$\hat{\beta} = H^+T \quad (10)$$

In the formula: There are many ways to calculate H^+ . In ELM, the orthogonal projection method (KKT) is usually used to solve H^+ . When $H^T H$ is a non-singular matrix, $H^+ = (H^T H)^{-1} H^T$; when HH^T is a non-singular matrix, $H^+ = H^T (HH^T)^{-1} H$.

In order to solve equation (2), a sufficiently small regular term $1/C$ is added to the diagonal of HH or HH^T , so that the classification model has better stability and generalization performance. The hidden layers output weight can be expressed as,

$$\hat{\beta} = \begin{cases} H^T (I/C + HH^T)^{-1} T, & N < 1 \\ (I/C + H^T H)^{-1} H^T T, & N \geq 1 \end{cases} \quad (11)$$

ELMs output function can be expressed as,

$$f(x) = h(x)\hat{\beta} = \begin{cases} h(x)H^T \left(\frac{I}{C} + HH^T\right)^{-1} T, & N < 1 \\ h(x) \left(\frac{I}{C} + H^T H\right)^{-1} H^T T, & N \geq 1 \end{cases} \quad (12)$$

The categorization issue demonstrates that the distribution of all the categorised sample data is not uniform. A technique based on ELM called Weighted Extreme Learning Machine (WELM) is used to address the classification problem of unbalanced data. According to the weighting scheme, assign a weight to each sample:

Weighting scheme one W1: Automatic weighting scheme:

$$W1 = \frac{1}{\text{Count}(t_i)} \quad (13)$$

where $\text{Count}(t_i)$ is the number of samples of class t_i in the training sample.

Push the minority class to majority class ratio in the direction of 0.618:1 using weighting system W2 (the golden ratio). This strategy sacrifices the majority class' classification accuracy in favour of the minority group's classification accuracy.

$$W2 = \begin{cases} \frac{0.618}{\text{Count}(t_i)}, t_i \text{ belong to the majority class} \\ \frac{1}{\text{Count}(t_i)}, t_i \text{ belong to the minority class} \end{cases} \quad (14)$$

The WELM hidden layer output weights can be expressed as,

$$\hat{\beta} = H^+T = \begin{cases} H^T (I/C + WHH^T)^{-1} WT, & N < 1 \\ (I/C + H^T WH)^{-1} H^T WT, & N \geq 1 \end{cases} \quad (15)$$

In the formula: diagonal matrix of $N \times N$ represents the weighting matrix; N samples represents the N main diagonal elements, and for various sample categories different weights are assigned, and the weighted weights of the same category are equal.

When the hidden layer feature map $h(x)$ is not known, the kernel matrix is defined as,

$$\Omega_{ELM} = HH^T: \Omega_{ELM_{i,j}} = h(x_i)h(x_j) = K(x_i, x_j) \quad (16)$$

In the formula: Ω_{ELM} is the kernel matrix, and the Mercer condition must be met by the kernel function K . Common kernel functions include the Gaussian kernel function, the polynomial kernel function, the radial basis kernel function, and the linear kernel function. It is possible to formulate the output function expression (12) as by formula. (16) $f(x) = h(x)\hat{\beta} = h(x)H^T \left(\frac{I}{C} +$

$$WHH^T)^{-1}WT = \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_n) \end{bmatrix} (I/C + W\Omega_{ELM})^{-1}WT \quad (17)$$

Therefore, the training process of the classification model based on the weighted extreme learning machine is:

- (1) Randomly set the weights of the input a_h and bias b_h of the hidden layer, where $h=1,2, \dots, L$;
- (2) According to the weighting scheme, assign weights to each sample, and calculate the weighting matrix W ;
- (3) Calculate the kernel matrix Ω_{ELM} according to the selected kernel function;
- (4) Calculate the output using equation (17).

Weighted Extreme Learning Machine Algorithm Based on Whale Optimization

The feed forward neural network WELM used in this study has the advantages of shorter training times, improved generalization capabilities, and global WOA optimisation. The WELM network's hidden layer input weight supports are optimized via the Whale Optimization method (WOA) method. Processing raises the minority attacks recall rate during network assaults, balances the data during network intrusion detection, and keeps WELM from opting for local best solutions. The algorithm flow in the study is shown in Figure 2.

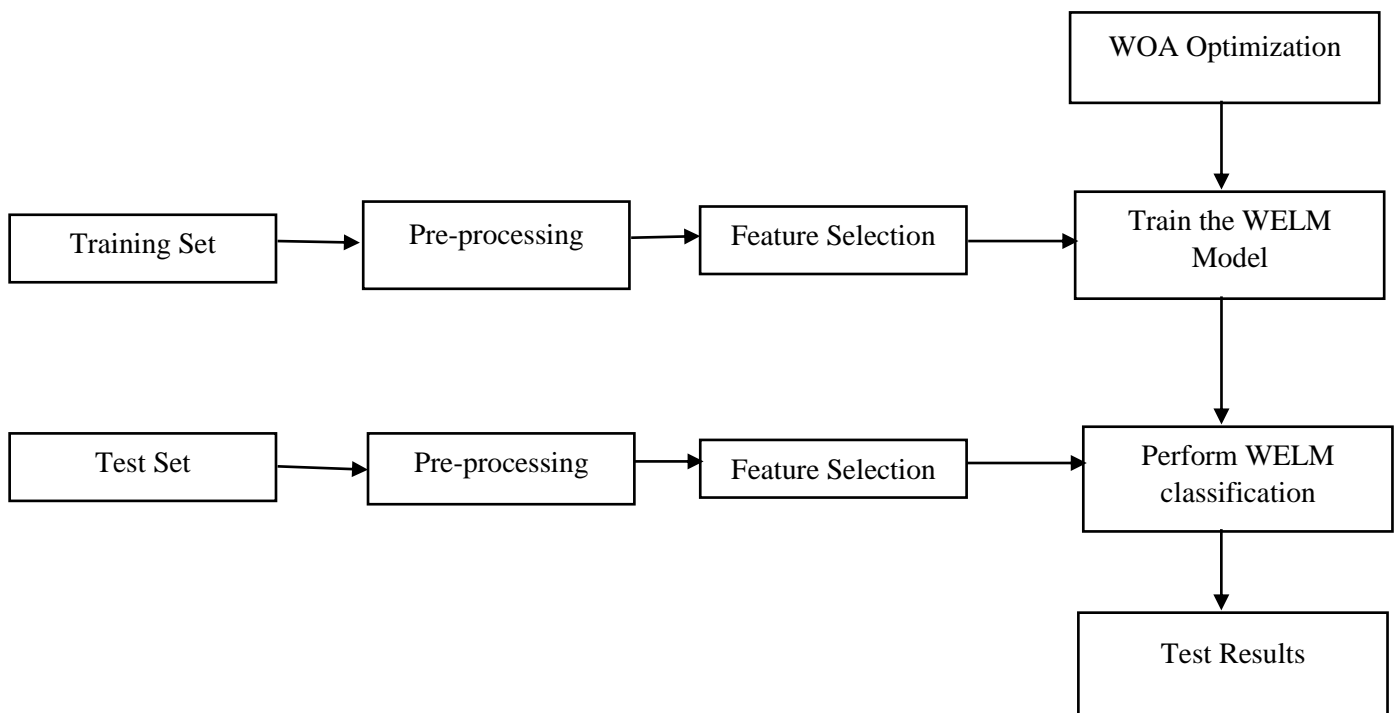


Figure. 4 Intrusion Detection Model using Weighted ELM Classification

3. Result Analysis and Algorithm Simulation

Algorithm simulation and result analysis is done in four stages preprocessing and data set selection, performance measures, data dimensionality reduction processing and finally simulation and result analysis.

3.1 Pre-Processing and Data Set Selection

In this study, the NSL-KDD dataset is chosen as the experimental data set, and the test set and training set for the experiment are KDDTest+ and KDDTrain+ from the NSL-KDD data package. Each dataset consists of 42 dimensions, the first 41 of which are dataset features and the final 12 of which are dataset label bits. The tag bits provide standard information as well as 39 distinct types of attacks, which are broken down into four groups: DOS, R2L, probe and U2R. The training set has twenty-one distinct forms of intrusion attacks, whereas the exam set contains eighteen different types. The test set that will be used to determine how effectively the intrusion detection algorithm can identify unknown assaults most likely includes these intrusion attempts. The distribution of each label class in the test sets and training sets and is shown in Table 1.

Table 1. Number of Features in Each Subclass in Training and Test Set

Class	Normal	DOS	U2R	R2L	Probe	Total
Training Set	67343	45927	52	995	11656	125973
Test Set	9711	7458	200	2754	2421	22544

Before training the model, the data in the dataset needs to be pre-processed:

Transform the character type feature into the numerical NSL-KDD dataset's forty-one-dimensional feature, as well as its third-dimensional feature (Service), fourth-dimensional feature (Connection) and second-dimensional feature (Protocol Type). It is necessary to transform the state (Flag), which is a character type characteristic, into a numerical type. In the second dimension feature, TCMP should be indicated as 1, UDP as 2, and TCP as 3. The 67 service types in the third dimension feature are respectively recorded as 1 to 67 in the alphabetical order by their names. In the fourth dimension feature, the 11 kinds of Flag states are recorded as 1~11 respectively. In the 42nd dimension label bits, there are 5 types of labels: DOS, Normal, Probe, U2R and R2L, which are recorded as 0~4 respectively.

Formulas (18) and (19) are utilised to analyse the numerical data set from the previous stage, and the measurement units between different eigenvalues are unified to lessen the detection results brought on by the difference in measurement units. influence.

$$\text{Normalized formula: } x_1 = \frac{x - \bar{x}}{\sigma} \quad (18)$$

In the formula: σ is the eigenvalues standard deviation; x is the eigenvalue; x_1 is the standardization result of the dimension of each data sample, \bar{x} is the eigenvalues mean value.

The normalization formula is:

$$\text{Normalized formula: } x_2 = \frac{x_1 - x_{1min}}{x_{1max} - x_{1min}} \quad (19)$$

In the formula: x_{1max} is the sample with highest value processed by formula for the dimension feature mentioned (18); x_{1min} is sample with highest value processed by formula for the dimension feature mentioned (18); x_2 is the value of each data sample. The result after dimension feature normalization.

3.2 Performance Measures

In this paper, the three performance analysis indicators of false alarm rate, recall rate and precision rate are used to analyze the advantages and disadvantages of the algorithm in this paper.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (20)$$

$$\text{False Positive Rate}(FPR) = \frac{FP}{FP+TN} \quad (21)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (22)$$

Where FN represents False Negative, TN represents True Neutral, FP represents False Positive and TP represents True Positive and respectively.

3.3 Data Dimensionality Reduction Processing

The pre-processed forty-one dimensional data features are then evaluated for correlation using the Pearson correlation coefficient method, yielding a weighted ranking diagram of each feature, in which the ranking order is organised in accordance with the NSL-KDD feature order as shown in Figure 3.

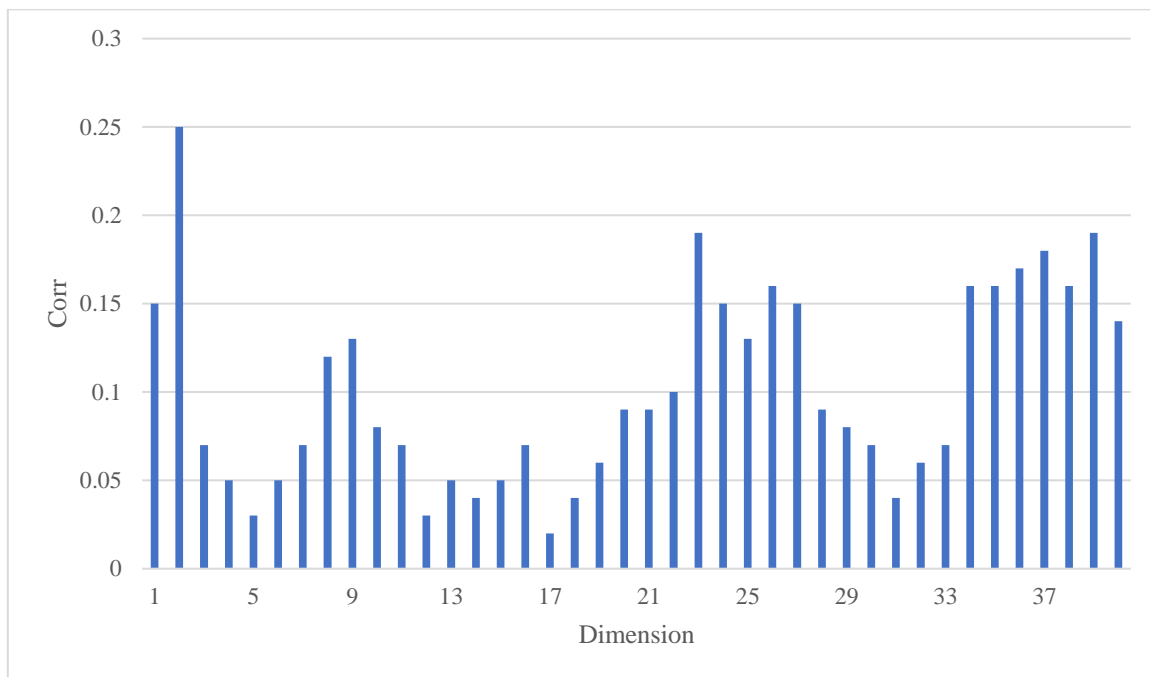


Figure.5 Sorted Feature Weight Graph

From the experiment, correlation coefficient feature ($\text{corr} > 0.05$) were taken as the test and training set after dimension is reduced. And hence, the dimensionality-reduced test and training set contain eighteen feature dimensions.

4. Simulation and Result Analysis

The simulation environment of this paper is the intel i7 processor 3.2GHz, 8GB memory, which is realized by compiling a simulation program using MATLAB.

(1) Comparison experiment between Weighted ELM and different ML algorithms

Eighteen layer nodes of Weighted ELM are set, the output node layers are 5, the hidden node layers are 250.

The kernel function for Case 2 is chosen to be the sigmoid kernel function. This study compares the approach with the detection outcomes of the random forest (RF), K-Nearest Neighbor (KNN), BP neural network, and extreme learning machine (ELM) algorithms. Table 2 displays the results of the comparison test.

Table 2 demonstrates that while the weighted extreme learning machine algorithm (WELM) has a higher recall rate than other machine learning algorithms, its detection effects for Dos, Normal, Probe, and U2R and R2L are comparable to those of other machine learning techniques. Because the training set's sample size is relatively small and the data set is unbalanced due to the two principal assaults' reduced appearance in real-world scenarios, U2R and R2L recall rates are still quite low.

(2) Relative analysis of WOA-WELM algorithm and WELM

Table 2. Prediction Results of Various Techniques

Detection Model	Recall (%)					Accuracy (%)	False Alarm Rate (%)	Detection Time(sec)
	Normal	DOS	U2R	R2L	Probe			
KNN	97	76	34	11	63	81	34	3.1
BPN	95	74	36	24	62	81	32	3.1
SVM	97	73	33	45	66	82	32	3.0
ELM	96	77	31	46	61	83	27	2.4
WELM	97	78	43	44	67	85	12	2.2
FOAELM	97	77	44	46	67	87	7	1.7
KHOELM	97	78	44	47	66	87	2	1.5
WOAELM	98	79	45	48	68	88	2	1.3

The global optimisation process that the WOA-optimized WELM algorithm underwent has boosted the recall rate of four distinct attack kinds, as seen in Table 2. In terms of classification results, WOA-WELM performs better than KHO-WELM, especially when it comes to the recall rate of U2R attacks, which is up 8% compared to WELM, the test set classification accuracy rate, which is up 4%, and the false alarm rate, which is down 4.5%. Under the identical experimental circumstances, the whale size is set to 50, the prey's direction and distance are set to [-20, 20], and the step size control factor is set to 0.95. The number of iterations is also set to 300. The FOA-WELM algorithm requires 1.7 seconds to train, followed by the KHO-WELM algorithm at 1.5 seconds and the WOA-ELM algorithm at 1.3 seconds. This is due to the adaptive capabilities of the recently implemented optimisation method increasing the time complexity, which increases training time.

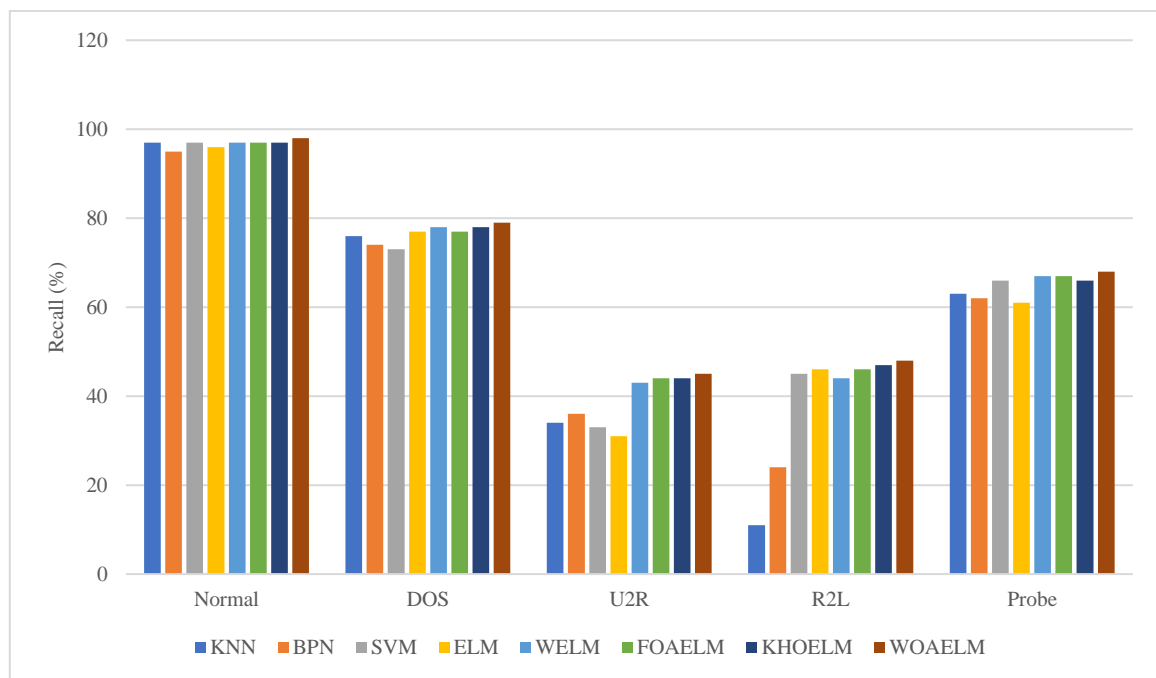


Figure. 6 Recall Percentage under Different Attacks

From the Figure 6, it is clear that when no attack is added WOAELM performed better than BPN and ELM by 3% and 2% respectively. When DOS attack is added WOAELM performed better than SVM and WELM by 8.21% and 1.2% respectively. When R2L attack is considered WOAELM performed better than FOAELM and KHOELM by 4% and 2% respectively. Finally, WOAELM is compared with other algorithms in probe attack. WOAELM performed better than KNN and ELM by 7.9% and 11.4%.

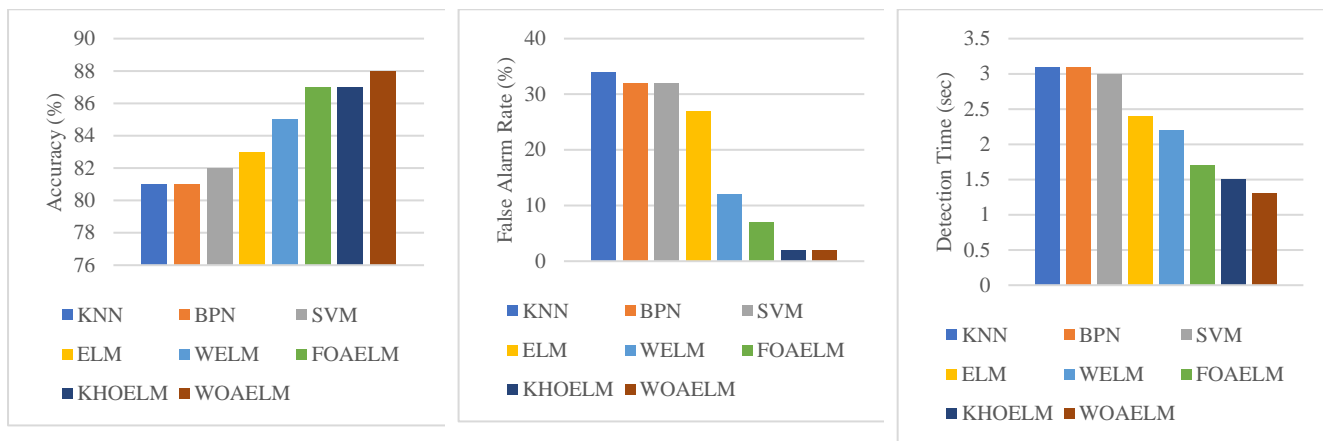


Figure. 7 Detection Time, Accuracy and False Alarm Rate under Different Attacks

It is obvious from the above figure that WOAELM outperformed other algorithms in parameters such as detection time accuracy and false alarm rate. WOAELM performed better than KNN in detection time accuracy and false alarm rate by 58%, 8% and 94% respectively, when SVM is taken for comparison WOAELM performed better in accuracy, false alarm rate and detection time by 56.5%, 6.8% and 93% and respectively. Next WELM is taken for comparison WOAELM performed better in detection time accuracy and false alarm rate by 86.3%, 3.4% and 83.33% respectively. Finally, KHOELM is taken for comparison WOAELM performed better in accuracy by 1.13%, false alarm rate is equal and detection time is better by 15%.

5. Conclusion

In order to boost WELM's performance, we applied the whale optimisation technique. It is evident from the comparison that the model put out in this research outperformed other conventional models in intrusion detection. The accuracy of intrusion detection is increased by reducing recall, false alarm rate, and detection time. Even if certain older approaches outperformed WOA based WELM in some assaults, the overall performance is still notable. By employing this method, the data imbalance is also lessened. Therefore, using WOA based WELM to find network intrusion is effective.

References

- [1] Zhu Zibin, Meng Anbo, Ou Zuhong, Wang Chenen, Zhang Zheng, Chen Mi, & Liang Ruduo. (2021). Ultra-short-term wind power forecasting based on a deep ensemble learning model based on multivariate mode decomposition and multi-objective algorithm optimization. *Modern Power*, 38, 1-12.
- [2] Wang Yanqi, Zhang Qiang, Zhu Liutao, & Yuan Heping. (2022). GBDT Regression Prediction Model Based on Improved Whale Optimization Algorithm. *Journal of Jilin University (Natural Science Edition)*, 60 (2), 401-408.

- [3] Li Shiyong, Zhang Weibin, Zhao Xinzhe, Liu Bin, & Zheng Yifei. (2021). Photovoltaic Maximum Power Point Tracking with Improved Whale Algorithm Optimizing Support Vector Regression. *Journal of Electrotechnical Society*, 36 (9), 1771-1781
- [4] Zhang Siqi, Ni Jing. Application of Hybrid Whale Algorithm in Flexible Job Shop System. *Journal of System Science*, 2020, 28(1): 131–136.
- [5] Guo Qicheng, Du Xiaoyu, Zhang Yanyu, & Zhou Yi. (2021). UAV 3D Path Planning Based on Improved Whale Algorithm. *Computer Science*, 48 (12), 304-311
- [6] Huang Q B, Li J X, Song C N, et al. Whale optimization algorithm based on cosine control factor and polynomial mutation [J]. *Control and Decision*, 2020, 35(3): 559-568
- [7] Du Liuqing, Li Renjie, & Li Baochuan. (2021). A Deep Learning Prediction Method for Thermal Errors of CNC Machine Tools Based on Attention Mechanism. *Engineering Science and Technology*, 53 (6), 194-203.
- [8] Guo Bin, Xing Jie, Yao Fei, & Jing Xiaomin. (2022). Optimal allocation of user-side hybrid energy storage based on bilevel programming model. *Energy Storage Science and Technology*, 11 (2), 615
- [9] Mei Wenjuan, Gao Yuan, Du Li, Liu Zhen, & Wang Houjun. (2022). Real-time flow prediction method of device degradation trend based on online correlation entropy extreme learning machine. *Journal of Instrumentation*, 40 (11), 212-224.
- [10] Chang Ningdong, Cheng Pengda, Feng Chun, Wang Lixiang, & Li Yuqiong. (2023). Research progress on optimization algorithms for equipment field traffic. *Optics and Precision Engineering*, 31 (5), 776-792
- [11] Shi Zhicheng. (2021). Emergency response to sudden toxic events: historical experience and lessons. Beijing Book Co. Inc
- [12] Han Xiao, & Li Pei. (2019). Analysis of B5G network architecture and key technologies. *Post and Telecommunications Design Technology*, 8, 30-33.
- [13] Zhang Yinghui, Hu Lingyun, Li Yixin, Ning Jianting, & Zheng Dong. (2022). A secure and efficient batch authentication scheme based on dynamic revocation mechanism in spatial information network. *Journal of Communications*, 43 (4), 164-176.
- [14] Zhang Yuxin, He Wenqin, Chen Hong, & An Peng. (2020). A Summary of Patent Analysis for Autonomous Vehicle Safety Technology. *Science in China: Information Science*, 50, 1732-1755.
- [15] Zhao Kaifeng. (2021). The road to digital transformation of petroleum geophysical prospecting: Towards the era of intelligent interpretation of real-time data collection and automatic processing. *Petroleum Geophysical Prospecting*, 60 (2), 175-189
- [16] Liang Jing, Liu Rui, Qu Boyang, & Yue Caitong. (2019). A review of the application of evolutionary algorithms in large-scale optimization problems. *Journal of Zhengzhou University: Engineering Edition*, 39 (3), 15-21.
- [17] Liu Jingsen, Ma Yixiang, & Li Yu. (2021). Improved Whale Algorithm for Engineering Design Optimization Problems. *Computer Integrated Manufacturing Systems*, 27 (7), 1884.
- [18] Liu Xiaolong. (2021). Multi-population vertical and horizontal two-way learning and whale optimization algorithm for information exchange. *Journal of Electronics and Information Technology*, 43 (11), 3247-3256.
- [19] Huang Fei. (2020). Application of an improved whale optimization algorithm in interval supply chain network (Master's thesis, Chengdu University of Information Science and Technology).