International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 16, No.1 (Oct-24)

http://dx.doi.org/10.12785/ijcds/1601115

Using a Grey Wolf Optimization and Multilayer Perceptron Algorithms for an Anomaly-Based Intrusion Detection System

Wathiq Laftah Al-Yaseen[1] **and Qusay Abdullah Abed**[1]

[1]*Kerbala Technical Institute, Al-Furat Al-Awsat Technical University, 56001, Kerbala, Iraq*

Received 14 Mar. 2024, Revised 29 Jun. 2024, Accepted 12 Jul. 2024, Published 1 Oct. 2024

Abstract: The swift development of information technology has led to an increase in the total number of electronic devices linked to the Internet. Additionally, there were more network attacks. Accordingly, it is crucial to create a defense system capable of identifying novel attack types. An intelligent system Intrusion detection system (IDS) is the most effective defense system, monitoring and analyzing network packets to spot any unusual activity. Moreover, there are a lot of useless and repetitive features in the network packets, that hurt the IDS system's performance and use up too many resources. The computation times will be shortened and computation complexity will be also simplified by choosing the suitable feature selection technique that helps to determine the most related subset of features. An enhanced anomaly IDS model based on a multi-objective grey wolf optimization technique has been proposed in this paper. Using the grey wolf optimization technique, the best features from the dataset were identified to achieve a considerable improvement in classification accuracy. However, a multilayer perceptron technique (MLP) was employed to assess the suitability of specific features that were properly for predicting attacks. Furthermore, to show the efficiency of the suggested approach using 20% of the NSL-KDD dataset, multiple attack scenarios were employed. The proposed approach achieves high detection rates (92.52%, 70.31%, 14.53%, and 2.87%) for DoS, Probe, R2L, and U2R categories, respectively, with classification accuracy reaching 85.43%. Our proposed model was evaluated against other current approaches and produced noteworthy results.

Keywords: Intrusion Detection, Grey Wolf Optimizer, Multilayer Perceptron, Feature Selection, Classification

1. Introduction

Due to the enormous improvements in the information technology and the widespread adoption of Internet apps., people are using the Internet more frequently. These days, using technology in daily life has become a necessity [1]. Additionally, several organizations and businesses use the network to convey crucial information, and this information needs to get to its destination undamaged [2][3]. Furthermore, surveillance and hacking methods have advanced and are now simple enough for even a layperson to use. To accurately monitor and check the massive volume of packets that transit across the network, it is necessary to create a security system [4].

Further, computer and network security's first line of defense is the use of current security methods like firewalls, authentication of clients, data encryption, and access controls; however, these techniques are unable to provide an ideal security scenario to completely protect the network [5]. In addition, to identify different types of new attacks and notify security personnel when action is required, many researchers are trying to develop security hardware and software that are capable of alerting users. Intrusion detection system (IDS) technologies are considered one of the most prominent forms of security systems that boost security in computer networks and prevent attacks [6]. Anderson first described the idea of IDS in a technical report in 1980. A defensive system IDS is in the position of detecting intrusions and suspicious activity. Monitoring and analyzing network traffic and client device activity is how this system is run. Furthermore, when malicious behavior is discovered on the network, the intrusion detection system generates a notification to alert the section of security and save the action to a log files that may have been utilized later for more analysis [7]. When an IDS is gathering a lot of packages and features which are obtained via network connections, they will be required to assess and determine in real time if everything is normal or abnormal. Several of these elements are unnecessary or repetitive, that has a substantial impact on the accuracy and responsiveness of IDS. By choosing the crucial features that improve performance, it is possible to eliminate this type of feature [8][9][10]. There are numerous methods to identify features

E-mail address: wathiq@atu.edu.iq, inkr.ks@atu.edu.iq

using AI and data mining techniques, however doing so does not always result in an increase in IDS performance. The effectiveness of the system's classification is impacted by poor feature selection. Additionally, it might lead to numerous false negative and positive results. Additionally, several of these techniques raise the cost of computing.

In the current research, the researcher proposed a feature- selection based Gery Wolf Optimization (GWO) combined with Multilayer Perceptron technique (MLP) for IDS to reduce the wrong alarm rate while improving detection accuracy. The performance of the proposed model is assessed using the NSL-KDD dataset. Prior to using classification algorithms, the researcher pre-processed the NSL-KDD dataset by converting and normalizing the data.

This article is structured as follows. A background is introduced in Section 2. Section 3 demonstrates the related works. In Section 4, the proposed methodology is presented in more detail. The experimental setup and results with discussion are shown in Section 5 and Section 6, respectively. In Section 7, the conclusion and future works are illustrated.

2. Background

A. Intrusion Detection System

To successfully identify the intrusion or not, the network's placement of IDS sensors is essential. Considering this, gathering data is crucial to the IDS detection process. Depending on where the IDS sensors are installed, this data can be gathered via network traffic or the client device and can be categorized into two categories: network-based IDS and host-based IDS. The host-based IDS runs on the client computer to analyze and inspects the system's local data, including log files, sign-in events, and commands, to find the intrusion. Additionally, it keeps track of how much RAM, CPU, and hard drive are being used by the device. Further, the IDS immediately notify the system administrator when any changes are made to the system or client files [11]. However, the network-based IDS observers and analyzes the network stream of traffic to discover the intrusion. The NIDS sensors are often placed throughout the network in various places. These sensors locate the intrusion by looking for any unusual activity in the network flow. As a result, it is incredibly challenging for the infringers to determine where they are in the network [12].

Depending on the method used for detection IDS can be divided into IDS based on anomalies and signature-based IDS. The recognition mechanism of based on the signature IDS strategy is based on a comparison between the actions of the client and predefined stored attack patterns. Additionally, the database includes descriptions of recognized attacks, including their signatures and characteristics [13]. In contrast, by using a matching algorithm the IDS analyzes incoming network traffic behavior and compares it to the database. If a match is discovered, the system will warn the security staff with an alarm [14]. This method is also capable of precisely identifying known attacks. However, to detect zero-day attacks, this model needs to be updated often. Contrariwise, anomaly-based IDS is establishing a profile for typical actions, this kind of inspects client or networking activity. It then compares system occurrences with the normal profile. The system will treat any occurrence that deviates from the expected profile as aberrant behavior, which will then cause a system alert. The wide distribution of Internet networks creates many difficulties in quickly identifying intrusions because the IDS requires monitoring and investigating the vast network packets. These packets include many attributes (features), such as the source and destination Internet Protocol (IP) addresses, and others, that are used to describe the packet's characteristics. Even though the analysis technique is extremely advanced, numerous repetitive and irrelevant aspects limit IDS's performance. As a result, the IDS must carefully manage each important piece of information to identify any anomalous activity [15]. The IDS's performance can be improved using a variety of methods. Feature selection is the approach that is most frequently utilized.

B. Feature Selection Techniques

The technique of selecting a subset of noteworthy features, or features, is known as feature selection. from a dataset and removing redundant and unnecessary information to create an effective learning strategy. Additionally, this method can reduce calculation complexity and time [16]. In general, a feature selection method involves several stages. In the first, a subset of features is extracted during the generating step from the original dataset. Next, the subset is assessed utilizing the objective function as the basis for evaluation (fitness function), which determines which subset of features is optimal. Thirdly, the effectiveness of the chosen features is evaluated using the stopping criterion. Lastly, the validation stage verifies if the chosen features satisfy the system need or not [17]. Moreover, three categories: wrapper, filter, and hybrid methods; can be used to group feature selection techniques. In this work, the feature selection techniques involve the usage of the wrapper approach. The wrapper technique selects the feature subset by evaluating machine learning algorithms. Additionally, the algorithms will produce and provide an efficient optimal subset of features that will yield metrics such as accuracy, detection rate, and so on, as well as seeking to minimize the initial set of characteristics. Though, the system's resources are exhausted and more processing time is required for these remarkable outcomes [18]. Figure 1 shows the process of wrapper feature selection steps.

3. Related Works

Recently, numerous researchers have employed machine learning approaches to deal with many issues of IDS. The feature selection technique solves most of these issues. In this section, the researcher focuses on GWO an algorithm for feature selection that is employed to enhance the IDS's performance.

A feature selection method based on the genetic approach, particle swarm optimization firefly optimization,

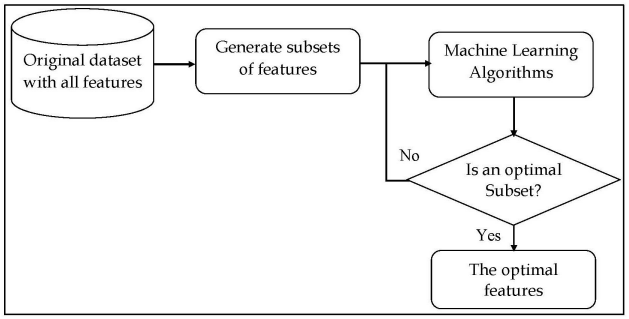

Figure 1. Wrapper feature selection method

and grey wolf optimization was introduced by researchers [19]. These methods were evaluated using the UNSW-NB15 dataset iteratively to identify which is the feature subset that would result in the best attack detection accuracy. A feature subset with thirty features was chosen after multiple attempts. In addition, the classifier procedure was carried out utilizing the SVM and J48 Tree-based models. The accuracy, false positive rate (FPR), and false negative rate (FNR) were the key performance metrics taken into account in this investigation. The UNSW-NB15 training subset was used for the experiments with the binary classification system. The findings showed that the suggested J48 model had a 14.95% FPR, a 90.17% FNR, and a 90.48% training accuracy. Additionally, the proposed SVM model obtained FPR of 15.39%, FNR of 3.13%, and training accuracy of 90.12%.

Grey wolf optimization (GWO) and particle swarm optimization (PSO) were suggested by Muhammad (2021) for IDS. Researchers developed the PSO-GWO-NB and PSO-GWO-ANN innovative FSs and IDS strategies. This work also assessed the PSO and GWO elements that were mostly repeated. the two classifiers ANN and NB were also used in evaluations. The outcomes of the test showed that MRF features produce good recalls and precisions. According to Mohammad's research, PSO-GWO-NB classifiers performed better in FSs and IDSs than PSO-GWO-ANN classifiers.

According to Kunhare, hybrid classification employing logistic regression (LR) and decision tree (DT) has also been carried out. Based on the selection of a relevant feature from the NSL-KDD dataset by a genetic algorithm (GA), the detection rate (DR) and accuracy (ACC) are improved. A range of meta-heuristic algorithms, including the Bat algorithm (BAT), Multiverse Optimization (MVO), Particle Swarm Optimization (PSO), and Grey Wolf Optimization (GWO), were used to improve the selections. The results revealed that the GWO approach, with twenty carefully chosen features, has offered a DR of 99.36% and the greatest accuracy of 99.44%. Though, GA takes longer to converge due to its stochastic character [20].

PSO and the grey wolf optimization (GWO) model were combined to create a new anomaly-based IDS model in [21]. To improve this model's identification abilities, the collected dataset was initially mined for strongly associated traits using the GA-based technique. Then, a hybrid PSO-GWO algorithm produced a BPNN. Finally, the original dataset was subjected to this combination methodology to address binary and multi-class classification difficulties. This paradigm was, nevertheless, susceptible to hidden or trivial issues.

Additionally, [22] introduced a multi-objective GWO to address in IDSs the FS problems. With reasonable important weights, the authors employed the fitness evaluation function's accuracy and decrease rate. In this instance, based on the population initialization stage of a heuristic search, the authors employed the random subset generation technique. The proposed work was assessed using a multi-class classification approach and the SVM and NSL-KDD dataset. The number of selected traits decreased more quickly than expected, according to the data. Instead of the DoS assault, the suggested model produced the highest classification accuracy across all categories.

In [23], to choose the best feature subset for intrusion categorization, an intelligent GWO technique was used. The researcher applied the Knowledge Discovery and Data Mining Cup 1999 (KDDcup99) dataset before applying the intelligent feature selection method to an informational index. As a result, this excellent FS decision leads to high grouping precision and minimizes computational cost. The application's outcomes include the diversity, accuracy, and detection rate of intrusion detection using different learning methods.

Moreover, the authers' area of expertise was cloud datacenter network anomaly detection. For anomaly detection, the researchers used CNN and the Gray Wolf Optimization (GWO) method. According to the authors, this technique, a significant amount of network log data may be analyzed in real time for anomaly ID. The efficiency of the approach is calculated using synthetic datasets, DARPA'98 and KDD'99, which demonstrates its superiority to previously published approaches. The accuracy of the approach described in this work was 97.92% on the DARPA'98 dataset and 98.42% on the KDD'99 dataset. The fascinating argument raised by the authors of this research is that current anomaly detection algorithms are ineffective for actual time anomaly identification in big data since they increase computing complexity and result in a significant number of false positives [24].

Additionally, it was advised to use a hybrid GWO strategy for the classification stage, combining the CS algorithm as a feature selection model with support vector machine (SVM). The authors employed in their pre-processing procedure the min-max method. Using DoS, Probe, U2R, and R2L attacks, the number of selected features was reduced to 18, 17, 34, and 8 in the experimental findings obtained

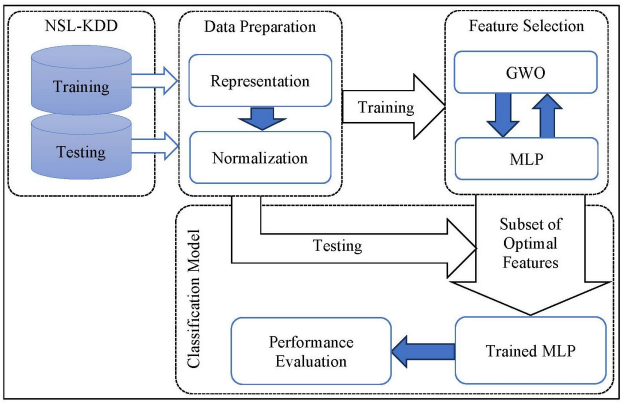

Figure 2. The proposed method's framework

using the suggested approach. The fitness function that was employed to identify the best-fitting subset of features was maximum mutual data [25].

4. Proposed Methodology

The researcher gives a general overview of the suggested approach's framework in this section. Figure 2 represents the total framework of the recommended method. The procedure of the proposed approach is as follows:

1) Obtain the datasets from traffic of network.
2) A preparation of datasets is achieved by represent the symbolic features as numeric, then a normalize method is applied to set the values of features between [0,1].
3) A feature selection is applied based on GWO technique to select the best features from datasets with employed MLP technique to evaluate the selected features.
4) A classifier model is built by training the MLP classifier with the best features which be selected from the previous step. Then, the trained model is used to classify the testing dataset.
5) An evaluation of the performance of proposed method is achieved by using the IDS metrics such as accuracy, detection rates, etc.

The following subsections provide a more details of the phases in the suggested model.

A. Data Preparation

One of the most important problems with machine learning and data analysis algorithms is data preparation. The goal of data preparation is to create and transform data in the right format, particularly when the data includes a variety of informational components and formats. The NSL-KDD dataset contains a wide range of features and data that are presented in several ways, such as by alphabet, number, symbol, etc. These attributes' investigation can require extra processing time and hardware resources. By employing the representation technique, symbolic features were transformed into numeric features to prevent such issues [26]. The protocol type characteristic with the three values, for instance (tcp, icmp, and udp) in NSL-KDD dataset can be represented in the proposed method to numeric features (0, 1, and 2), respectively. The other symbolic features will be re-represented as the protocol feature. Table I presents the symbolic features of NSL-KDD dataset with representation in the selected method.

To provide a range of feature values that is proportionate, a process for calibrating feature values is used. In the feature record, each value is scaled in this work via Equation (1).

$$X_{new} = \frac{X_{old} - X_{min}}{X_{max} - X_{min}} \tag{1}$$

Where X_{old} is the record's current value before normalized, X_{max}, X_{min} denote to the maximal and the minimal values in the feature of X_{old} value, consecutively. X_{new} is the normalized value. Lastly, the range of record values that falls between values of one and zero.

B. Grey Wolf Optimization Technique based Feature Selection

To select the optimal collection of features for this work, the GWO technique was modified. Prior to that, the random subset generation technique was used to create a subset of features [27]. Mirjalili indicated a swarm-based algorithm known as the GWO was applied. Grey wolves' natural social behaviors serve as the inspiration for GWO. The hunting and pursuing methods used by grey wolves to catch their prey are an example of the search process that results in the best outcome. Grey wolves in the wild like to live in packs, which often consist of five to twelve wolves [28]. The wolves of these packs can be divided into four groups depending on the position of wolves in the pack which leads to improve the hunting and chasing process [29]. The first group which name Alpha (α) comprises both male and female wolves who serve as leaders to make decisions regarding hunting, waking, sleeping, and other related matters. However, A second pack of wolves known as Beta (β) oversees helping the other wolves in the packs make decisions. These wolves can be either male or female. The third group, Delta (δ), fulfills several significant responsibilities including caregiver, sentinel, pack elder, and hunter. The last group in the hierarchical paradigm is Omega (ω), the weakest wolf, and by following other wolves' orders, it acts as a scapegoat [30].

When hunting, the grey wolves start to circle around their prey. Equation (2) provides the mathematical expression for this step:

$$\vec{X}(t+1) = \vec{X}(t) - \vec{A}.\left|\vec{C}.\vec{X}_p(t) - \vec{X}(t)\right| \tag{2}$$

where the vector $\vec{X}_p$ represents the prey's position, the

TABLE I. REPRESENT SYMBOLIC FEATURES

Protocol_type		**Service**						**Flag**	
Values	***Represent***	***Values***	***Represent***	***Values***	***Represent***	***Values***	***Represent***	***Values***	***Represent***
tcp	0	private	0	netbios_ns	22	shell	43	REJ	0
icmp	1	ftp_data	1	link	23	hostnames	44	SF	1
udp	2	eco_i	2	Z39_50	24	echo	45	RSTO	2
		telnet	3	sunrpc	25	daytime	46	S0	3
		http	4	auth	26	pm_dump	47	RSTR	4
		smtp	5	netbios_dgm	27	IRC	48	SH	5
		ftp	6	uucp_path	28	netstat	49	S3	6
		ldap	7	vmnet	29	ctf	50	S2	7
		pop_3	8	domain	30	nntp	51	S1	8
		courier	9	name	31	netbios_ssn	52	RSTOS0	9
		discard	10	pop_2	32	tim_i	53	OTH	10
		ecr_i	11	http_443	33	supdup	54		
		imap4	12	urp_i	34	bgp	55		
		domain_u	13	login	35	nnsp	56		
		mtp	14	gopher	36	rje	57		
		systat	15	exec	37	printer	58		
		iso_tsap	16	time	38	efs	59		
		other	17	remote_job	39	X11	60		
		csnet_ns	18	ssh	40	ntp_u	61		
		finger	19	kshell	41	klogin	62		
		uucp	20	sql_net	42	tftp_u	63		
		whois	21						

vector of the grey wolf's position is $\vec{X}$, the current iteration is denoted by t, the coefficient matrix vectors $\vec{A}$ and $\vec{C}$ are defined as:

$$\vec{A} = 2\vec{\alpha}.\vec{r_1} - \vec{\alpha}$$

$$\vec{C} = 2.\vec{r_2}$$

where $\vec{r}_1$, $\vec{r}_2$ are random vectors and α decreases over iterations from 2 to 0.

Whereas the beta and the delta wolves are deeply knowledgeable about the prey potential location, alpha wolves presume to have the best solution. Consequently, to determine the locations of the remaining wolves, including the omega wolf, the placements of the alpha, beta and delta wolves will be used as indicated by the following Equation (3):

$$\vec{X}(t+1) = \frac{1}{3}\vec{X}_1 + \frac{1}{3}\vec{X}_2 + \frac{1}{3}\vec{X}_3 \tag{3}$$

where $\vec{X}_1$, $\vec{X}_2$, and $\vec{X}_3$ are given by the following:

$$\vec{X}_1 = \vec{X}_\alpha(t) - \vec{A}_1.\left|\vec{C}_1.\vec{X}_\alpha - \vec{X}\right|$$

$$\vec{X}_2 = \vec{X}_\beta(t) - \vec{A}_2.\left|\vec{C}_2.\vec{X}_\beta - \vec{X}\right|$$

$$\vec{X}_3 = \vec{X}_\delta(t) - \vec{A}_3.\left|\vec{C}_3.\vec{X}_\delta - \vec{X}\right|$$

Where the alpha, beta, and delta wolves' locations in each iteration, or the first three optimal solutions to the issue, are shown by the symbols $\vec{X}_\alpha$, $\vec{X}_\beta$, and $\vec{X}_\delta$ indicate in essence, the prey's position is indicated by alpha, beta, and delta wolves, while the remaining wolves roam randomly around it.

To conclude, the wolves attack their victim when it stops moving. Over the duration of the iteration, the number decreases from 2 to 0, indicating that the wolves are getting closer to the prey. The $\vec{\alpha}$ values can be computed by the following Equation (4).

$$\vec{\alpha} = 2 - \frac{2 \times t}{MaxIter} \tag{4}$$

Where *MaxIter* is the problem's maximum iteration and t is the current iteration. As mentioned above, this proposed used GWO technique due to the behavior of meta-heuristic as well as the capability to identify the ideal outcome that is possible with avoid the local minima problem. In addition, it is simple to construct and has very few tuning parameters.

Furthermore, after one algorithm iteration, the alpha, beta, and delta wolves ($\vec{X}_\alpha$, $\vec{X}_\beta$, and $\vec{X}_\delta$) in the pack will become interested in the top three spots. Alpha's location is the best answer (position) in terms of classification accuracy, followed by beta and delta. Moreover, with every iteration, the positions of the grey wolves are convergent toward the prey. The optimal solution is the wolf in the alpha position, which is closest to the prey. The classifier is trained and validated in each algorithm iteration, after which the classifier's accuracy is calculated for each position matrix subset (or solution) [31]. The GWO algorithm is shown as the following.

GWO Algorithm

Input: *Max_Iteration, Population_Size*
Output: *The best solution* X_α
1: *Initialize the population of grey wolves* X_i $(i = 1, 2, \ldots, Population\ Size)$
2: *Prepare* α, A, C
3: *Compute the position of each grey wolf by using MLP (as fitness)*
4: *Find the best three solutions* $X_\alpha, X_\beta, X_\delta$
5: ***While*** *(t < Max_Iteration)*
6: *Compute* X_1, X_2, X_3
7: *Update* α, A, C
8: *Compute the position of all grey wolves by using MLP (as fitness)*
9: *Update* $X_\alpha, X_\beta, X_\delta$
10: $t = t + 1$
11: ***end while***
12: ***return*** X_α

C. Multilayer Perceptron Technique based Evaluation of Selected Features

Multilayer Perceptron (MLP) was employed in this phase to assess the chosen features from the previous phase. It is known as a feedforward neural network. Its construction consisted of three levels: one or more hidden layers, the output layer, and the input layer [32]. A MLP mechanism consists of two sorts of data flows: data forward propagation and error backpropagation [33]. The relationship in forward propagation of an MLP with a single hidden layer between the input $x = [x_1, x_2, \ldots, x_n]$ and the output $y = [y_1, y_2, \ldots, y_k]$ may be expressed as Equation (5):

$$y_k = \sum_{j=1}^{m} \left[f\left(\sum_{i=1}^{n} \omega_{ij} x_i + b_j \right) \omega_{jk} + b_k \right] \tag{5}$$

Where b_k symbolizes the bias from the hidden layer to the output layer, b_j symbolizes the bias from the input layer to the hidden layer, and $f()$ is the hidden layer's activation function. Additionally, ω_{ij} and ω_{jk} are the weights that connect the input layer and hidden layer, respectively. The discrepancy between the observation data and the MLP output is known as the prediction error E of an MLP. The prediction error of backward propagation algorithm enables the parameters (bias and weight) in an MLP to be repeatedly trained, or refined, when the MLP is first started to model a complex process [34]. In details, each neuronś weight and bias are modified in the manner described below:

The weight ω'_{ij} that the hidden layer has updated from the input layer as

$$\omega'_{ij} = \omega_{ij} - \mu \frac{\partial E}{\partial \omega_{ij}}$$

The weight ω'_{jk} that has been updated from the hidden layer to the output layer as

$$\omega'_{jk} = \omega_{jk} - \mu \frac{\partial E}{\partial \omega_{jk}}$$

The bias b'_j that has been updated from the input layer to the hidden layer as

$$b'_j = b_j - \mu \frac{\partial E}{\partial b_j}$$

The bias b'_k that has been updated from the hidden layer to the output layer as

$$b'_k = b_k - \mu \frac{\partial E}{\partial b_k}$$

where μ is the learning rate.

The researcher noted from GWO algorithm that MLP will be used to evaluate every wolf in the population (i.e., computed the fitness of wolf as accuracy of MLP). As a result, the dataset was divided into two sets: X_{train}, which stands for the features in the training set, and Y_{train}, which in the training set, stands for the class label. While the features in the testing set comprise X_{test}, the testing set's class label features are contained in Y_{test}. X_{train} and Y_{train} will be used to train the MLP classifier, and X_{test} will subsequently be fed into the model. Subsequently, the model's output will be cross-checked against the Y_{test} results; a matching show that the classifier perfectly expected the behavior of the dataset record.

5. Experimental Setup

The datasets, performance metrics, experiment, and findings are presented in this section. The proposed approach is conducted using python language. Additionally, the effectiveness of the suggested model was evaluated using 20% of the NSL-KDD dataset. The experiments were carried out on a 2.80 GHz Core i7 CPU running Windows 11 with 8

GB of RAM.

A. Benchmark Dataset

Researchers had been using the KDD'99 set of data extensively in recent years to assess intrusion detection systems; however, the collection had some issues, such as duplicate and unnecessary entries, which had a major negative impact on the system's performance. Therefore, Tavallaee suggested a unprecedented set of data called NSL-KDD, which was chosen from the original KDDCup'99 data but with its issues resolved. The NSL-KDD dataset was the most reliable in the field and works well for comparison and assessment [35]. Generally speaking, it contains two datasets for training and testing (125973 and 22544 samples, respectively), each including 41 features. The distribution of data samples in each training and testing dataset is shown in the third Figure 3.

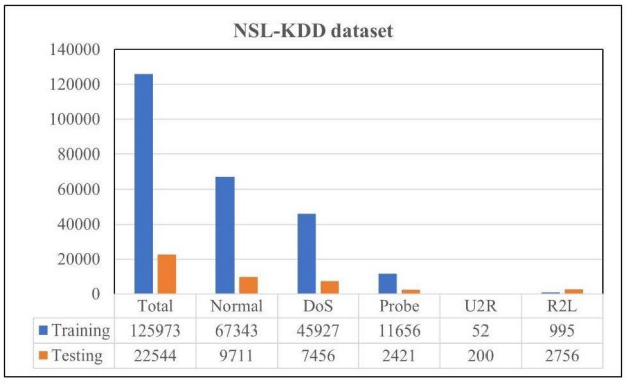

Figure 3. Sample distribution from the NSL-KDD dataset

All samples are categorized as either normal or attack; the training dataset was corresponding of 22 attacks while the testing dataset has 39 attacks, and they fall into one of four categories:

- DoS: To keep resources from being available to the user, more queries were made to the system.
- Probe: To find out more information about the target host, use network scanning.
- User to Root (U2R): Tries to guess the password to gain unauthorized entry into the account that controls and change system data.
- Remote to Local (R2L): Entry to the system as an authorized user.

B. Performance Metrics

The researcher used the commonly used evaluation metrics such as accuracy, F-Score, detection rate, precision and false alarm rate to gauge how effective the suggested methodology was.

Equation (6) was used to calculate the accuracy, which is the ratio of correct predictions (attack and normal) to the total size of the data set.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

Equation (7) can be used to calculate the detection rate that represents the ratio of accurately estimated attack cases to the attack class's actual size.

$$Detection\ rate = DR = Recall = \frac{TP}{TP + FN} \tag{7}$$

Equation (8) can be used to calculate the precision that represents the ratio of accurately anticipated attack cases to the estimated attack class size.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

Equation (9), which was used to determine the false alarm rate, takes into account the ratio of correctly predicted normal instances that are classified to the whole number of normal cases as attack cases.

$$False\ Alarm\ rate = FAR = \frac{FP}{TN + FP} \tag{9}$$

The F-Score evaluates the ratio of detection rate to precision which will be calculated by using equation (10):

$$F - Score = \frac{2 \times Recall \times Precsion}{Recall + Precision} \tag{10}$$

Where TP is the classifier accurately predicts the class as intrusion, which is the actual class of the dataset, TN is the dataset's actual class is normal, as predicted accurately by the classifier, FP is the dataset's actual class is normal, despite the classifier's prediction that it is an incursion, and FN is the actual class of the dataset is an intrusion, despite the classifier's prediction that it will be normal.

6. Results and Discussions

The main findings of the suggested technique were covered in this section. The accuracy of the suggested binary classification-based technique is achieved. Compared to hybrid Cuckoo search and the GWO algorithm by Xu et al., [36], which obtained 83.57% accuracy with only 6 features, the proposed technique can achieve a better performance, reaching 88.67% accuracy with only 5 features. However, using 10 features in multiclass, the proposed technique achieved 85.43% accuracy, whilst a modified GWO and extreme learning machine (ELM) by [37] with 17 features was introduced 81% accuracy in multiclass. Table

TABLE II. PERFORMANCE AFTER APPLIED PROPOSED FEATURE SELECTION METHOD

Measure	41 features (Multiclass)	Feature Selection	
		10 features (Multiclass)	*5 features (Binary)*
Accuracy	77.13%	85.43%	88.67%
FAR	0.54	0.12	0.04

II presents the enhanced performance after applying the proposed feature selection approach.

To provide further evidence that the proposed approach performs well, the researcher compared it with MLP 41 features in categories level as it is shown in Table III. Moreover, it is logical that the false alarms with binary classification are less than with multi-class classification because the proposed classification model, like any model, requires (e.g. a simple linear function) to separate between the classes, while with multiclass it requires a more complex function to separate between the classes (such as polynomial and RBF kernel functions, etc.), thus the number of false alarms are increases. Figure 4, Figure 5, and Figure 6 show the performance of the proposed model which be used 10 features in comparison with pure MLP which be used full the features of dataset in terms of detection rate, precision, and f-score, respectively.

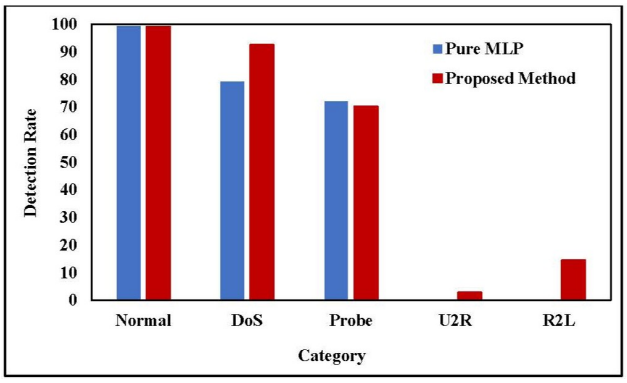

Figure 4. Comparison proposed method with pure MLP based on detection rates

Table III and Table IV show the suggested method that improves performance in terms of detection rates and accuracy. The frequencies of false alarms are reasonable. nevertheless, coming down to 0.12% using NSL-KDD. Consequently, the results demonstrate that the suggested technique greatly enhances intrusion detection systems' performance. Furthermore, it is more reliable than state-of-the-art methods as it's shown in Table IV. The used method is superior to the closest, which is 84.29% for NSL-KDD. The reason for the superiority of the proposed method over the rest of the methods is that it has the ability to detect attacks within a U2R, R2L, and Probe

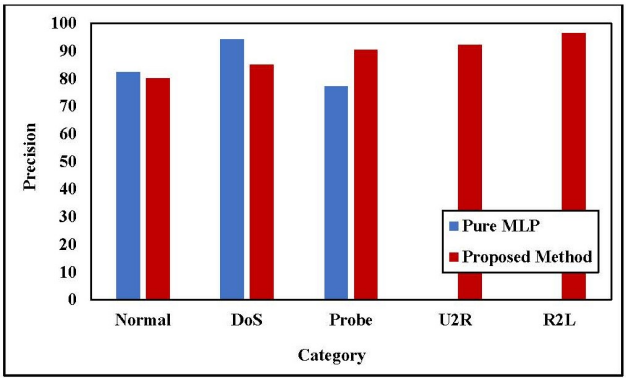

Figure 5. Comparison proposed method with pure MLP based on precision

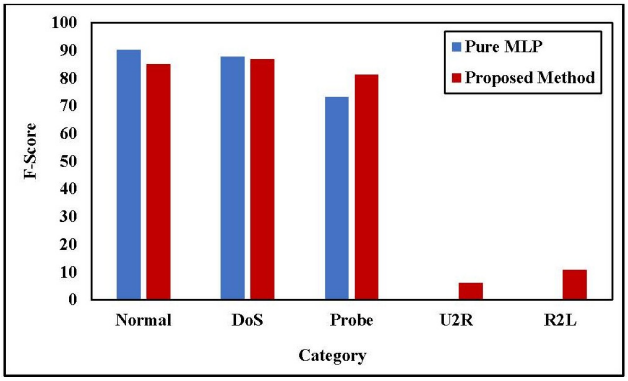

Figure 6. Comparison proposed method with pure MLP based on F-Score

in the good proportions, which makes the results balanced with the rest of the categories like DoS and Normal, which reflected positively on the system's performance in general. This is evident when the researcher compares the overall accuracy of the model that have been suggested, which is 85.43%. Moreover, the proposed method outperforms others with a false alert rate of 0.12%. Accordingly, this study is outperforming others with the remaining results. Figure 7 shows the performance of proposed method comparing with the state-of-the-arts methods in terms of accuracy, false alarm rate, precision, and f-score. However, in terms of computational complexity and runtime performance of the proposed method, our method can be achieved testing time reached to 1.266 ms while the pure MLP without feature selection reach to 3.232 ms.

A statistical test (t-test) is used to show how the findings of the suggested technique differ significantly from the earlier ones. The GWO+MLP considerably increased accuracy, as demonstrated by a t-test (one-tail) with a p-value of 0.0041545. This suggests that GWO+MLP can significantly improve intrusion detection system performance.

TABLE III. COMPARE THE PROPOSED METHOD'S PERFORMANCE (10 FEATURES) WITH MLP (41 FEATURES) IN MULTICLASS

Measure	MLP 41 features					Proposed Method 10 features				
	Normal	*DoS*	*Probe*	*U2R*	*R2L*	*Normal*	*DoS*	*Probe*	*U2R*	*R2L*
DR	99.32	79.33	72.14	0	0	99.12	92.52	70.31	2.87	14.53
Precision	82.32	94.23	77.23	0	0	80.22	84.98	90.34	92.16	96.53
F-Score	90.15	87.82	73.12	0	0	85.16	86.91	81.23	6.11	10.76

TABLE IV. A COMPARISON OF THE SUGGESTED METHOD WITH THE STATE-OF-THE-ART

Method	Accuracy	FAR	Precision	F-Score
Proposed Method	85.43	0.12	94.43	86.22
multi-level Hybrid kNN+ELM [38]	84.29	6.3	94.18	84.83
ResNet 50 [39]	79.14	N/A	91.97	79.12
GoogLeNet [39]	77.04	N/A	91.66	76.5
CNN-BiLSTM [40]	83.58	N/A	85.82	85.14
MLP+IGRF-RFE [41]	84.24	4.03	83.6	82.85

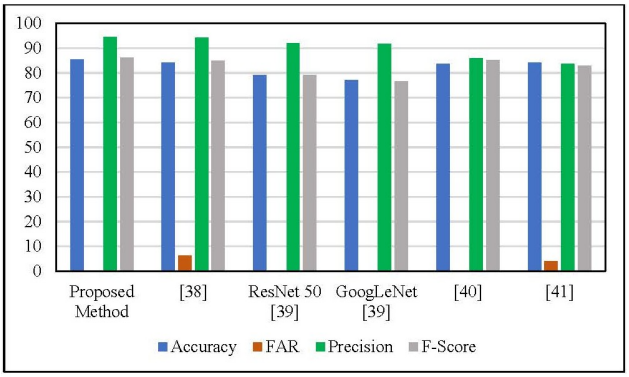

Figure 7. Comparison the performance of proposed method with the state-of-the-arts methods

7. Conclusions and Recommendations

This paper proposes a unique IDS model that makes use of GWO+MLP approaches. The GWO+MLP's performance evaluation reveals that for the NSL-KDD dataset, accuracy and false alarm rate can achieve 85.43% and 0.12%, respectively for multiclass classification, while they reach 88.67% and 0.04% for binary classification. The entire training and testing datasets are used for the studies, employing NSL-KDD datasets, by using KDDTrain+ for training and KDDTest+ for examination. In comparison to the previous researches, the proposed GWO+MLP performs better and yields findings that are balanced across all categories. The performance will be improved in the future by utilizing actual data frameworks to expand the proposed algorithm to various datasets and implementing deep learning-based feature selection techniques. Additionally, in the future works we will be to build hybrid model based on optimization techniques with GWO to solve some limitations of GWO such as local optima, slow rate of convergence, and weak exploration. Moreover, we will be to used other datasets with the modification of our proposed such as CICIoMT2024, CICIoT2023, and UNSW-NB15.

Conflicts of Interest

The authors have no conflicts of interest to disclose.

References

[1] A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab, "Machine learning for anomaly detection: A systematic review," *Ieee Access*, vol. 9, pp. 78 658–78 700, 2021.

[2] A. H. B. AIghuraibawi, S. Manickam, R. Abdullah, Z. A. A. Alyasseri, H. M. Jasim, and N. S. Sani, "Modified flower pollination algorithm for icmpv6-based ddos attacks anomaly detection," *Procedia Computer Science*, vol. 220, pp. 776–781, 2023.

[3] T. A. Alamiedy, M. F. Anbar, B. Belaton, A. H. Kabla, and B. H. Khudayer, "Ensemble feature selection approach for detecting denial of service attacks in rpl networks," in *Advances in Cyber Security: Third International Conference, ACeS 2021, Penang, Malaysia, August 24–25, 2021, Revised Selected Papers 3*. Springer, 2021, pp. 340–360.

[4] A. Aldweesh, A. Derhab, and A. Z. Emam, "Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues," *Knowledge-Based Systems*, vol. 189, p. 105124, 2020.

[5] S. Jayabharathi and V. Ilango, "Anomaly detection using machine learning techniques: A systematic review," in *International Conference on Advances in Data-driven Computing and Intelligent Systems*. Springer, 2022, pp. 553–572.

[6] T. A. Al-Amiedy, M. Anbar, B. Belaton, A. H. H. Kabla, I. H. Hasbullah, and Z. R. Alashhab, "A systematic literature review on machine and deep learning approaches for detecting attacks in rpl-based 6lowpan of internet of things," *Sensors*, vol. 22, no. 9, p. 3400, 2022.

[7] G. F. Scaranti, L. F. Carvalho, S. Barbon, and M. L. Proença, "Artificial immune systems and fuzzy logic to detect flooding attacks in software-defined networks," *IEEE Access*, vol. 8, pp. 100 172–100 184, 2020.

[8] N. Acharya and S. Singh, "An iwd-based feature selection method for intrusion detection system," *Soft Computing*, vol. 22, pp. 4407–4416, 2018.

[9] R. Zuech and T. M. Khoshgoftaar, "A survey on feature selection for intrusion detection," in *Proceedings of the 21st issat international conference on reliability and quality in design*, 2015, pp. 150–155.

[10] A. Nazir and R. A. Khan, "A novel combinatorial optimization based feature selection method for network intrusion detection," *Computers & Security*, vol. 102, p. 102164, 2021.

[11] J. M. Kizza, "System intrusion detection and prevention," in *Guide to computer network security*. Springer, 2024, pp. 295–323.

[12] 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.

[13] Z. Chkirbene, S. Eltanbouly, M. Bashendy, N. AlNaimi, and A. Erbad, "Hybrid machine learning for network anomaly intrusion detection," in *2020 IEEE international conference on informatics, IoT, and enabling technologies (ICIoT)*. IEEE, 2020, pp. 163–170.

[14] S. Kumar and R. Joshi, "Design and implementation of ids using snort, entropy and alert ranking system," in *2011 International Conference on Signal Processing, Communication, Computing and Networking Technologies*. IEEE, 2011, pp. 264–268.

[15] R. Zhang, F. Nie, X. Li, and X. Wei, "Feature selection with multi-view data: A survey," *Information Fusion*, vol. 50, pp. 158–167, 2019.

[16] Y. Zhu, W. Li, and T. Li, "A hybrid artificial immune optimization for high-dimensional feature selection," *Knowledge-Based Systems*, vol. 260, p. 110111, 2023.

[17] H. Alazzam, A. Sharieh, and K. E. Sabri, "A feature selection algorithm for intrusion detection system based on pigeon inspired optimizer," *Expert systems with applications*, vol. 148, p. 113249, 2020.

[18] M. Kadhum, S. Manaseer, A. Dalhoum *et al.*, "Evaluation feature selection technique on classification by using evolutionary elm wrapper method with features priorities [j]," *Journal of Advances in Information Technology Vol*, vol. 12, no. 1, 2021.

[19] O. Almomani, "A feature selection model for network intrusion detection system based on pso, gwo, ffa and ga algorithms," *Symmetry*, vol. 12, no. 6, p. 1046, 2020.

[20] N. Kunhare, R. Tiwari, and J. Dhar, "Intrusion detection system using hybrid classifiers with meta-heuristic algorithms for the optimization and feature selection by genetic algorithm," *Computers and Electrical Engineering*, vol. 103, p. 108383, 2022.

[21] S. Sheikhi and P. Kostakos, "A novel anomaly-based intrusion detection model using psogwo-optimized bp neural network and ga-based feature selection," *Sensors*, vol. 22, no. 23, p. 9318, 2022.

[22] T. A. Alamiedy, M. Anbar, Z. N. Alqattan, and Q. M. Alzubi, "Anomaly-based intrusion detection system using multi-objective grey wolf optimisation algorithm," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 9, pp. 3735–3756, 2020.

[23] D. Srivastava, R. Singh, and V. Singh, "An intelligent gray wolf optimizer: a nature inspired technique in intrusion detection system (ids)," *Journal of Advancements in Robotics*, vol. 6, no. 1, pp. 18–24, 2019.

[24] S. Garg, K. Kaur, N. Kumar, G. Kaddoum, A. Y. Zomaya, and R. Ranjan, "A hybrid deep learning-based model for anomaly detection in cloud datacenter networks," *IEEE Transactions on Network and Service Management*, vol. 16, no. 3, pp. 924–935, 2019.

[25] E. Roopa Devi and R. Suganthe, "Enhanced transductive support vector machine classification with grey wolf optimizer cuckoo search optimization for intrusion detection system," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 4, p. e4999, 2020.

[26] W. L. Al-Yaseen, Z. A. Othman, and M. Z. A. Nazri, "Multi-level hybrid support vector machine and extreme learning machine based on modified k-means for intrusion detection system," *Expert Systems with Applications*, vol. 67, pp. 296–303, 2017.

[27] D. S. Kim, H.-N. Nguyen, S.-Y. Ohn, and J. S. Park, "Fusions of ga and svm for anomaly detection in intrusion detection system," in *Advances in Neural Networks–ISNN 2005: Second International Symposium on Neural Networks, Chongqing, China, May 30-June 1, 2005, Proceedings, Part III 2*. Springer, 2005, pp. 415–420.

[28] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in engineering software*, vol. 69, pp. 46–61, 2014.

[29] Q. Al-Tashi, H. Md Rais, S. J. Abdulkadir, S. Mirjalili, and H. Alhussian, "A review of grey wolf optimizer-based feature selection methods for classification," *Evolutionary machine learning techniques: algorithms and applications*, pp. 273–286, 2020.

[30] H. Faris, I. Aljarah, M. A. Al-Betar, and S. Mirjalili, "Grey wolf optimizer: a review of recent variants and applications," *Neural computing and applications*, vol. 30, pp. 413–435, 2018.

[31] J. Tang, G. Liu, and Q. Pan, "A review on representative swarm intelligence algorithms for solving optimization problems: Applications and trends," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 10, pp. 1627–1643, 2021.

[32] X. Huang, Y. You, X. Zeng, Q. Liu, H. Dong, M. Qian, S. Xiao, L. Yu, and X. Hu, "Back propagation artificial neural network (bp-ann) for prediction of the quality of gamma-irradiated smoked bacon," *Food Chemistry*, vol. 437, p. 137806, 2024.

[33] Z. Zhang and J. Li, "Big data mining for climate change," 2019.

[34] A. H. Fath, F. Madanifar, and M. Abbasi, "Implementation of multilayer perceptron (mlp) and radial basis function (rbf) neural networks to predict solution gas-oil ratio of crude oil systems," *Petroleum*, vol. 6, no. 1, pp. 80–91, 2020.

[35] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the kdd cup 99 data set," in *2009 IEEE symposium on computational intelligence for security and defense applications*. Ieee, 2009, pp. 1–6.

[36] H. Xu, X. Liu, and J. Su, "An improved grey wolf optimizer algorithm integrated with cuckoo search," in *2017 9th IEEE in-*

ternational conference on intelligent data acquisition and advanced computing systems: technology and applications (IDAACS), vol. 1. IEEE, 2017, pp. 490–493.

[37] A. Alzaqebah, I. Aljarah, O. Al-Kadi, and R. Damaševičius, "A modified grey wolf optimization algorithm for an intrusion detection system," *Mathematics*, vol. 10, no. 6, p. 999, 2022.

[38] M. Latah and L. Toker, "An efficient flow-based multi-level hybrid intrusion detection system for software-defined networks," *CCF Transactions on Networking*, vol. 3, no. 3, pp. 261–271, 2020.

[39] Z. Li, Z. Qin, K. Huang, X. Yang, and S. Ye, "Intrusion detection using convolutional neural networks for representation learning," in *International conference on neural information processing*. Springer, 2017, pp. 858–866.

[40] K. Jiang, W. Wang, A. Wang, and H. Wu, "Network intrusion detection combined hybrid sampling with deep hierarchical network," *IEEE access*, vol. 8, pp. 32 464–32 476, 2020.

[41] 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, no. 1, p. 15, 2023.

Wathiq Laftah Al-Yaseen assistance professor at Al-Furat Al-Awsat Technical University, Iraq. My PhD in computer science from UKM, Malaysia, mater of computer science from Babylon University, Iraq. My interest in machine learning, deep learning, multiagent systems, data mining, and network security.

Qusay Abdullah Abed assistance professor at Al-Furat Al-Awsat Technical University, Iraq. My Master in computer science from UUM, Malaysia. My interest in Information Technology, Data base, machine learning, Internet of Things, Communication Technology and network security.