Journal of Ambient Intelligence and Humanized Computing (2020) 11:3735–3756
https://doi.org/10.1007/s12652-019-01569-8

ORIGINAL RESEARCH

Anomaly-based intrusion detection system using multi-objective grey wolf optimisation algorithm

Taief Alaa Alamiedy[1] · Mohammed Anbar[1] · Zakaria N. M. Alqattan[1] · Qusay M. Alzubi[1]

Received: 21 May 2019 / Accepted: 29 October 2019 / Published online: 9 November 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

The rapid development of information technology leads to increasing the number of devices connected to the Internet. Besides, the amount of network attacks also increased. Accordingly, there is an urgent demand to design a defence system proficient in discovering new kinds of attacks. One of the most effective protection systems is intrusion detection system (IDS). The IDS is an intelligent system that monitors and inspects the network packets to identify the abnormal behavior. In addition, the network packets comprise many attributes and there are many attributes that are irrelevant and repetitive which degrade the performance of the IDS system and overwhelm the system resources. A feature selection technique helps to reduce the computation time and complexity by selecting the optimum subset of features. In this paper, an enhanced anomaly-based IDS model based on multi-objective grey wolf optimisation (GWO) algorithm was proposed. The GWO algorithm was employed as a feature selection mechanism to identify the most relevant features from the dataset that contribute to high classification accuracy. Furthermore, support vector machine was used to estimate the capability of selected features in predicting the attacks accurately. Moreover, 20% of NSL–KDD dataset was used to demonstrate effectiveness of the proposed approach through different attack scenarios. The experimental result revealed that the proposed approach obtains classification accuracy of (93.64%, 91.01%, 57.72%, 53.7%) for DoS, Probe, R2L, and U2R attack respectively. Finally, the proposed approach was compared with other existing approaches and achieves significant result.

Keywords Intrusion detection system · Feature selection · Multi-objective optimisation · Swarm intelligence · Grey wolf algorithm · Support vector machine · Classification

1 Introduction

The vast advances in information technology and the wide spread of Internet applications have led to increased use by people. Nowadays the use of technology becomes a prerequisite for people's daily life, for example, pay the bills online, flight bookings, watching TV and so on (Kim et al. 2010). In addition, there are many organisations and companies that transfer important information over the network and this information must be delivered to the destination without any modification (Gholipour Goodarzi et al. 2014; Alamiedy et al. 2019). Besides, the spying and hacking techniques become more sophisticated and easily use by an ignorant person. Therefore, there is a need to implement a security system that is able to monitor and inspect the enormous number of packets that pass through the network accurately (Liao et al. 2013).

Furthermore, the existing security techniques like data encryption, client authentication, firewalls, and access controls are utilised as the first line of defence for computer and network security, nevertheless, these techniques cannot furnish an idealistic security circumstance to protect the network entirely (Kim et al. 2010). Moreover, various researchers work on developing a security software/hardware that can reveal various kinds of new attacks and alert to the security staff to take action. One of the most popular security systems that provide higher security in computer

✉ Mohammed Anbar
anbar@nav6.usm.my

Taief Alaa Alamiedy
taiefalaa@gmail.com

Zakaria N. M. Alqattan
zakaria@nav6.usm.my

Qusay M. Alzubi
qusaizoubi@nav6.usm.my

[1] National Advanced IPv6 Centre of Excellence (NAv6), Universiti Sains Malaysia, 11800 USM Penang, Malaysia

networks and to thwart attacks is an intrusion detection system (IDS) (Alamiedy et al. 2019). The concept of IDS was identified first in a technical report by Anderson in 1980. The paper is structured as follows. Section 1.1 presents the concept of intrusion detection system. Section 1.2 illustrates principle of feature selection technique. Section 2 discusses the literature review related to this work. Section 3 presents the description of the benchmark dataset that is used in this work. Section 4 describes the methodology of the proposed approach. The experiment setup and analysis technique are explained in Sect. 5. The result and discussion are covered in Sect. 6, and finally, Sect. 7 concludes the paper indicating future research directions.

1.1 Intrusion detection system (IDS)

IDS is a defensive system that's responsible for identifying intrusions and suspicious activities. This system is operated by monitoring and inspecting the behavior of the client device or network traffic. Beside that, the IDS issues an alarm to notify the security team and register the action into a log file to be used later for further investigation when there is a malicious activity detected in the network (Shen and Wang 2011). In addition, the IDS can be categorised into different classes based on certain criteria; such, as the source of collecting information, detection approach and IDS response type (Liao et al. 2013). Figure 1 demonstrates a typical IDS taxonomy.

The deployment of the IDS sensors in the network is crucial to detect the intrusion successfully or not. Accordingly, the collecting data play an important role in the IDS detection process. This information can be collected either from the client device or network traffic depending on the installation location of IDS sensors. This type of the IDS can be classified into two types, namely host-based IDS and network-based IDS.

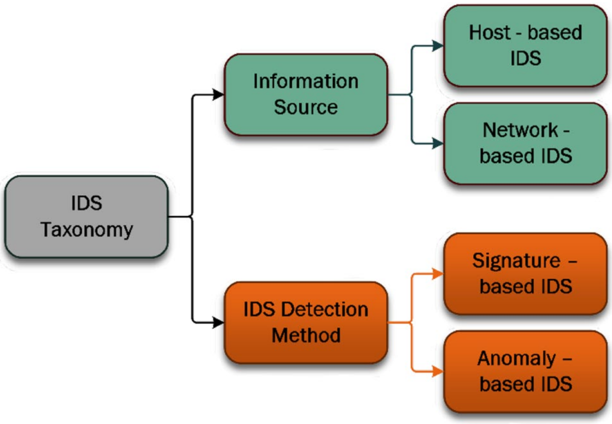

Fig. 1 IDS taxonomy (Wolf, Wolf and Shashua 2005)

1.1.1 Host based-IDS (HIDS)

This approach operates on the client machine and detects the intrusion by reviewing and inspecting the local files in the system such as log files, commands executed and sign in events. In addition, it monitors the usage of hardware resources; like memory, central processor unit (CPU) and hard drive (Vithalpura and Diwanji 2015). Besides, when there is any modification in the system or client files, the IDS directly inform the system administrator.

1.1.2 Network-based IDS (NIDS)

This model detects the intrusion by observing and inspecting the network packets. The NIDS sensors are usually deployed in various locations in the network. These sensors identify the intrusion by scanning the network traffic for any abnormal behavior. In addition, these sensors operate in an inconspicuous mode. Consequently, it is very difficult for the infringers to diagnose their place in the network (Lotfi Shahreza et al. 2011). Additionally, another classification of IDS is based on the detection approach. This type can be classified into signature-based IDS and anomaly-based IDS which are illustrated in the next sections.

1.1.3 Signature-based IDS (SIDS)

The detection process in this approach is based on the comparison between the client activities with predestined attack patterns kept in the database. In addition, the database contains the description of known attacks such as their signatures and attributes (Kumar and Prakash Sangwan 2012). In contrast, the IDS inspect the behavior of the inbound network traffic and matches it with the database through the matching technique. In the case there is a match, then the system will trigger an alarm to notify the security staff (Kumar and Joshi 2011). Furthermore, this approach is competent for detecting known attacks accurately. However, this model must be updated constantly to reveal zero-day attacks.

1.1.4 Anomaly-based IDS (AIDS)

This type inspects the client or networking activities by creating a profile for the regular activities, then it will compare the system events with the normal profile. If any event veers away from the normal profile, the system activity will be treated as abnormal behavior which in turn will trigger a system alert. Beside that, there are various methods used for building the profile; such as statistical data mining, and machine learning methods (Alomari and Othman 2012).

The extensive spread of Internet networks brings about numerous challenges to identify the intrusions expeditiously. As mentioned before, the main duty of the IDS is to monitor and examine the network packets. However, these packets consist of a lot of attributes (features) used to describe the characteristics of the packet, for example, source/destination IP addresses, protocol type, and so on. Besides, there are various repetitious and irrelevant features that curtail the performance of IDS even though the analysis technique is highly sophisticated. Consequently, the IDS must handle meticulously each significant information to detect the abnormal behavior (LIU et al. 2011). In fact, there are several techniques employed to increase the performance of the IDS. One of the most prevalent techniques is feature selection. The following section demonstrates the principle of feature selection technique.

1.2 Feature selection technique

Feature selection is a method of taking a subset of significant features (attributes) by eliminating the superfluous and repetitive features from the dataset for building an adequate learning approach. In addition, this process can shortcut the computation time and complexity (Dastanpour et al. 2014).

A feature selection technique in general comprises many steps as presented in Fig. 2. Firstly, the subset generation step produces a subset of features that are extracted from the original dataset, then, the subset evaluated in the evaluation step based on the objective function (fitness value) to determine the optimum subset of features. Thirdly, the stopping criteria are to decide whether the selected features realise the best result or not. Finally, the validation step checks if the selected features achieve the system requirement or not (Acharya and Singh 2018). Additionally, this technique can be classified into three categories known as wrapper, filter, and hybrid methods. In this work, we use wrapper method during the feature selection stage, the following subsequent provides more details on wrapper method.

1.2.1 Wrapper method

The feature subset in this approach is selected based on the evaluation of machine learning algorithms. These algorithms are employed for the generation and evaluation of the subcategory of features. In addition, the optimum subset of features produced after the algorithms will produce some specific metrics like accuracy, detection rate and so on. Furthermore, this approach aims to diminish the original set of features for producing an effective subset of features. However, these significant outcomes need more processing time and exhaust the system resources (Kumari and Swarnkar 2011). The selection process is launched by generating a subset of features through the initialisation step, then the machine learning algorithm evaluates the selected features with the release of the classification algorithm. Figure 3 illustrates the steps of the wrapper method.

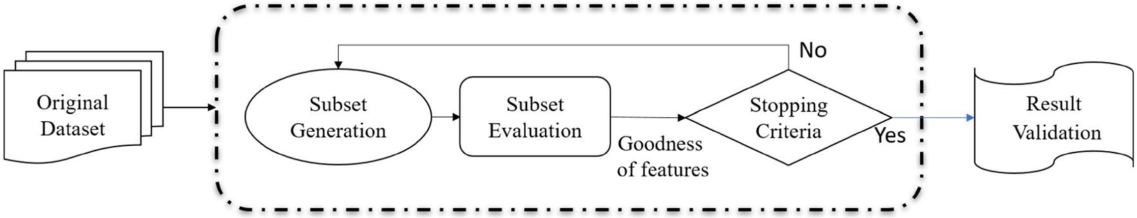

Fig. 2 Feature selection process (Shen and Wang 2011)

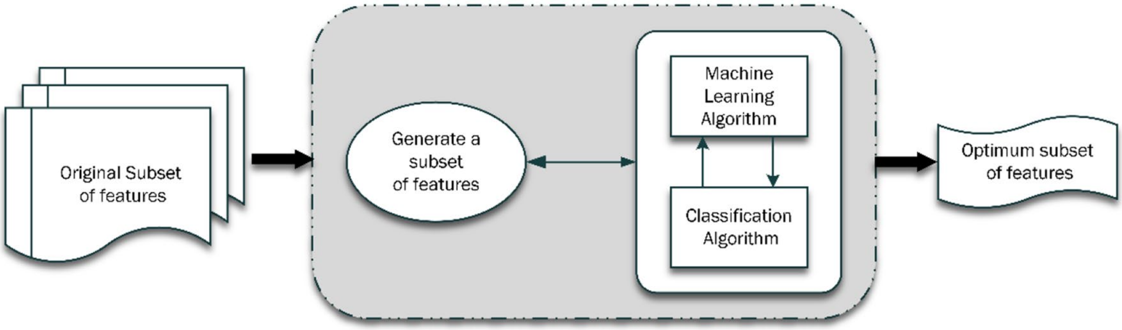
Original Subset of features
Generate a subset of features
Machine Learning Algorithm
Classification Algorithm
Optimum subset of features

Fig. 3 Wrapper method (Kumari and Swarnkar 2011)

2 Literature review

In recent years, many researchers use machine learning algorithms to solve different optimisation problems. The solution for these problems is signified by finding the shortest path (optimum solution). In Internet security, especially, anomaly-based IDS, the optimisation problems like precision, huge datasets, lopsided circulation of information and most troublesomely, to distinguish the limits among typical and unusual parameters. Most of these problems are solved by feature selection technique. In this section, we present various algorithms and methods employed as feature selection to improve the performance of the IDS.

Alomari and Othman (2012) anticipated a wrapper-based component choice approach utilising the bee's algorithm (BA) as an exploration technique for subcategory generation, and also utilising SVM as the classifier. The analyses used four arbitrary subsets gathered from KDD99. Every subset contains around 4000 records. The performance of the anticipated method is assessed by means of the standard IDS estimations. The evaluation criteria in their work based on the balance between the average accuracy with the average of selected features. The experiential result shows that the detection accuracy achieved (99%) and the feature set reduced to (8) features, while the false alarm rate was (0.004).

Alternatively, ant colony optimisation (ACO) and SVM choice feature weighting of network interruption recognition strategy proposed by (Xingzhu 2015). They combined ACO to choose the components with a component weighting SVM. In the first place, they utilised SVM grouping precision and highlight subset measurement to develop a complete fitness weighting index. Subsequently, they used the ACO for an optimisation that is global and numerous exploration capabilities to accomplish for the optimal solution feature search feature. The multi-objective function based on the combination of classification error with weighting features was used in their approach during subset evaluation. Finally, the results exhibited that the proposed approach can successfully reduce the dimension of features and have enhanced network intrusion detection accuracy to (95.75%).

Rani and Xavier (2015) proposed a detection system that is hybrid intrusive. The system is likewise cantered on C5.0 decision tree, which also uses a One-Class SVM. C5.0 is used to train the misuse discovery model in the hybrid intrusion detection system. The mismanagement detection model can distinguish recognised attacks with a low false alarm rate. One-class SVM was applied to the anomaly detection (trained using normal training traffic). In addition, during the training procedure, decision boundaries are chosen with normal data contained in the original dataset. The outliers are detected as using the decision function and the model classifies outliers as attack connection. The experimental results were performed on NSL–KDD Dataset. The overall performance of the planned method was enhanced in terms of the discovery rate and low false alarm rate in the evaluation of this methodical approach. Furthermore, the experimental result shows that their approach was able to reduce a subset of features and improve the classification accuracy to (99%) and reduce the processing time. Beside that, the solution evaluated based on classification accuracy which is considered as a single objective function.

Ghanem and Jantan (2016) anticipated a novel method based on multi-objective artificial bee colony (ABC) for feature selection, particularly for intrusion detection systems. Their approach is classified into two stages: generating the feature subsets of the Pare to front of non-dominated solutions in the first stage and using the hybrid ABC and particle swarm optimisation (PSO) with a feed-forward neural network (FFNN) as a classifier to evaluate feature subsets in the second stage. Thus, the anticipated method consisted of two-fold steps: the first one, using a new feature selection technique called multi-objective ABC feature selection to diminish the number of features of network traffic data and the second one, used a new classification technique called hybrid ABC–PSO optimised FFNN to classify the output data from the previous stage, determine an intruder packet, and detect known and unknown intruders. The proposed approach did not only provide a new approach for feature selection, but also proposed a new fitness function for feature selection to diminish the number of features and achieve the minimum rate of classification errors and false alarms.

Acharya and Singh (2018) proposed an intelligent water drop (IWD) algorithm that is based on the feature selection technique. The method is characterised by the IWD algorithm. Inspired by nature, the method is an optimisation algorithm, is applicable in the selection of feature subset while a vector machine plays the role of a classifier in the evaluation process of selected features. SVM was the classifier used. Amongst parameters used as evaluation were the size of feature subset, false alarm rate and the rate of the classifiers detection. Furthermore, the IWD is a meta-heuristic optimisation algorithm, yielded and optimised procedure of selecting features for SVM. From 41 to 9, the model substantially reduced the features. Parameters found to have been better improved as presented in the new model with a proposed method (IWD + SVM) are precision, false alarm rate, accuracy and rate of detection. This outcome measured improvement over other prevailing models. A precision rate of 99.40% and an accuracy score of 99.09% were recorded in the new model. While a low false rate of 1.4% and a precision rate of 99.10% were also recorded. The period used by this prototype to do the training was remarkably minimised to 1.15 min. Moreover, the score of detection rate was used for subset evaluation during feature selection stage.

In the work of (Negandhi et al. 2019), an intrusion detection system using random forest on the NSL–KDD Dataset was proposed. In their work, the supervised learning algorithm random forests were employed to train a model to detect various networking attacks. In addition, smart feature selection using Gini importance was used to reduce the number of features. The NSL–KDD dataset was used to evaluate the performance of the proposed approach. The experimental results show that the proposed model runs faster and obtained 99.88% of classification accuracy. Beside that, the proposed approach reduced the number of selected features from 41 to 25.

In the study of (Çavuşoğlu 2019), a new hybrid approach for intrusion detection using machine learning methods, a hybrid and layered intrusion detection system was proposed. In their work, they used a combination of different machine learning and feature selection techniques to provide high-performance intrusion detection in different attack types. In the proposed system, firstly data pre-processing is performed on the NSL–KDD dataset, then by using different feature selection algorithms, the size of the dataset was reduced. In addition, they proposed two approaches for feature selection operation which are cfs subset eval and wrapper subset eval with different classification algorithms. The layered architecture is created by determining appropriate machine learning algorithms according to attack type. The NSL–KDD dataset was used to evaluate the performance of the proposed approach. Besides, to demonstrate the performance of the proposed system, it was compared with the studies in the literature. The experimental outcomes show that the proposed system achieves high accuracy and low false-positive rates in all attack types. Table 1 shows a comparison of bio-inspired feature selection algorithms and presents a summary of existing studies.

3 Existing studies based on GWO algorithm

The following studies present different types of feature selection based on GWO algorithm which used to solve the optimisation problems in the intrusion detection system.

Devi and Suganthe (2017) proposed a wrapped feature selection method based on GWO algorithm, they used a multi-objective fitness function to evaluate a subset of features in the feature selection stage, then for classification stage, they combine SVM with a Naive Bayes classifier. In addition, they used NSL–KDD to evaluate their system performance. In addition, they utilised the mutual information to evaluate the candidate solution selected through feature selection stage. Finally, the experimental result of their approach had shown taking a faster time to detect attacks and produced good positive rates significantly and they were able to reduce the feature set of 18 features. Furthermore, they achieved 99.89% for classification accuracy of DoS attacks.

In the work of (Seth and Chandra 2016), a key feature selection based on GWO algorithm was proposed. In their approach, GWO was used to reduce the original set of features. In addition, NSL–KDD benchmark dataset was used to evaluate the proposed approach. Furthermore, the experimental result shows that their approach was able to reduce a subset of features and improve the classification accuracy to (99%) and reduce the processing time. Beside that, the solution evaluated based on classification accuracy which is considered as a single objective function.

An improved GWO algorithm integrated with cuckoo search (CS) proposed by (Xu et al. 2017). In their approach, they combined GWO with CS in order to improve the performance of the GWO. In addition, they observed that the feature (service) contributes high false positive rate, therefore, they eliminated it from the dataset. Furthermore, the experimental results show that the proposed CS–GWO algorithm achieve a better result compared to the standard version of both algorithms. Moreover, the proposed algorithm achieved 83.54% for classification accuracy and reduced the feature number into 6 features.

In the study of (Roopa Devi and Suganthe 2018) proposed a hybrid GWO algorithm with CS algorithm as a feature selection model combined with transudative support vector machine (SVM) for classification stage, in the approach they used the min–max method during the pre-processing step. The optimal subset of features extracted based on maximum mutual data, they used maximum mutual data used as the fitness function, the experimental result shows the proposed approach reduces the number of selected features to (18,17,34,8) for DoS, Probe, U2R and R2L attacks respectively.

Zawbaa et al. (2018) proposed a hybrid bio-inspired heuristic approach for large-dimensionally small-instance set feature selection. In their work, they hybridised antlion optimisation with grey wolf optimisation. The proposed system evaluated by using 50,000 features and 200 instances. The results were compared to the genetic algorithm and particle swarm optimisation; however, the proposed system produced better performance in terms of high accuracy of prediction, and the process was complex.

Velliangiri (2019) proposed hybrid intrusion detection model based on binary GWO (BGWO) with kernel principal component analysis for intrusion detection, in their approach, they used KPCA for select the optimum subset of features and multi-class SVM for classification stage. In addition to that, they combined KPCA with the SVM classifier to improve the classification process, the GWO algorithm employed to select the best values for SVM classifier.

Table 1 Feature selection approaches based on various bio-inspired optimisation algorithms

Authors and year	Feature selection algorithms	Dataset	Feature length	Classification algorithms	Performance evaluation	Objective function	Work limitations
Alomari and Othman (2012)	Bees algorithm	KDD CUP99	8	Support vector machine	Detection rate: 98.38 False alarm rate: 0.004	Multi—objective (average accuracy with average feature numbers)	Result not appreciable for all classes
Xingzhu (2015)	Ant colony + feature weighting support vector machine	KDD CUP99	13	Support vector machine	Detection rate: 95.75	Multi—objective (classification error with weighting features)	Select a large set of features and focusing only on detection rate
Rani and Xavier (2015)	Cuttle fish algorithm	NSL–KDD	–	C5.0 + one class SVM	Accuracy: 98.20 False alarm rate: 1.405	–	Result calculated only for accuracy and error rate only
Ghanem and Jantan (2016)	Artificial bee colony optimisation	NSL–KDD	–	Feed-forward neural network	–	Multi—objective (classification error rate with false alarm rate with feature rate)	The result was not appreciable for all classes
Acharya and Singh (2018)	Intelligent water drop algorithm	KDD CUP99	9	Support vector machine	Detection rate: 99.40 false alarm rate: 1.405	Single—objective (detection rate)	False alarm rate is high and select one type of attack
Negandhi et al. (2019)	Gini importance	NSL–KDD	25	Random forest	Classification accuracy: 99.88%	–	Select a large set of features
Çavuşoğlu (2019)	Cfs subset eval and wrapper subset eval	NSL–KDD	25	Naïve bayes, random forest, J48, random tree	DoS Accuracy: 99.98%	Single—objective (accuracy)	Selecting a large set of features and complexity of analysis

Furthermore, the KDD-99 dataset was used to evaluate the performance of the proposed model, and they were indicated that the proposed method can reduce the training time and testing time. Finally, the proposed approach obtains accuracy performance of (96.82%, 95.38%, 75.502%, 74.56%) for (Probe, DoS, U2R, and R2L) attacks Consecutively.

The authors (Srivastava et al. 2019a) proposed a nature-inspired technique for intrusion detection system (IDS), in their work, they use grey wolf optimiser as feature selection and they applied different types of classification technique lie k-nearest (KNN), support vector machine (SVM) and generalized regression neural network (GRNN). In addition, they utilised 10% of KDD-99 dataset for testing the model. The experimental result clarifies that the combination model of (GWO–KNN) achieves the best result in term of accuracy, sensitivity, and specificity compared to the other approaches.

The authors (Srivastava et al. 2019b) proposed and implemented different hybrid methods for intrusion detection system, in their work, they used grey wolf optimisation (GWO) algorithm with several classification techniques like entropy basic graph (EBG), support vector machine (SVM), generalised regression neural network (GRNN) and k-nearest neighbor (KNN), the KDD-99 dataset was utilised to assess the classification of data into normal or intrusion using different hybrid classification techniques. Besides, the authors divide the testing data into different volumes and measure the performance of the proposed approach. The outcomes show that the GWO–EBG classification approach obtains the higher result compared to the other approaches.

In addition to that, the grey wolf optimisation (GWO) algorithm also implemented in other fields like science, medicine, industry, education and so on. The following studies present examples for using GWO algorithm to solve different types of optimisation problems.

The authors (Makhadmeh et al. 2018) proposed a multi-objective grey wolf optimisation (GWO) algorithm to solve the power scheduling problem in smart homes, they used GWO to achieve an optimal schedule. In addition, they evaluated their approach using seven consumption profiles and seven real-time electricity prices with different characteristics. Moreover, in their work, they used three factors for evaluated the proposed approach which are electricity bill, peak-to-average ratio (PAR), and user comfort level. The experimental result shows that the proposed approach obtains a significant impact on the final schedule.

A hybrid genetic grey wolf algorithm (HGWO) for large scale global optimisation (LSGO) was proposed by (Gu et al. 2019), In their work, they combined GWO with three genetic factors to improve the demerit of GWO when solving the LSGO issues, three genetic operators are embedded into the standard GWO and a hybrid genetic grey wolf algorithm (HGGWA) was proposed. The performance of HGGWA was verified by ten benchmark functions. Finally, the simulation results show that the HGGWA was greatly improved in convergence accuracy, which proves the effectiveness of HGGWA in solving LSGO problems.

In the work of (Garg et al. 2019), a hybrid deep learning-based model for anomaly detection in cloud data centre networks was presented. In their research, a hybrid data processing model for network anomaly detection was proposed that powers the performance of grey wolf optimisation (GWO) and convolutional neural network (CNN). The proposed model works in two phases for efficient network anomaly detection. In the first phase, improved GWO was used for feature selection. In the second phase, improved CNN was used for classification stage. The efficacy of the proposed model was validated on the benchmark (DARPA’98 and KDD’99) and synthetic datasets. The results obtained demonstrate that the proposed cloud-based anomaly detection model was superior in comparison to the other state-of-the-art models. In average, the proposed model exhibits an overall improvement of 8.25%, 4.08% and 3.62% in terms of detection rate, false positives, and accuracy, consistently; relative to standard GWO with CNN.

A study of Experienced grey wolf optimiser through reinforcement learning and neural networks was presented by (Emary et al. 2018). In their work, a variant of GWO that uses reinforcement learning principles combined with neural networks to enhance the performance of the system. In addition, they utilised reinforcement learning to set it on an individual basis. The resulted algorithm is called experienced GWO (EGWO) and its performance was assessed on solving feature selection problems and on finding optimal weights for neural networks algorithm. Beside that, they used a set of performance indicators to evaluate the efficiency of the proposed method. The Result shows that the proposed over various datasets demonstrate an advance of the EGWO over the original GWO and other meta-heuristics such as genetic algorithms and particle swarm optimisation.

In the work of (Emary et al. 2015), a feature subset selection approach by GWO was presented, in their study, a classification accuracy-based fitness function was proposed by GWO to find optimal feature subset. The aim of the GWO in this work was to find optimal regions of the complex search space through the interaction of individuals in the population. The proposed approach proves better performance in both classification accuracy and feature size reduction compared with particle swarm optimisation (PSO) and genetic algorithm (GA) over a set of UCI machine learning data repository, Moreover, the gray wolf optimisation approach proves much robustness against initialisation in comparison with PSO and GA optimisers.

The authors (Emary et al. 2017), proposed a method of multi-objective retinal vessel localisation using flower pollination search algorithm with pattern search. In this work, the proposed multi-objective fitness function uses flower pollination search algorithm (FPSA) to find optimal clustering of the given retinal image into compact clusters under some constraints. In addition, the pattern search (PS) method also used to enhance the segmentation results using another objective function based on shape features. The database namely DRIVE dataset was used to evaluate the performance of the proposed approach. The proposed approach also compared with state-of-the-art techniques in terms of accuracy, sensitivity, and specificity.

A study on the impact of chaos functions on modern swarm optimisers was identified by (Emary and Zawbaa 2016). In their study, they used chaos-based control of exploration/exploitation rates against using systematic native control. Three recent algorithms were used in their work namely grey wolf optimiser (GWO), antlion optimiser (ALO) and moth-flame optimiser (MFO) in the domain of machine learning for feature selection. In addition, they used a set of standard machine learning data with a set of assessment indicators. The experimental outcomes proved that the performance of optimisation algorithm enhanced by using variational repeated periods of declined exploration rates overusing systematically decreased exploration rates.

The authors (Lu et al. 2017) investigated and proposed a true unique welding booking issue explored from the hypothesis and handy application points of view. In the first place, they figured out a multi-objective scientific model which considered three dynamic occasions comprised of machine breakdown, work with discharge time postponement and employment with low quality at the same time. This model additionally includes succession subordinate setup time, work subordinate transportation times and controllable preparing times. At that point, the author builds up a crossbreed multi-objective grey wolf optimisation agent (HMO–GWO) to address this dynamic issue with the goal to limit the make span, machine load, and precariousness at the same time. It effectively minimizes the make span, machine load, and instability simultaneously. The weaknesses of this method are, it does not use information on problem properties and consumes a long time to attain a set of non-dominated solutions.

4 NSL–KDD dataset description

NSL–KDD is a simulated dataset proposed by (Dhanabal and Shantharajah 2015) to solve most of the inseparable issues in the KDD-99 dataset. The NSL–KDD dataset is an improved version of KDD-99, which contains a smaller number of irrelevant and repetitive records in comparison to the original dataset. In addition, the NSL–KDD was used in this work since it is the most popular dataset in intrusion detection field (Özgür and Erdem 2017). Besides, we use 20% of NSL–KDD dataset to evaluate the proposed model. Furthermore, this part of the dataset contains 25,192 samples for the training set and 11,850 samples for the testing set. The benefits of the NSL-KDD that stands out to the original KDD-99 dataset illustrated in (Tavallaee et al. 2009).

4.1 Dataset attacks types

This part presents the main classes of attacks in the NSL-KDD dataset. The dataset mainly contains four types of attacks which are illustrated in the following list of points:

- Denial of Service (DoS): In this type of attack, the attacker tries to keep the system or memory resources too busy. Therefore, this process will make the machine unable to handle any request from legitimate users or perform any other services.
- User to Root (U2R): The attacker starts the attack by obtaining some legitimate user credits. Then, the attacker exploits the system vulnerabilities for getting permission to user root rights.
- Remote to Local (R2L): The attacker sends a packet to the machine that is connected to the network. Afterwards, the attacker attempts to observe the vulnerabilities and exploit privileges in the host system to gain access. Then, the intruder becomes an administrator of the remote machine.
- Probing Attacks: The attacker scans the client/network machine to collect information. This information is very useful to determine the weaknesses and vulnerabilities in the system that may be used later to compromise the client's machine system. Figure 4 shows number of normal and attacks instances in dataset training and testing sets.

In addition, each type of main attacks contains many sub-attacks, these types contain some features that help to identify the main type of attack. Figure 6 presents the distribution of attacks in the training and testing set.

5 Methodology of the proposed model

In this section, we provide an overview on the framework of the proposed approach. The overall framework is shown in Fig. 5.

The next subsections provide more detail on the proposed model stages.

5.1 Stage 1—dataset preparation

This stage plays an important aspect in the machine learning and data analysis algorithms. The task of preparing data is to transform and present the data in appropriate format. Especially, when the data comprises of different formats and

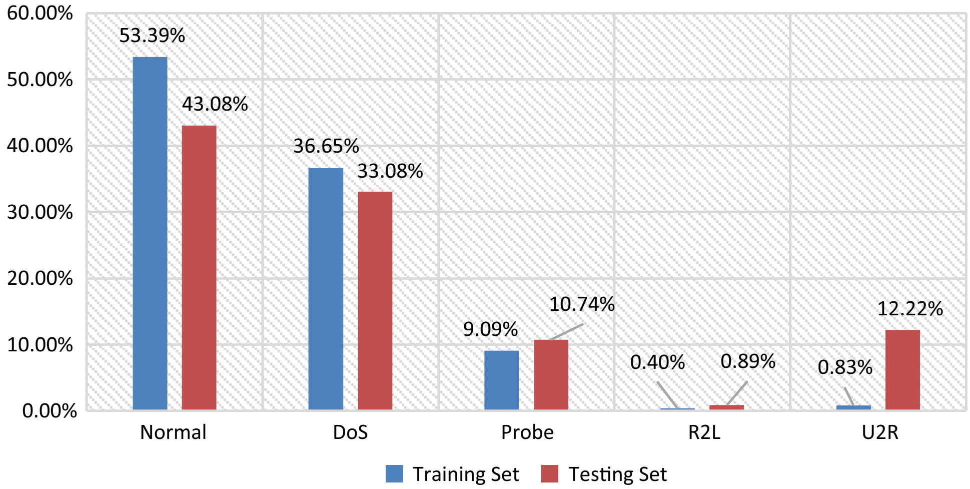

Fig. 4 Number of normal and attacks instances in dataset training and testing sets

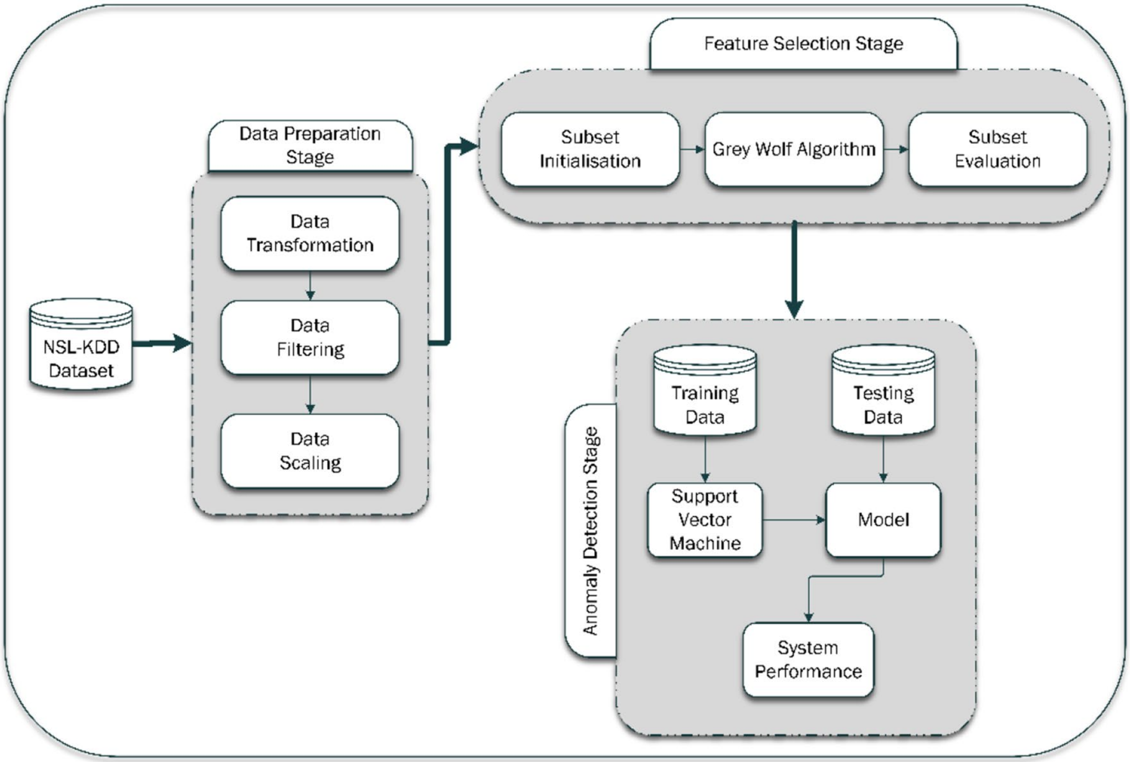

Fig. 5 Architecture of the proposed model

a wide range of information values. This stage consists of the subsequent steps.

5.1.1 Step 1—data transformation

The NSL-KDD dataset comprises many features and data presented in various formats like alphabet, numbers, symbols and so on. The analysis of these features might take more processing time and consume a lot of hardware resources. Consequently, to avoid these dilemmas, the transformation process was implemented to map symbolic features to numeric features (Shah and Trivedi 2013). Figure 6 shows an example of data transformation process.

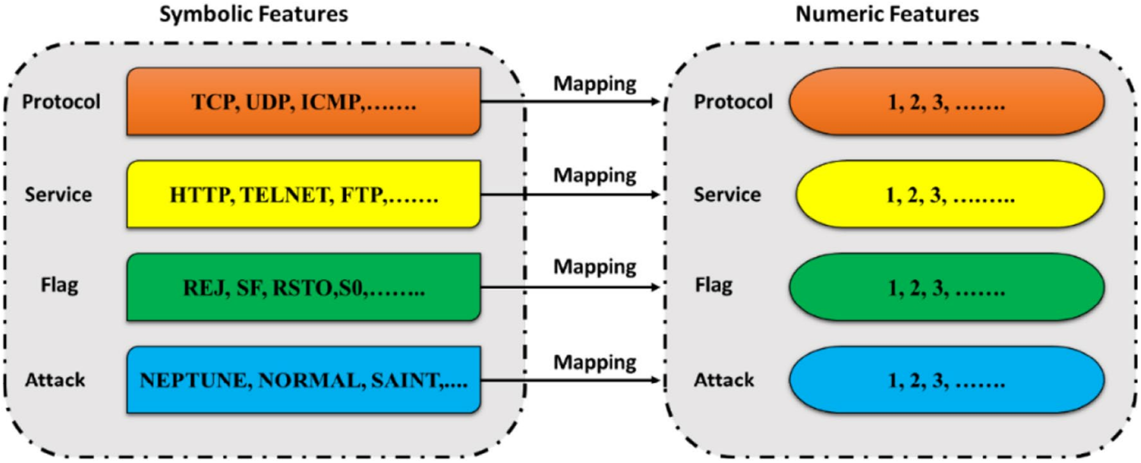

Fig. 6 Example of data transformation process

Table 2 List of features adjusted in normalization step

Feature no.	Feature name	Description
1	Duration	Length (number of seconds) of the connection
2	Protocol_type	Type of protocol, e.g. tcp, udp, etc
3	Service	Network service on destination, e.g. HTTP, telnet, etc
4	Src_bytes	Number of data bytes from source to destination
5	Dst_bytes	Number of data bytes from destination to source
6	Flag	Normal or error status of the connection
8	Wrong_fragment	Number of "wrong" fragments
9	Urgent	
10	Hot	
11	Num_failed_logins	Number of logins failed attempted
13	Num_compromised	Number of "compromised" conditions
16	Num_root	Number of "root" access
17	Num_file_creations	Number of files creation operations
18	Num_shells	Number of sell prompts
19	Num_access_files	Number of operations on access control files
23	Count	Number of connections to the same host as current connection in the past 2 s
24	Serror_rate	% of connection that have "SYN" errors
30	Srv_rerror_rate	% of connection that have "REJ" errors
31	Srv_diff_host_rate	% of connection to different hosts
32	Dst_host_count	Destination host count
33	Dst_host_srv_count	Service count for destination host
41	Class	Describe the type of traffic (normal or attack)

5.1.2 Step 2—data normalisation

This process is defined as a method of calibrating the range of feature values into a well-proportioned range. In this work, each value in the feature record is scaled using Eq. (1).

$$X\text{\`} = \frac{X}{X\_maximum}. \quad (1)$$

where $X\text{\`}$ is the normalised value, X is the current value in the feature's record and $X\_maximum$ refer to the maximum values in the feature record. Finally, the range of record values falling between zeros and one values. Table 2 presents a list of features adjusted in this step.

5.1.3 Step3—data filtering

The filtering step is typically used to select or eliminate some information from the dataset. In this work, the filtering method was utilised to extract and detach different classes of attacks to test the proposed approach in different types of attack scenarios. Figure 7 illustrates the filtering process.

From Fig. 7, it could be noted that the NSL–KDD dataset contains numerous classes of attacks. Additionally, every sub-attack refers to the main category of the dataset attacks; like: Denial of Service (DoS), Probe, Remote to Local (R2L) and User to Root (U2R). Consequently, in this step, each class of attack mapped to the main attack category. Lastly, the output is different dataset containers and each container has a different kind of dataset attacks.

5.2 Stage 2: feature selection

After the data is prepared, then the feature selection step is utilised to select the optimal subset of features. In this work, the GWO algorithm was adapted to select the optimal set of features. The following subsections give further details for the steps of stage 2.

5.2.1 Step 1: subset generation

Subset generation is a technique of heuristic search, within which every sample in the search area specifies a candidate solution for subset evaluation. In this work, the random subset generation technique (Kim et al. 2010) was used to generate a subset of features. The following Equation used in the initialisation step to generate the solution.

$$X_{(i,j)} = x_j^{\min} + \delta\left(x_j^{\max} - x_j^{\min}\right). \tag{2}$$

where x_{ij} is dimension of matrix generated in the initialisation step, x_j^{max} and x_j^{min} represent the upper and lower bound of the matrix, and the values of parameter i form 1 to N, and the j parameter from (1 to D). Where N represents the number of solutions and D refer to the dimension of the solution in the matrix.

5.2.2 Step 2: grey wolf optimisation algorithm

The GWO is swarm-based algorithm, which is proposed by (Mirjalili 2014). The GWO is motivated by the social behavior of grey wolves in nature. The chasing and hunting

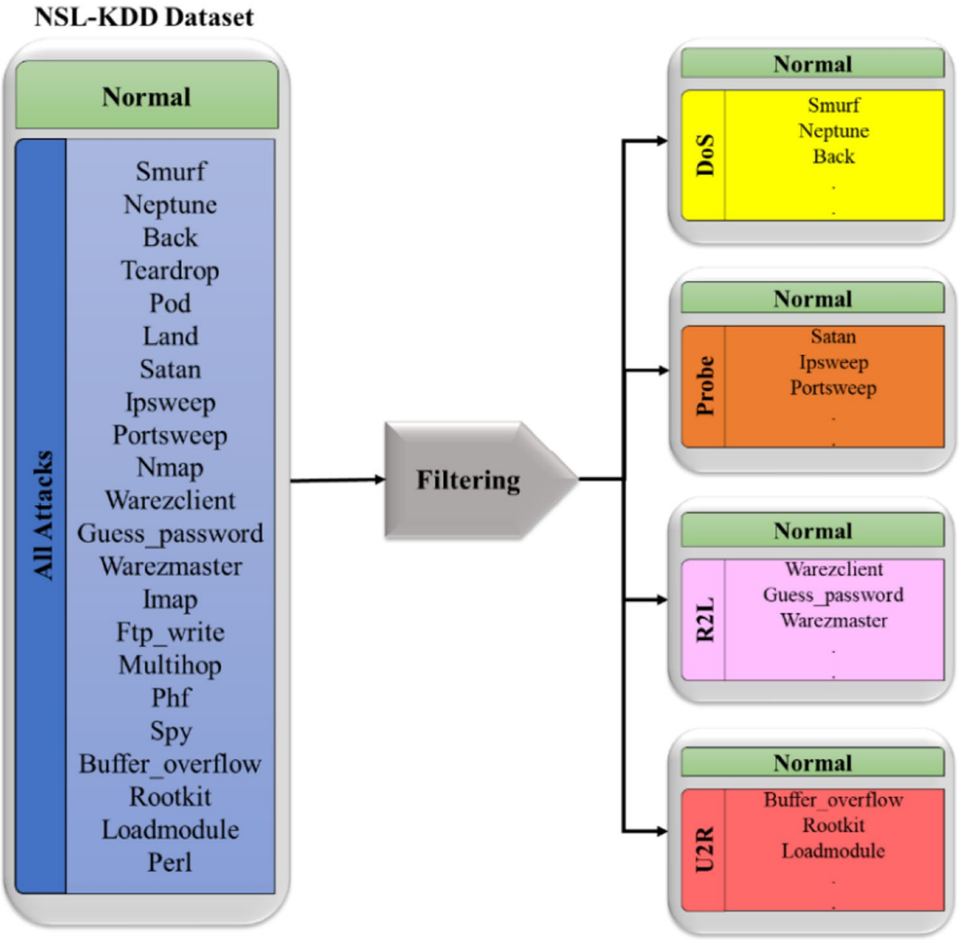

Fig. 7 Data filtering process

behavior of wolves to catch the prey represent the searching path to the optimal solution. In nature, grey wolves prefer to live in a pack. The average size of the pack varies from 5 to 12 wolves (Mirjalili 2014). In addition, the packs' members are classified into four groups based on the level of the wolf's position in the pack that assists in improving the hunting process (Alzubi et al. 2019). These groups are named as follows: Alpha (α) consist of a male or a female, these wolves are the leaders in the pack and responsible for decision making, for example, hunting, waking, sleeping time and place. Besides, beta (β) is a second level which consists of male or female wolves and responsible for helping in some decisions for the other wolves in the packs. Delta (δ) is the third level and they perform some important roles such as caretaker, sentinels, an elder in the pack and hunter. The last level is omega (ω). This level is the weakest of the lves in the hierarchal model and plays a role of scapegoat and should obey other individuals' orders (Emary et al. 2016).

5.2.2.1 The mathematical model of GWO algorithm is section provides details of the mathematical model encircling, hunting and attacking the prey as follows. Firstly, encircling the prey: as discussed above, the grey wolves start to encircle the prey during the hunting process. The mathematical expression for this step is presented in the following Equations.

$$\vec{D} = \left|C.\vec{X}_p(t) - \vec{X}(t)\right|. \tag{3}$$

$$\vec{X}(t+1) = \vec{X}(t) - \vec{A}.\vec{D} \tag{4}$$

where t is the existing iteration, (A and C) are coefficient matrix vectors, X_p. is the position vector of the prey and X is the position vector of the grey wolf. The vectors (A and C) are described as follows:

$$\vec{A} = 2\vec{d}.\vec{r}_1 - \vec{d}. \tag{5}$$

$$\vec{C} = 2.\vec{r}_2. \tag{6}$$

where (a) decreases from 2 to 0 over iterations and r_1, r_2 are random vectors. Secondly, hunting the prey: The grey wolves have the knowledge to determine the location of prey and hunt it, at this step, they regularly follow the alpha wolf. In addition, the beta, and delta might also participate in hunting infrequently. The alpha wolf assumes to be the best solution, whereas beta and delta have excellent knowledge about the possible position of prey. Therefore, the position of alpha, beta and delta will be utilised to adjust the positions of the other wolves including omega wolf as described in the following Equations:

$$\vec{x}(t+1) = \frac{1}{3}\vec{X}_1 + \frac{1}{3}\vec{X}_2 + \frac{1}{3}\vec{X}_3 \tag{7}$$

where X_1, X_2 and X_3 are given by following Equations:

$$\vec{X}_1 = \vec{X}_\alpha(t) - \vec{A}_1.\vec{D}_\alpha \tag{8}$$

$$\vec{X}_2 = \vec{X}_\beta(t) - \vec{A}_2.\vec{D}_\beta \tag{9}$$

$$\vec{X}_3 = \vec{X}_\delta(t) - \vec{A}_3.\vec{D}_\delta \tag{10}$$

where $X\alpha$, X_β and X_δ are the positions of alpha, beta and delta wolves in iteration; i.e., the first three best solutions of our problem. A_1, A_2 and A_3 are presented in Eqs. 8, 9, and 10 respectively, and D_α, D_β and D_δ are given by the following Equations:

$$\vec{D}_\alpha = \left|\vec{C}_1.\vec{X}_\alpha - \vec{X}\right| \tag{11}$$

$$\vec{D}_\beta = \left|\vec{C}_2.\vec{X}_\beta - \vec{X}\right| \tag{12}$$

$$\vec{D}_\delta = \left|\vec{C}_3.\vec{X}_\delta - \vec{X}\right| \tag{13}$$

where C_1, C_2 and C_3 are shown in Eqs. 11, 12 and 13 respectively. Figure 8 presents how the search wolves change their location with respect to alpha, beta, and delta in the search space. It can be noted that the last location would be in a random position within a circle which is described by the positions of alpha, beta, and delta in the search space. More simply, alpha, beta, and delta define the location of the prey, and the other wolves renew their locations randomly around the prey (Mirjalili 2014).

Finally, attacking the prey: when the prey stops moving, the wolves start to attack it. The value of a decrease from 2 to 0 over the course of the iteration that means the wolves are approaching the prey. The following Equation describes the value of *a:*

$$\vec{a} = 2 - \frac{2 \times t}{MaxItr} \tag{14}$$

where t is the existing iteration and $MaxItr$ is the maximum amount of iterations. The GWO algorithm was selected as a feature selection mechanism in this work. Due to its behavior based on meta-heuristics and has the ability to find the optimum solution and avoid the stack on one solution. Furthermore, it shows a powerful performance in the unexplored and challenging search areas. Moreover, it has very few control parameters and is easy to implement. Besides, the GWO takes the decision based on the three best leading search agents (Liao et al. 2013). The following Equation is used to convert the solution generated by the GWO algorithm to binary values in the initialisation step and selection of the best solution in the final step (Kiran 2015).

$$Z_i = \text{round}\left(\left|y_i \text{mod } 2\right|\right)\text{mod}2 \tag{15}$$

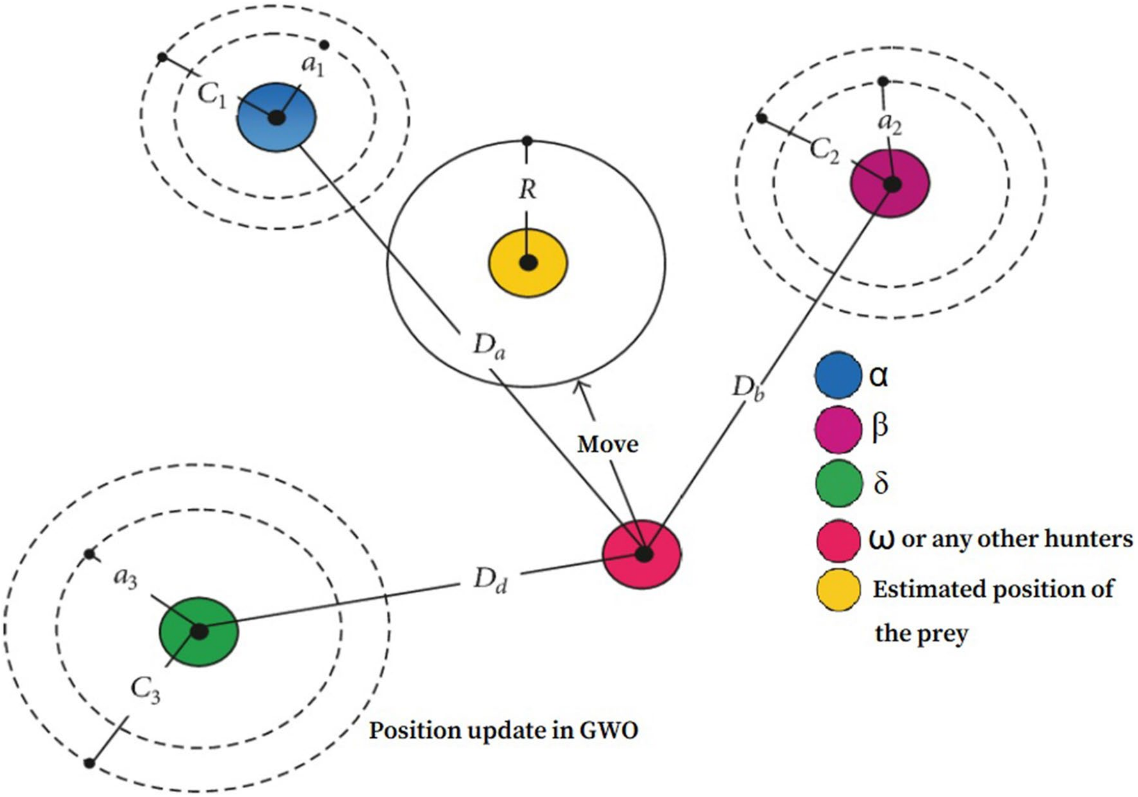

Fig. 8 Position updating in GWO

where Z_i is the binary value (discreate value) represented by 0 or 1, i refer to number of solutions and y_i is the value of solution (continues values) generated through initialisation and final steps. Besides, by using Eq. (15), if the absolute value of remainder is between 0 and 0.4999 or 1.5 and 1.9999, the binary number is obtained as 0. Else if the absolute value of the remainder is between 0.5 and 1.4999, the binary number is obtained as 1.

Moreover, after completing one iteration by the algorithm, the first three best solutions x_α, x_β and x_δ as positions of alpha, beta and delta wolves which will attract the other wolves in the pack. The solution (position) that has the best classification accuracy is alpha's position, and then beta and followed by delta. In each iteration of the algorithm, the classifier is trained and validated, then the accuracy of the classifier is computed toward each subset (solution) of the position matrix (Mirjalili 2014).

Furthermore, in each iteration of the algorithm, the position of each wolf in the pack changes, hence the change in the positions of alpha, beta and delta wolves. All solutions are on the corner of a hypercube. The grey wolf positions are converging towards the prey in each iteration. The wolf nearest to the prey is the best solution; i.e., alpha position. Figure 9 illustrates the flowchart of the proposed GWO algorithm.

5.2.3 Step 3: objective function (subset evaluation)

The objective function (fitness value) is used to evaluate each candidate solution selected by GWO algorithm. The following points give more details on the standard and the proposed multi-objective function.

- *Single Objective Function:* The standard objective function which was used by many researchers to evaluate a subset of features based on classification accuracy and ignoring the number of selected features. The classification accuracy can be calculated based on Eq. (16).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{16}$$

where the false positive (FP) represents the number of samples incorrectly predicted as attack class, false negative (FN) refer to the number of samples incorrectly predicted as a normal class, whereas true positive (TP) is the number of samples correctly predicted as attack class, and true negative (TN) indicates to the number of samples correctly predicted as a normal class. Beside that, the benefit of this objective function is that it performs high classification accuracy. However, it will not give attention to the number of selected

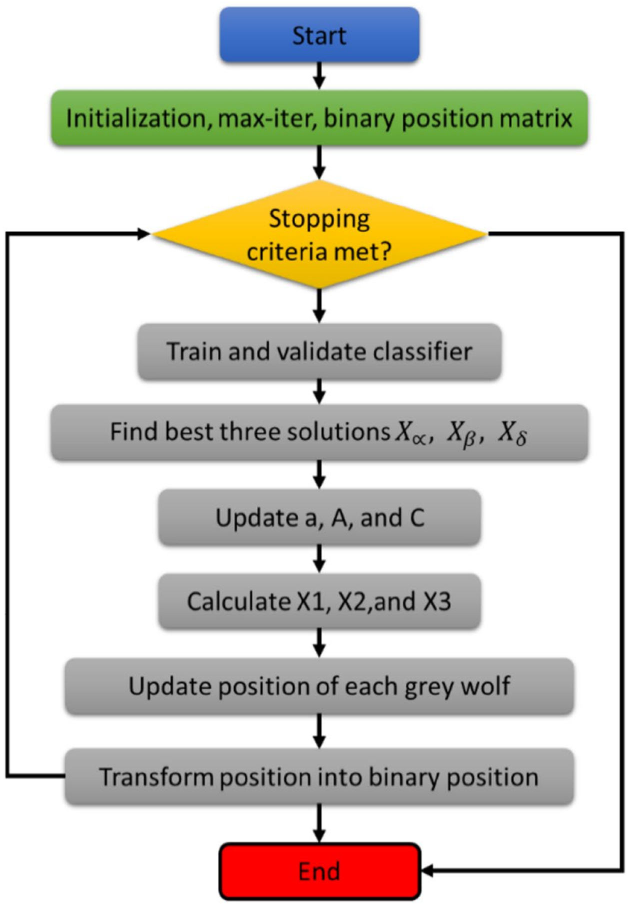
Start
Initialization, max-iter, binary position matrix
Stopping criteria met?
Train and validate classifier
Find best three solutions X_α, X_β, X_δ
Update a, A, and C
Calculate X1, X2,and X3
Update position of each grey wolf
Transform position into binary position
End

Fig. 9 Flowchart of the proposed GWO algorithm (Mirjalili 2014)

features. Therefore, it will perform high classification accuracy, but with a large subset of features which will need high computational time and more hardware resources. To overcome the limitation of standard objective function, a multi-objective function was proposed in this paper.

- Proposed Multi-Objective Function: The proposed multi-objective function is known as weighted sum fitness function. The aim of this fitness evaluation in this work is to reduce the number of selected features and meanwhile, attain high classification accuracy. The fitness of the candidate solution is evaluated based upon the accuracy and number of features present in a solution. The accuracy is calculated using Eq. (16), and the fitness value for proposed multi-objective function calculate using the following formula.

$$\mathit{Proposed\ Fitness\ Function} = W_1 \times \mathit{Accuracy} + W_2 \times \frac{1}{\mathit{Number\ of\ Selected\ Features}} \tag{17}$$

Each feature subset contains a list of features. If two subsets achieve the same accuracy, while having a different number of features, the subset with fewer set of features will be selected. In addition, the values of (W_1) and (W_2) in the Eq. (17) are configurable, and the condition for that (W_1) is multiplied by *Accuracy* and (W_2) multiplied by the inverse *Number of Selected Features*. Beside that, among *Accuracy* and *Number of Selected Features*, the accuracy is the key concern, so more weight to accuracy (W_1) is given than a number of selected features (W_2) to be selected. Therefore (W_1) should be larger than (W_2) in all situations.

Finally, the drawbacks of the proposed technique could be noted from the results of some other performance measures like the false positive, true negative, true positive and false negative does not have any improvement or might obtain a worse result in some kinds of attacks. The reason for getting these outcomes due to the fact that the scope of our work is to achieve high classification accuracy with the lowest number of selected features as stated in Eq. 17, Beside that, another reason for that, in some type attacks the number of unknown attacks in the testing set might exceed the number of known attacks in the training set. As a consequence, the proposed technique will negatively affect the performance of the classification algorithm to identify the attacks accurately.

5.3 Stage 3: anomaly detection

In this stage, support vector machine (SVM) was used to evaluate the selected features from Stage 2. SVM is a machine learning technique produced by (Cortes 1995). SVM proves beneficial without the need for any prior experience. Also, it does not undergo the local minimum trap. In addition, the SVM has a quick execution time even for huge dimensional noisy datasets. These characteristics match the requirements needed for achieving an efficient IDS (Tribak et al. 2012).

SVM can be classified into two classes known as learner and non-learner. The learner approach is used to classify simple information. Whereas, the non-learner type handles the complicated dataset classification. Beside that, the operation of non-learner approach is based on the kernel function which are polynomial, gaussian and gaussian radial basis function (GRBF). Furthermore, in this work. The radial basis function (RBF) was used. This kernel function is a good solution due to its fewer controllable parameters and an excellent nonlinear forecasting performance.

From the figure above, it could be noted that the output from stage 2 will be input to the classifier. In addition, the dataset was split into a training set (X_{train}, Y_{train}) and testing set (X_{test}, Y_{test}), X_{train} represents the features in the training set and Y_{test} refers to the class label in the training set. While X_{test} consists of features in the testing set and Y_{test} contains

Table 3 System parameter settings

items	Values
Number of populations	10
Number of iterations	40
Number of runs	20
Dimension of the solution	41
Range of search space	[0, 1]
Weight of the proposed fitness function	$W_1 = 0.7$, $W_2 = 0.3$

the class label for the features in the testing set. The feature set is represented by X_1, X_2, and the class label is represented by Y, the classifier will train using X_{train} and Y_{test}, then X_{test} will be used as input to the model. After that, the output of the model will be compared to the Y_{test}, if there is a match, that means the classifier accurately predicted the behavior of the record in the dataset.

6 Experiment setup and analysis

There are many platforms used for data analysis process like data preparation, feature selection, clustering, classification and so on. The most popular programming tools used in the machine learning field are Weka, knime, RapidMiner, MATLAB and so on. In this work, we use MATLAB to conduct the proposed approach. In addition, to clarify the performance of the proposed model, 20% of NSL–KDD dataset was used and the experiments are performed on 3.2 GHz Core i7 processor with 12 GB of memory running on windows 10 platform. Table 3 presents more details for the system parameter settings.

Moreover, different types of attack scenarios were used to evaluate the proposed approach and the performance of the IDS was measured by classification accuracy and number of selected features.

6.1 Performance metrics

In order to evaluate the effectiveness of the proposed approach model, we use the most popular evaluation metric which is the classification accuracy. In addition, the classification accuracy computed in different cases which is average, best and worst case and the number of selected features for each status is also recorded. The average was calculated using the following Equation:

$$\frac{\sum_{n=1}^{n=20} A(n)}{N} \tag{18}$$

where n refers to the range of value of experiments run, N represents the total number of the experiment runs, and $A\ (n)$ represent the final value of classification accuracy obtained by the standard and proposed multi-objective GWO algorithms in each run of the experiment. Beside that, the best accuracy value represents the maximum value reached through each run of the experiment, whereas the worst value for classification accuracy represent the lowest values obtained by the standard and proposed multi-objective GWO algorithms in each run of the experiment.

7 Result and discussion

This section presents the experimental result and discussion of the proposed model in all attacks scenarios. Table 4 shows the experimental result of the confusion matrix parameters for SVM classifier with full set of features, standard GWO algorithm and proposed GWO algorithm for all attack scenarios. It observed that for TP parameter, the multi-objective GWO algorithm achieves the best result in Probe and R2L attack scenario. However, the standard GWO algorithm obtains the best result in DoS attack scenario, whereas in U2R attack, the standard GWO algorithm and multi-objective GWO algorithm attain the same result. Therefore, it could be concluded that the multi-objective GWO algorithm increases the detection of abnormal instances that classified successfully in Probe and R2L attacks.

Beside that, regarding to the result of FP parameter, the multi-objective GWO algorithm accomplish the highest result in DoS attack scenario, while in Probe, R2L and U2R attacks, the standard GWO algorithm and multi-objective GWO algorithm gain the same result. This result was expected because of the fitness function in the standard and multi-objective GWO algorithm focus on increasing the classification accuracy and did not focus on minimizing the system false alarm rate. With respect to the result of TN parameter, the standard GWO algorithm and multi-objective GWO algorithm acquire the best result for U2R attack. However, for the other types of attacks, the performance of standard GWO algorithm and multi-objective GWO algorithm achieve the lowest result. Finally, for FN parameter, the multi-objective GWO algorithm obtains the best result in probe and R2L attacks scenario and these results shows the impact of multi-objective function in reducing the number of abnormal instances that are classified incorrectly. However, in DoS attack, the standard GWO algorithm achieves the best result.

Table 5 presents the result of maximum classification accuracy with a number of selected features obtained by the SVM classifier with full set of features, standard GWO algorithm and multi-objective GWO algorithm, the maximum accuracy calculated through the experiment rounds as we mentioned before is in Table 3. In addition, the classification accuracy of the standard GWO algorithm is determined

Table 4 Experimental results of confusion matrix parameters

Confusion matrix parameters	Type of attack	Algorithms		
		Full set of features + SVM classifier	Standard GWO algorithm + SVM classifier	Multi—objective GWO algorithm + SVM classifier
TP	DoS	2318	**4243**	4215
	Probe	950	2194	**2231**
	R2L	0	115	**124**
	U2R	200	**184**	**184**
FP	DoS	**27**	265	286
	Probe	**62**	258	258
	R2L	0	**1**	**1**
	U2R	2150	**1072**	**1072**
TN	DoS	**2125**	1887	1866
	Probe	**2090**	1894	1914
	R2L	**400**	399	399
	U2R	0	**1078**	**1078**
FN	DoS	2024	**99**	127
	Probe	1452	208	**171**
	R2L	506	391	**382**
	U2R	**0**	16	16

Bold values refer to the best result achieved during the experiments

TP true positive, *FP* false positive, *TN* true negative, *FN* false negative, *SVM* support vector machine, *GWO* grey wolf optimisation, *DoS* denial of service, *R2L* remote to local, *U2R* user to root

Table 5 Result of maximum classification accuracy with the number of selected features

Class of attack	Algorithms					
	Full set of features + SVM classifier		Standard GWO algorithm + SVM classifier		Multi—objective GWO algorithm + SVM classifier	
	Classification accuracy	No. of selected features	Classification accuracy	No. of selected features	Classification accuracy	No. of selected features
DoS	68.41%	41	**94.39%**	11	93.64%	**9**
Probe	66.75%	41	89.77%	12	**91.02%**	**10**
R2L	44.15%	41	56.73%	11	**57.73%**	**5**
U2R	8.51%	41	**53.70%**	4	**53.70%**	**2**

Bold values represent to the best result achieved during the experiments

SVM support vector machine, *GWO* grey wolf optimisation, *DoS* denial of service, *R2L* remote to local, *U2R* user to root

using Eq. (16), whereas in the multi-objective GWO algorithm computed in Eq. (17). From the Table above, it is clear that our proposed multi-objective GWO algorithm achieved superior result in the number of selected features for all attack scenarios. Furthermore, the number of selected features in R2L and U2R attack scenarios was impressive. Also, in terms of classification accuracy, the proposed GWO algorithm obtains superior result compared to the others. In spite of that the proposed GWO algorithm result in DoS attack was approximately equal to the standard GWO algorithm. Finally, these results prove that the proposed approach could efficiently select the significant set of features that achieve high classification accuracy. Figure 10 shows the types of input data to the SVM classifier.

Figure 11 displays the number of selected features achieved by the standard and proposed GWO algorithm for 20 runs. In this DoS attack scenario, the number of selected features in each run produced after executing the algorithms for 40 iterations as indicated in Table 3. From the chart, it can be seen that the performance of the proposed GWO algorithm achieved a fewer number of features during the experiment runs. Whereas the standard

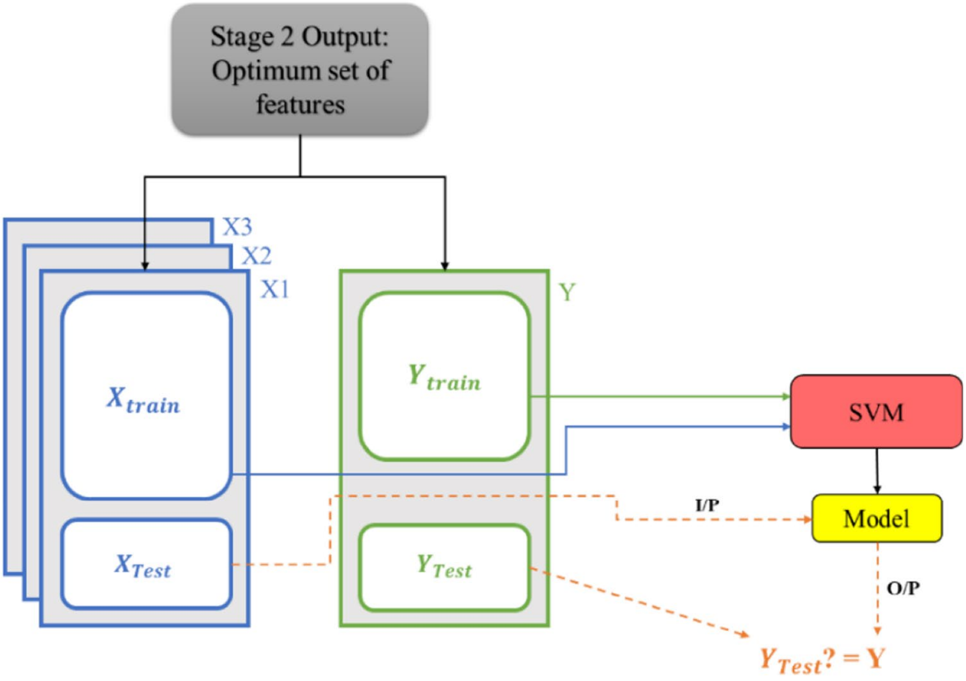

Fig. 10 Types of input data to the SVM classifier (Emary et al. 2016)

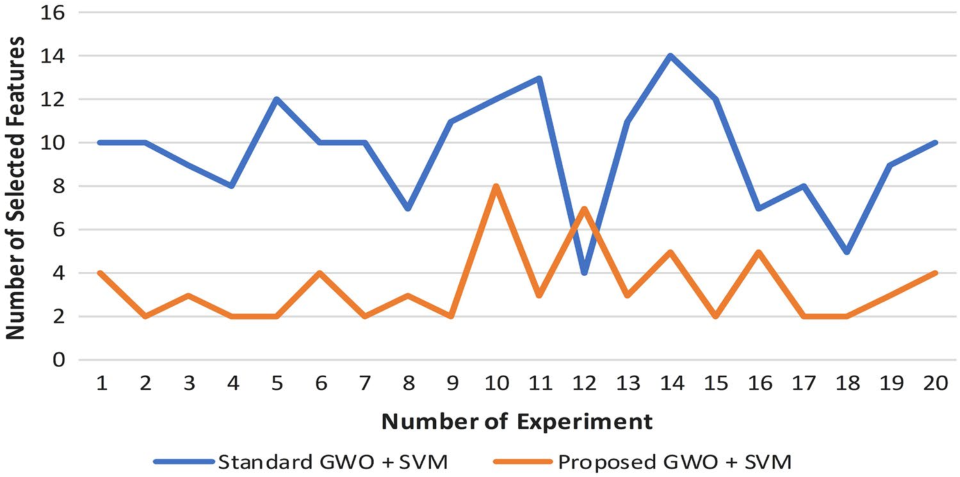

Fig. 11 Number of selected features for each run of the experiment in DoS attack scenario

Table 6 Result of average and worst accuracy with number of selected features for DoS attack scenario

Algorithms	Average classification accuracy	Selected features	Worst classification accuracy	Selected features
Standard GWO algorithm + SVM classifier	87.52	10	69.83	**20**
Multi—objective GWO algorithm + SVM classifier	**89.18**	**3**	**70.43**	23

Bold values refer to the best result achieved during the experiments

GWO algorithm obtained a higher number of features. Therefore, the multi-objective function enhanced the performance of the proposed GWO algorithm in obtaining the lowest number of selected features.

Table 6 presents the result of both average and worst accuracy with the number of selected features for DoS attack scenario which is obtained after 20 runs of the experiment. From the Table, it could be observed that the proposed GWO algorithm exceeded the highest result

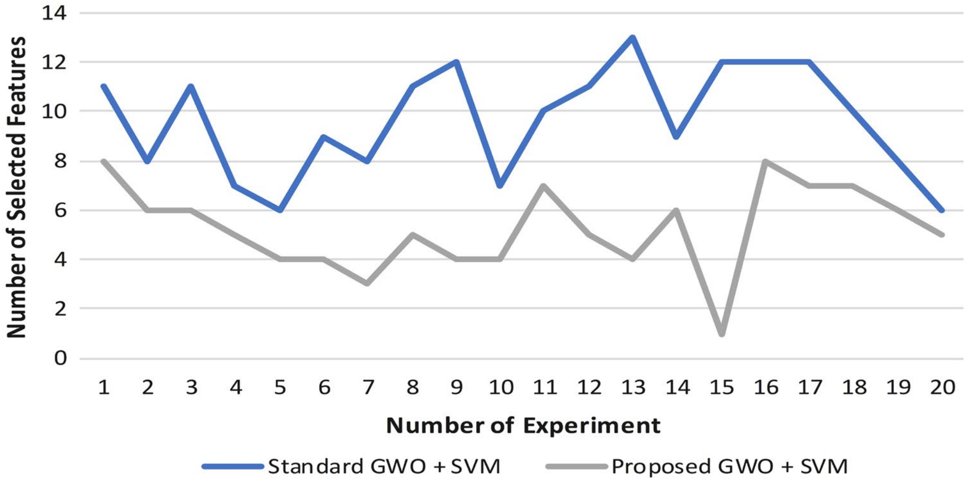

Fig. 12 Number of selected features for each run of the experiment in probe attack scenario

for average and worst accuracy compared to the standard GWO algorithm. Beside that, in terms of selected features, the proposed GWO algorithm achieved the minimum number of features with the average classification accuracy case. Whereas the standard GWO algorithm obtained the lowest number of selected features for the worst classification accuracy case.

Figure 12 presents the number of selected features obtained by the standard GWO algorithm and proposed GWO algorithm for 20 runs. In this probe attack scenario, the number of selected features in each run produced after executing the algorithms for 40 iterations as mentioned in Table 3. From the graph, it could be noted that the proposed GWO algorithm performed the lowest number of features during the experiment runs. While the standard GWO algorithm gained a higher number of features compared to the proposed GWO algorithm. Thus, it could be concluded that the proposed multi-objective function was effective to enhance the performance of the proposed GWO algorithm to obtain the lowest number of significant features.

Table 7 displays the result of both average and worst classification accuracy with the number of selected features for Probe attack scenario which is achieved after 20 runs of the experiment. From the Table, it could be noted for the best and worst classification accuracy, the proposed GWO algorithm obtained close result compared to the standard GWO algorithm. However, with respect to the selected features, the proposed GWO algorithm exceeded the standard GWO algorithm in achieving an optimum subset of features.

Figure 13 displays the number of selected features achieved by the standard GWO algorithm and proposed GWO algorithm for 20 runs. In this R2L attack scenario, the number of selected features in each run produced after executing the algorithms for 40 iterations as shown in Table 3. From the chart, it can be recognised that the proposed GWO algorithm achieved the minimum number of features during the experiment runs. While the standard GWO algorithm obtained a higher number of features.

Table 7 Result of average and worst accuracy with number of selected features for Probe attack scenario

Algorithms	Average classification accuracy	Selected Features	Worst classification Accuracy	Selected Features
Standard GWO algorithm + SVM classifier	**86.86**	10	**71.87**	22
Multi—objective GWO algorithm + SVM classifier	85.59	**5**	69.12	**13**

Bold values refer to the best result achieved during the experiments

Table 8 clarifies the result of both average and worst classification with a number of selected features for R2L attack scenario which obtained after 20 runs of the experiment. From the Table, it could be recognised that the proposed GWO algorithm achieved a higher result for best accuracy compared to the standard GWO algorithm. Beside that, for the average classification accuracy, the proposed GWO algorithm achieved a close result compared to the standard GWO algorithm. In addition, the proposed GWO algorithm was superior from the standard GWO algorithm in the number of chosen features.

Figure 14 displays the number of selected features produced by the standard GWO algorithm and proposed GWO algorithm for 20 runs. In this R2L attack scenario, the number of selected features in each run produced after executing the algorithms for 40 iterations as shown in Table 3. From the Figure, it can be seen that the proposed GWO algorithm produced the lowest number of features during the experiment runs. While the standard GWO algorithm reached a

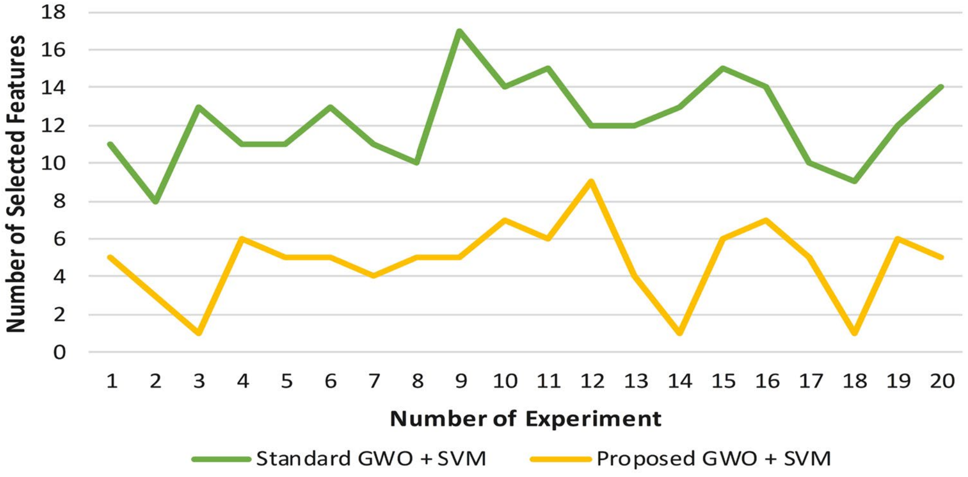

Fig. 13 Number of selected features for each run of the experiment

Table 8 Result of average and worst accuracy with number of selected features for R2L attack scenario

Algorithms	Average classification accuracy	Selected features	Worst classification accuracy	Selected features
Standard GWO algorithm + SVM classifier	**54.95**	12	**48.01**	20
Multi—objective GWO algorithm + SVM lassifier	53.18	**5**	44.15	**2**

Bold values refer to the best result achieved during the experiments

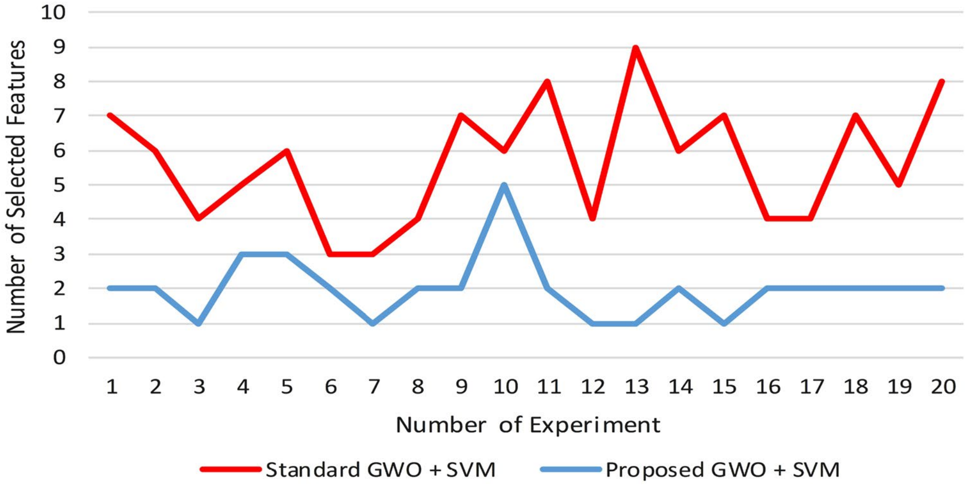

Fig. 14 Number of selected features for each run of the experiment

higher number of features compared to the proposed GWO algorithm. Therefore, the multi-objective function was successful in improving the performance of the proposed model to produce the significant subset of features.

Table 9 presents the result of both average and worst classification accuracy cases with a number of selected features for U2R attack scenario which is obtained after 20 runs of the experiment. From the Table, it could be observed that the proposed GWO algorithm exceeded the highest value for average accuracy compared to the standard GWO algorithm. However, for worst classification accuracy case, the proposed GWO algorithm produces almost the same result for accuracy compared to the standard GWO algorithm. Beside that, regarding to the number of selected features, the proposed GWO algorithm obtains the optimal subset of features experiment compared to the standard GWO algorithm.

Table 10 illustrates the comparison of our proposed approach with other existing approaches. In this comparison, all types of attacks are considered as one attack scenario. This scenario shows the efficiency of the proposed

Table 9 Result of average and worst accuracy with number of selected features for U2R attack scenario

Algorithms	Average classification accuracy	Selected features	Worst classification accuracy	Selected features
Standard GWO algorithm + SVM classifier	34.29	6	**8.59**	18
Multi—objective GWO algorithm + SVM classifier	**36.58**	**2**	8.51	**15**

Bold values refer to the best result achieved during the experiments

Table 10 Comparison of proposed approach with other existing algorithms

Author & year	Feature Selection algorithms	Dataset	Classification algorithms	Classificatiosn accuracy	No. of selected features	Analysis technique
Our proposed approach	Multi—objective GWO algorithm	NSL–KDD	SVM classifier	87.59%	**4**	Splitting dataset
Xu et al. (2017)	Hybrid cuckoo search with GWO algorithm	NSL–KDD	–	83.57%	6	Splitting dataset
Seth and Chandra (2016)	GWO algorithm	NSL–KDD	Neural network	99%	24	Cross validation
(Negandhi et al. 2019)	Gini importance	NSL–KDD	Random forest	99.80%	25	–
(Çavuşoğlu 2019)	Cfs subset eval and wrapper subset eval	NSL–KDD	Naïve bayes, random forest, J48, random tree	DoS accuracy:99.98%	25	Cross validation

Bold values indicate the best result achieved during the experiments

GWO to detect different types of attacks at the same time. In addition, we implemented and evaluated the other existing approaches using the same parameters and values that was obtained in our proposed approach. Due to the inability to use the same values and characteristics that were utilised in these approaches as a consequence of the limited hardware resources. It could be observed from the Table that the proposed multi-objective GWO algorithm obtained significant result compared to the other existing approaches in terms of classification accuracy and number of selected features. Beside that, the authors (Seth and Chandra 2016; Çavuşoğlu 2019) achieved 99% for accuracy. However, they used cross-validation method for analysis. Whereas, we used splitting dataset analysis which clarifies the real performance of the system in detecting new types of network attacks.

8 Conclusion and future research direction

This study sets out to investigate the impact of a new multi-objective function GWO algorithm to improve the performance of the IDS. The proposed multi-objective GWO was used in this study through the feature selection process to choose an optimal subset of features with high classification accuracy. In addition, 20% portion of the NSL—KDD dataset was used to test the performance of the proposed approach. The analysis technique in this study is based on data separation. This technique plays a key role in evaluating the effectiveness of the IDS system and disclosing the real performance through testing it against new types of network attacks. Furthermore, the findings conducted on the proposed approach were able to produce high classification accuracy with an optimal subset of features with different types of attack scenario. Moreover, the effectiveness and feasibility of the proposed approach were verified by comparing it with recent approaches and shows better results.

For future research directions, it is suggested that the researchers perform the proposed multi-objective function with other bio-inspired algorithms to solve different optimisation problems in addition, these algorithms could be applied to improve the performance of the SVM classifier by selecting the optimal RBF parameters. Furthermore, we will expand on this area in our future work through the implementation of new parameters for the multi-objective function such as detection rate, classification error and so on. Finally, we will apply other benchmark datasets that contain new kinds of network attacks.

References

Acharya N, Singh S (2018) An IWD-based feature selection method for intrusion detection system. Soft Comput 22:4407–4416. https://doi.org/10.1007/s00500-017-2635-2

Alamiedy TA, Anbar M, Al-Ani AK et al (2019) Review on feature selection algorithms for anomaly-based intrusion detection system. Adv Intell Syst Comput 843:605–619. https://doi.org/10.1007/978-3-319-99007-1_57

Alomari O, Othman ZA (2012) Bees algorithm for feature selection in network anomaly detection β-Hill climbing for optimization problems view project feature selection on high-dimensional data view project. Artic J Appl Sci Res 8:1748–1756

Alzubi QM, Anbar M, Alqattan ZNM et al (2019) Intrusion detection system based on a modified binary grey wolf optimisation. Neural Comput Appl 1:1–13. https://doi.org/10.1007/s00521-019-04103-1

Çavuşoğlu Ü (2019) A new hybrid approach for intrusion detection using machine learning methods. Appl Intell 49:2735–2761. https://doi.org/10.1007/s10489-018-01408-x

Cortes C (1995) Support|[ndash]|vector networks. Mach Learn 20:273–297. https://doi.org/10.1023/A:1022627411411

Dastanpour A, Ibrahim S, Mashinchi R (2014) Using genetic algorithm to supporting artificial neural network for intrusion detection system. J Commun Comput 11:1–13

Devi EMR, Suganthe RC (2017) Feature selection in intrusion detection grey wolf optimizer. Asian J Res Soc Sci Humanit 7:671. https://doi.org/10.5958/2249-7315.2017.00197.6

Dhanabal L, Shantharajah DSP (2015) A Study On NSL–KDD dataset for intrusion detection system based on classification algorithms. Int J Adv Res Comput Commun Eng 4:446–452. https://doi.org/10.17148/IJARCCE.2015.4696

Emary E, Zawbaa HM (2016) Impact of chaos functions on modern swarm optimizers. PLoS One 11:1–26. https://doi.org/10.1371/journal.pone.0158738

Emary E, Zawbaa HM, Grosan C, Hassenian AE (2015) Feature subset selection approach by gray-wolf optimization. In: Afro-European Conference for Industrial Advancement. Springer, Cham, pp 1–13

Emary E, Zawbaa HM, Hassanien AE (2016) Binary grey wolf optimization approaches for feature selection. Neurocomputing 172:371–381. https://doi.org/10.1016/j.neucom.2015.06.083

Emary E, Zawbaa HM, Hassanien AE, Parv B (2017) Multi-objective retinal vessel localization using flower pollination search algorithm with pattern search. Adv Data Anal Classif 11:611–627. https://doi.org/10.1007/s11634-016-0257-7

Emary E, Zawbaa HM, Grosan C (2018) Experienced gray wolf optimization through reinforcement learning and neural networks. IEEE Trans Neural Networks Learn Syst 29:681–694. https://doi.org/10.1109/TNNLS.2016.2634548

Garg S, Kaur K, Kumar N et al (2019) A hybrid deep learning-based model for anomaly detection in cloud datacenter networks. IEEE Trans Netw Serv Manag 16:924–935. https://doi.org/10.1109/tnsm.2019.2927886

Ghanem WAHM, Jantan A (2016) Novel multi-objective artificial bee colony optimization for wrapper based feature selection in intruction detectoin. Int J Adv Soft Comput its Appl 8:70–81

Gholipour Goodarzi B, Jazayeri H, Fateri S et al (2014) Intrusion detection system in computer network using hybrid algorithms (SVM and ABC). J Adv Comput Res 5:43–52

Gu Q, Li X, Jiang S (2019) Hybrid genetic grey wolf algorithm for large-scale global optimization. Complexity 2019:2653512. https://doi.org/10.1155/2019/2653512

Kim DS, Nguyen H-N, Ohn S-Y, Park JS (2010) Fusions of GA and SVM for anomaly detection in intrusion detection system. In: International Symposium on Neural Networks. pp 415–420

Kiran MS (2015) The continuous artificial bee colony algorithm for binary optimization. Appl Soft Comput J 33:15–23. https://doi.org/10.1016/j.asoc.2015.04.007

Kumar S, Joshi RC (2011) Design and implementation of IDS using snort, entropy and alert ranking system. In: 2011—international conference on signal processing, communication, computing and networking technologies, ICSCCN-2011. pp 264–268

Kumar V, Prakash Sangwan O (2012) Signature based intrusion detection system using SNORT. Int J Comput Appl Inf Technol I, Issue III 1:2278–7720

Kumari B, Swarnkar T (2011) Filter versus wrapper feature subset selection in large dimensionality microarray: a review. Int J Comput Sci Inf Technol 2:1048–1053

Liao HJ, Richard Lin CH, Lin YC, Tung KY (2013) Intrusion detection system: a comprehensive review. J Netw Comput Appl 36:16–24. https://doi.org/10.1016/j.jnca.2012.09.004

Liu R, Rallo R, Cohen Y (2011) Unsupervised feature selection using incremental least squares. Int J Inf Technol Decis Mak 10:967–987. https://doi.org/10.1142/s0219622011004671

Lotfi Shahreza M, Moazzami D, Moshiri B, Delavar MR (2011) Anomaly detection using a self-organizing map and particle swarm optimization. Sci Iran 18:1460–1468. https://doi.org/10.1016/j.scient.2011.08.025

Lu C, Gao L, Li X, Xiao S (2017) A hybrid multi-objective grey wolf optimizer for dynamic scheduling in a real-world welding industry. Eng Appl Artif Intell 57:61–79

Makhadmeh SN, Khader AT, Al-Betar MA, Naim S (2018) Multi-objective power scheduling problem in smart homes using grey wolf optimiser. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-018-1085-8

Mirjalili S (2014) Grey wolf optimizer MATLAB code. Adv Eng Softw 69:46–61

Negandhi P, Trivedi Y, Mangrulkar R (2019) Intrusion detection system using random forest on the NSL–KDD dataset. Emerging research in computing. Information communication and applications. Springer, Berlin, pp 519–531

Özgür A, Erdem H (2017) The impact of using large training data set KDD99 on classification accuracy. PeerJ Prepr 5:e2838v1

Rani MS, Xavier SB (2015) A hybrid intrusion detection system based on C5. 0 decision tree and one-class SVM [J]. Int J Curr Eng Technol 5:2001–2007

Roopa Devi EM, Suganthe RC (2018) Enhanced transductive support vector machine classification with grey wolf optimizer cuckoo search optimization for intrusion detection system. Concurr Comput 1–11. https://doi.org/10.1002/cpe.4999

Seth JK, Chandra S (2016) Intrusion detection based on key feature selection using binary GWO. In: 2016 3rd international conference on computing for sustainable global development (INDIACom). pp 3735–3740

Shah B, Trivedi BH (2013) Data set normalization: for anomaly detection using back propagation neural network. In: IEEE-international conference on research and development prospectus on engineering and technology (ICRDPET)

Shen J, Wang J (2011) Network intrusion detection by artificial immune system. In: IECON proceedings (industrial electronics conference). pp 4716–4720

Srivastava D, Singh R, Singh V (2019a) An intelligent gray wolf optimizer: a nature inspired technique in intrusion detection system (IDS). J Adv Robot 6:18–24

Srivastava D, Singh R, Singh V et al (2019b) Analysis of different hybrid methods for intrusion detection system. 757–764

Tavallaee M, Bagheri E, Lu W, Ghorbani AA (2009) A detailed analysis of the KDD CUP 99 data set. In: 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications. IEEE, pp 1–6

Tribak H, Delgado-Márquez BL, Rojas P et al (2012) Statistical analysis of different artificial intelligent techniques applied to intrusion detection system. In: Proceedings of 2012 international conference on multimedia computing and systems, ICMCS 2012. pp 434–440

Velliangiri S (2019) A hybrid BGWO with KPCA for intrusion detection. J Exp Theor Artif Intell 00:1–16. https://doi.org/10.1080/0952813x.2019.1647558

Vithalpura JS, Diwanji HM (2015) Analysis of fitness function in designing genetic algorithm based intrusion detection system. J Sci Res Dev 3:86–92

Wolf L, Shashua A (2005) Feature selection for unsupervised and supervised inference: the emergence of sparsity in a weighted-based approach. J Mach Learn Res 6:378–384. https://doi.org/10.1109/iccv.2003.1238369

Xingzhu W (2015) ACO and SVM selection feature weighting of network intrusion detection method. Int J Secur its Appl 9:259–270. https://doi.org/10.14257/ijsia.2015.9.4.24

Xu H, Liu X, Su J (2017) An improved grey Wolf optimizer algorithm integrated with cuckoo search. In: Proceedings of the 2017 IEEE 9th international conference on intelligent data acquisition and advanced computing systems: technology and applications, IDAACS 2017. pp 490–493

Zawbaa HM, Emary E, Grosan C, Snasel V (2018) Large-dimensionality small-instance set feature selection: a hybrid bio-inspired heuristic approach. Swarm Evol Comput 42:29–42. https://doi.org/10.1016/j.swevo.2018.02.021

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.