Journal of Big Data

RESEARCH

Open Access

Hybrid wrapper feature selection method based on genetic algorithm and extreme learning machine for intrusion detection

Elijah M. Maseno[1*†] and Zenghui Wang[2†]

†Elijah M. Maseno and Zenghui Wang contributed equally to this work.

*Correspondence: masenoelijah@gmail.com

[1] Department of Computer Science, University of South Africa, Florida, South Africa
[2] College of Science, Engineering and Technology, University of South Africa, Florida 1709, South Africa

Abstract

Intrusion detection systems play a critical role in the mitigation of cyber-attacks on the Internet of Things (IoT) environment. Due to the integration of many devices within the IoT environment, a huge amount of data is generated. The generated data sets in most cases consist of irrelevant and redundant features that affect the performance of the existing intrusion detection systems (IDS). The selection of optimal features plays a critical role in the enhancement of intrusion detection systems. This study proposes a sequential feature selection approach using an optimized extreme learning machine (ELM) with an SVM (support vector machine) classifier. The main challenge of ELM is the selection of the input parameters, which affect its performance. In this study, the genetic algorithm (GA) is used to optimize the weights of ELM to boost its performance. After the optimization, the algorithm is applied as an estimator in the sequential forward selection (wrapper technique) to select key features. The final obtained feature subset is applied for classification using SVM. The IoT_ToN network and UNSWNB15 datasets were used to test the model's performance. The performance of the model was compared with other existing state-of-the-art classifiers such as k-nearest neighbors, gradient boosting, random forest, and decision tree. The model had the best quality of the selected feature subset. The results indicate that the proposed model had a better intrusion detection performance with 99%, and 86% accuracy for IoT_ToN network dataset and UNSWNB15 datasets, respectively. The model can be used as a promising tool for enhancing the classification performance of IDS datasets.

Keywords: Extreme learning machine, Genetic algorithm, Feature selection, Wrapper methods, IoT, ToN network data set

Introduction

Due to the increased security incidences in IoT devices, researchers have increased their efforts in developing efficient and effective intrusion detection systems. In most cases, IDS acts as the first line of defense against intrusions into any network infrastructure. The IDS checks network traffic to detect any suspicious activity within a given network, preventing cyber security threats. The performance of any IDS is as good as its training dataset. The integration of many devices within the

Springer Open

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

IoT environment has triggered the generation of massive amounts of data. This data mostly consists of irrelevant and redundant features which reduce the effectiveness of machine learning algorithms and classification accuracy. In addition, these unwanted features affect the computation power of the available resources. Research shows that most of the datasets used in the training and testing of IDS suffer from the problem of high dimensionality, which reduces their performance [1–3]. To overcome these challenges, researchers have adopted feature selection as a measure of feature reduction. Feature selection aims at reducing the number of input features through the selection of the most important or relevant features in each domain [4]. This process increases the effectiveness and improves the model's classification accuracy. The main problem in feature selection is the elimination of the less important feature to keep the most relevant feature. Feature selection techniques can be classified as filter, wrapper, embedded, and hybrid techniques. The choice of feature selection techniques can have either a positive or negative impact on the model's overall performance [5]. A poor choice of feature selection techniques will reduce the detection rate of an intrusion detection system.

One of the most effective feature selection techniques is the wrapper method. The goal of this method is to select the best features that will yield the best performance of a model. This technique is further subdivided into recursive feature selection (RFE), sequential feature selection (SFS), and exhaustive search. In recursive feature selection, the goal is to eliminate the features with the lowest coefficient. The principle of operation of this technique is based on the filter selection technique in the ranking of feature importance. Exhaustive search evaluates all possible combinations of the features within a data set. It has a high computational cost and takes a lot of time to run. Despite the listed challenges, this method produces the best feature combination with high accuracy. Lastly, sequential feature selection adds or removes features from a feature subset in a greedy manner. This method is faster with less computational cost compared to an exhaustive search.

The study by [5], observed that many AI techniques have been applied in this field of feature selection to improve the performance of IDS. In this paper, we propose a new wrapper sequential feature selection based on GA and ELM. This method, known as GAELMSFS, aims at reducing the features of IoT data sets to improve the performance of intrusion detection systems. To achieve this, first, we will improve the performance of ELM through optimal weight selection using GA, and then use the optimized ELM in the sequential forward selection. The key contributions of this work are:

- Application of GA-ELM for feature selection. GA is applied to optimize ELM input parameters.
- Development of a novel hybrid intrusion detection system based on GA-ELM and SVM. The first phase of the model performs feature reduction while the second phase does classification.

- Evaluation of the model using two datasets (IoT_ToN dataset and UNSWNB15 dataset). IoT_ToN is a current dataset that reflects the current digital environment.
- Performance comparison of the model with other existing state-of-the-art classifiers.

The rest of the paper is organized as follows: "Related Work" section of the paper explains earlier works related to the current research. "The proposed classifier" section introduces the proposed classifier, "Experiments" section, the experiments, In "Results and Discussion" section, discussion of the results, "Limitation of the proposed approach" section limitations of the study, "Threats to validity" section discusses the validity of the experiment, Lastly, "Limitation of the proposed approach" section concludes the paper with a discussion of the contributions and prospects for future work.

Related work

To extract the best features [6], meta-heuristic-based sequential forward selection (MH_SFS) was proposed to reduce the number of features for anomaly classification. The researchers deployed a meta-heuristic search to solve the problem of the "nesting effect" found in the original SFS. The model outperformed both the original wrapper based SFS and the Improved Forward Floating Selection algorithms. The researchers focused only on anomaly detection. Research done by [2] proposed a wrapper feature selection for an IDS-based Pigeon Inspired Optimizer. In this research, the pigeon-inspired optimizer was used to select the most relevant features. To evaluate the model, the researchers used three datasets: KDDCUP 99, NLS-KDD, and UNSW-NB15. The researchers reported that the model reduced the data set features significantly. This reduced the model's learning rate and improved its performance in terms of TPR, FPR, accuracy, and F-score. In the future, this model can be tested using current intrusion detection system data sets that capture the state of the current digital environment. This research [7] proposed a genetic algorithm wrapper-based feature selection and a naive bayes classifier for intrusion detection in a fog environment. The model aims to reduce the redundant features and increase the detection accuracy of the model. The researchers claimed to have selected the most relevant features in the NSL-KDD dataset, which improved the performance of the model. The model had a detection accuracy of 99.73% and a false positive rate of 0.6%. The model recorded a lower F-score compared to SVM, Random Forest, and Decision tree algorithms.

In [8], the authors propose a hybrid feature selection technique. In the first phase, the authors use filter feature techniques to extract relevant features from cancerous microarray datasets. In the second phase, they apply a wrapper-based sequential forward selection method to select key features. The research applies four types of estimators: support vector machine (SVM), decision tree (DT), random forest (RF), and K-nearest neighbor (KNN). The estimators reduced the number of features significantly in all the datasets. SVM achieved the highest accuracy compared to the other three estimators. The researchers proposed that the model be tested in the future using microarray datasets of larger size. Also [9], a hybrid ensemble-filter wrapper selection technique is proposed. The wrapper stage was based on a sequential forward selection technique for feature reduction. The model was evaluated using twenty medical datasets from diverse sources. The model had better performance compared to the other fourteen algorithms. This study [10] proposes wrapper-based feature selection techniques and intrusion detection systems for Wi-Fi networks. The authors reported a detection accuracy of 99.95% using the Aegean Wi-Fi Intrusion Dataset. One major concern raised by the researchers is the time taken to develop the model. Another study [11] proposed a feature reduction technique using two wrapper-based ML algorithms, namely SVM and J48, for the classification of impersonation attacks. This work aimed to develop semi-distributed and distributed IDS to solve the limitations of centralized IDS. Centralized IDS are slow and have a single point of failure, which limits their operation. With the application of a multi-layer perceptron (MLP) classifier, distributed IDS had the lowest CPU running time of 73.52 s and the best detection accuracy of 97.80%. However, the authors noted the need for up-to-date datasets for further evaluation of the model. A similar version [12], introduced a novel wrapper feature selection based on a new feature selection metric known as CorrACC (Correlation attribute evaluation (CAE, and classifier accuracy (ACC) metric) to select the key features. Before the application of the wrapper feature selection, the authors apply the bijective soft set technique for the extraction of key features. Moreover, four machine learning classifiers were applied in their model, which enhanced the performance with an average accuracy of 95% on the BoT-IoT dataset using only 7 features out of the 39 original features. However, the study focused only on anomaly intrusion detection in an IoT environment.

The study done by [5], proposed a wrapper feature selection based on differential evolution (DE) with an extreme learning machine (ELM) classifier to improve the performance of IDS. The experiment applied the NSL-KDD intrusion detection dataset; 9 notable features were selected from the 42 original features. The selected features improved the accuracy of the model and reduced its running time. The researchers suggest the future application of more powerful classifiers to improve the performance of the model. Research done by [13], proposed the whale optimization algorithm-genetic algorithm (WOA-GA) wrapper technique for feature selection in wireless mesh networks. In this study, a support vector machine (SVM) was used to classify the selected features. The study applied CICIDS2017 and ADFA-LD datasets to evaluate the model. 77 features in CICIDS2017 were reduced to 35, and 44 features in ADFA-LD were reduced to 25. The improved WOA performed better compared with the traditional WOA in terms of detection rate and accuracy. Investigations can be done in the future to further reduce the features. In [14] proposed a new wrapper feature selection technique for IDS using a multi-objective BAT algorithm for feature selection and optimized neural networks for classification. In this study, the researchers adopt the optimized bat algorithm (EBAT) as a multi-objective and binary variation of the bat algorithm (MOB-BAT) as the base estimator for the feature selection. MLP was optimized using EBAT (EBATMLP) for classification purposes. The model was evaluated using NLS-KDD, ISCX2012, UNSW-NB15, KDD CUP 1999, and CICIDS2017 datasets. The researchers concluded that the model produced better performance compared to other existing models.

Research by [1] combined both filter and wrapper feature methods for the best feature selection. The aim was to drop the features with the smallest value in the first stage while keeping the features with high values using filter methods. The wrapper method in the second phase used evolutionary algorithms for the selection of the best feature subsets for the model to increase the classification accuracy. To select features with high coefficients, the researchers used correlation-based feature selection (CFS) and minimum redundancy and maximum relevance (mRMR) algorithms. The binary genetic algorithm (BGA) and binary particle swarm optimization (BPSO) were used in this study as wrapper models. The researchers used several data sets to evaluate the performance of the model. The researchers saw that the integration of the two types of feature selection techniques improved the model's efficiency. In a smiler context [15], introduced a feature reduction technique based on the maximum relevance minimum redundancy (mRMR) algorithm and the improved dragonfly algorithm (IDA). In this study, mRMR is used for the selection of high-ranking features to minimize redundancy. The selected subsets form the input for the wrapper algorithm (IDA) for best feature selection. The researchers observed that the hybrid model has high computational complexity compared to the wrapper algorithm. This study by [16], proposed a hybrid feature reduction technique (wrapper–filter). The genetic and PSO algorithms (HGPFS) are integrated to extract the most relevant features in the filtering phase. Genetic and PSO algorithms are used independently as estimators in the wrapper phase to further reduce the selected feature subset into optimal features. The output of these two wrapper algorithms is analyzed, and the best feature subset is selected. The authors claimed the model had satisfactory performance in feature reduction and classification accuracy. The authors did not compare the computational requirements of the model with other models.

Research by [17], presented a wrapper-based feature reduction method using a binary version of the hybrid grey wolf optimization (GWO) and particle swarm optimization (BGWOPSO). The authors deploy BGWOPSO as the wrapper estimator for the selection of best features and K-nearest neighbors for classification. The proposed model outperformed other models in accuracy, best feature selection, and computational time. The authors propose that the model be confirmed with other classification algorithms such as SVM and artificial neural networks (ANN) to see if there will be any variation in performance. The authors [18] presented a wrapper approach for best feature selection and a genetic algorithm for further reduction of the selected best features for sentiment classification. The authors concluded that this technique had the capability of reducing the feature subsets up to 61.95% without affecting the accuracy level of the model. The authors evaluated the model using only one dataset, in the future, as proposed by the authors, the model can be validated using other datasets.

Research by [19], developed a hybrid wrapper feature selection technique based on Genetic Algorithm and Permutation Importance (GA-PI). To test the classification accuracy of the selected features the study used support vector machines (SVM). Compared with other models the proposed model yielded better performance in terms of accuracy and execution time. The UNSWNB-15 dataset was used for the evaluation of the model. The main limitation of the model is that it focused on the detection of only two types of attacks. In an equivalent manner, [20] proposed a hybrid feature technique based on embedded and wrapper methods. In this research the two feature reduction techniques

Table 1 Summary of related work

Reference	Objective	Feature selection method	Dataset	Advantage	Disadvantage
[6]	Solving the problem of the "nesting effect" found in the original SFS	Wrapper	KDD Cup 99 dataset	The high detection rate of anomaly intrusion with reduced features	Focused only on anomaly detection
[2]	Designing a new technique of binarizing binarize a continuous pigeon-inspired optimizer	Wrapper	KDDCUP 99, NLS-KDD and UNSW-NB15	The model had a better learning rate and outperformed other models in terms of TPR, FPR, accuracy, and F-score	The model was evaluated using outdated datasets
[7]	To develop IDS in a fog environment	Wrapper	NSL-KDD dataset	The excellent detection rate of 99.73%	Lower F-score compared to SVM, Random Forest, and Decision tree algorithms
[8]	Developing a model to diagnose different cancer diseases from big data	Filter-base + Wrapper	Four cancerous microarray datasets (Leukemia, ovarian cancer, small round blue cell tumor, and lung cancer datasets)	The model selected the few relevant genes with high accuracy	The model was tested using microarray datasets of smaller sizes
[9]	Proposed ensemble-filter-based hybrid feature selection model for disease detection	Filter-base + Wrapper	Twenty benchmark medical datasets	The model was evaluated using four classifiers namely, Naïve Bayes, Support Vector Machine with Radial Basis Function, Random Forest, and k-Nearest Neighbor	The study used only two performance metrics namely, accuracy and AUROC. The authors propose other metrics to be used in future work
[10]	Investigation of various feature selection techniques	Wrapper	Aegean Wi-Fi Intrusion Dataset (AWID)	The model reported a high detection accuracy of up to 99.95%	It takes longer to build the model
[11]	To develop semi-distributed and distributed IDS	Wrapper	AWID	Using a multi-layer perceptron (MLP) classifier, distributed IDS had the lowest CPU running time of 73.52 s and the best detection accuracy of 97.80%	The authors noted the need for up-to-date datasets for further evaluation of the model
[12]	To select the best features for exact classification of smart IoT anomaly and intrusion traffic identification	Wrapper	Bot-IoT dataset	The research reduced the 39 original features to 7 without affecting the model's accuracy	The research focused only on Bot-IoT attacks
[5]	To develop a feature selection technique based on a differential evaluation algorithm	Wrapper	NSL-KDD dataset	The selected features improved the accuracy and the running time of the model	The results are still not optimistic
[13]	To develop a wrapper-based feature selection method based on a modified whale optimization algorithm (WOA)	Wrapper	CICIDS2017 and ADFA-LD standard datasets	The improved WOA performed better compared with the traditional WOA in terms of detection rate and accuracy	Investigations can be done in the future to further reduce the features

Table 1 (continued)

Reference	Objective	Feature selection method	Dataset	Advantage	Disadvantage
[14]	To improve the performance of IDS through the development of a two-phase framework to increase the detection rate as well reducing the false alarm rate	Wrapper	NLS-KDD, ISCX2012, UNSW-NB15, KDD CUP 1999, and CICIDS2017 datasets	Introduction of a new metaheuristic algorithm (MOBBAT), a binary version of the BAT algorithm	The researchers did not consider computational cost as a metric measure
[1]	Selection of key features using an evolutionary algorithm	Filter and Wrapper	Wine, Ada Sonar, Sylva, Madelon and Gina datasets	Evaluation of the model was done using several datasets to drop any bias	The model was evaluated using only one metric measure
[15]	To develop a novel feature selection algorithm named hybrid improved dragonfly algorithm (HIDA)	Filter and Wrapper	10 gene expression datasets and 8 UCI data sets	HIDA has an excellent performance in resolving imbalanced classification problems	High computational complexity compared to the wrapper algorithm
[16]	Combination of genetic algorithms (GA) and particle swarm optimization (PSO) for best feature selection	Filter and Wrapper	Lung, Hill-Valley, Gas 6, Musk 1, Madelon, and Isolet 5	The model had a superior performance in feature reduction as well as classification accuracy	The authors did not compare the computational requirement of the model with other models
[17]	Implementation of binary version of the hybrid grey wolf optimization (GWO) and particle swarm optimization (PSO) for feature selection	Wrapper	18 standard benchmark datasets from UCI	The proposed model outperformed other models in accuracy, best features selection, and the computational time	The model used on one classification algorithm (KNN)
[18]	Implementation of a new feature selection named GAWA	Wrapper	Tweeter datasets	The technique had the capability of reducing the feature subsets up to 61.95% without affecting the accuracy level of the model	The model was evaluated using only one dataset, in the future as proposed by the others the model can be confirmed using other datasets
[19]	Implementation of novel feature selection based on GA and PI	Wrapper	UNSWNB 15 dataset	The proposed model yielded better performance in terms of accuracy and execution time	The main limitation of the model is that it focused on the detection of only two types of attacks
[20]	Combination of embedded and wrapper for feature selection	Embedded and Wrapper	NSL-KDD dataset	Use of two classification algorithms to test the selected feature subset	The model used the NSL-KDD dataset, a classical dataset that does not capture the current intrusion threats

we combined to leverage their advantages. SVM and Naive Bayesian were used as base classifiers in this research. The model used NSL-KDD dataset a classical dataset that does not capture the current intrusion threats Table 1.

Feature selection technics

Feature reduction techniques can be broadly classified into two categories: unsupervised and supervised. Supervised techniques are further classified as filter, wrapper, embedded, and hybrid techniques [8]. Supervised techniques select features by their relation to the target variable [4]. Filter techniques statistically evaluate the performance of each feature against the output variable and keep scores as the basis of performance. These techniques are computationally inexpensive but have low accuracy [4, 8]. On the other hand, wrapper techniques evaluate the performance of different combinations of the input variables to figure out the best combination for the model. This technique is computationally expensive but has high accuracy. The hybrid technique is an integration of the filter and wrapper techniques [4, 8]. Embedded feature selection techniques have their own inbuilt mechanisms for selecting key features as part of the learning [21]. Hybrid feature selection techniques take advantage of both the filter and wrapper methods for feature selection. In most cases, hybrid feature selection techniques have two or more phases. Filter techniques are usually applied in the first phase to reduce the features according to their importance, and in the second phase, these techniques apply wrapper techniques for best feature selection to improve classification accuracy [22]. Most of these techniques suffer from high computational capacity compared to their counterparts.

This study focused on wrapper feature selection techniques that offer high accuracy compared to filter feature selection techniques. The major categories of wrapper feature selection techniques are sequential feature selection and recursive feature elimination. According to [6], the traditional sequential selection algorithms are sequential forward selection (SFS) and sequential backward selection (SBS). Sequential feature selection automatically selects key features from the original dataset (X) to reduce the dataset to an optimal feature subset (K) [23]. According to [6, 23], SFS eliminates or adds one feature at a time until the predefined feature subset is reached. The predefined features subset is smaller than the original features ($X > K$).

Sequential forward selection (SFS) begins with an empty set and adds one feature at a time to optimize the performance of the classifier. This is repeated until the required features are generated [24]. The authors [24] evaluated the performance of three wrapper feature selection techniques, namely, sequential forward selection, sequential backward selection, and recursive feature elimination, in the selection of an optimal feature subset. The researchers observed that sequential forward selection could select the optimal feature subset that increased the classification performance of ANN compared to the other two techniques. In this study, we propose a novel SFS based on a hybrid method (GA-ELM). The aim is to improve the model's accuracy, reduce computational costs, and improve its running time.

The proposed classifier

Motivation

As discussed earlier, the IoT environment consists of the integration of diverse types of devices, leading to enormous data generation. The main challenge of the generated data set is that it suffers from the problem of high dimensionality in that it consists of irrelevant features that affect the accuracy and performance of intrusion detection systems. The selection of key features is a key element for the optimal performance of any machine learning algorithm. The choice of dataset in the development of an intrusion detection system plays a key role in the overall performance of an intrusion detection model [25]. To evaluate the proposed model, the researchers adopted two publicly available datasets namely the IoT_ToN dataset and UNSWNB15 dataset. These two datasets capture the current attacks in the cyber world. That captures current intrusions. Like other forms of datasets, the two datasets are enormous in nature, consisting of irrelevant features that affect the effectiveness of an intrusion detection system. This study proposes the development of effective IDS based on GA-ELM and SVM. The proposed technique aims to improve the effectiveness and accuracy of IDS through feature reduction in the two selected datasets using GA-ELM.

The first phase of the model is data preprocessing or cleaning. This phase's aim is to ensure the data set is suitable for the proposed model. To solve the problem of class imbalance, which is found in many datasets [26], this study adopted a hybrid technique known as SMOTE-Tomek Links which combines both oversampling and under sampling. In this technique, SMOTE is used to increase the minority class through oversampling. When the desired proportion of the minority class is achieved, Tomek Links is applied to aliment data samples from the majority class that are identified to be near the minority class. This process is applied to the training data set only. The second phase of the model is the feature selection phase. This study integrates GA and ELM to develop a more efficient and effective wrapper feature reduction technique. This technique has been used in many fields with immense success but has not yet been tested as a feature-reduction technique in an IoT environment. GA aims to optimize the performance of ELM. To optimize ELM performance, GA is used in selecting input weights. The first task in this process is setting the number of hidden neurons and activation function. After the setup, the initial input weights are randomly generated. The ELM will be trained to produce the first GA population. GA will be trained to produce the best input weights for ELM through evolution principles. This process aims to improve the performance of ELM by selecting the best input weights. Research shows that the performance of ELM is highly affected by randomly generated input weights. The optimized ELM will be used in this study as the base classifier in the sequential forward selection. The main goal of the optimized ELM is to select the best features and drop the irrelevant features. The third and final phase of the model is the feature classification. The best feature subset from the second phase forms the input of the third phase. This study adopted SVM as the base classifier due to its effectiveness in classification.

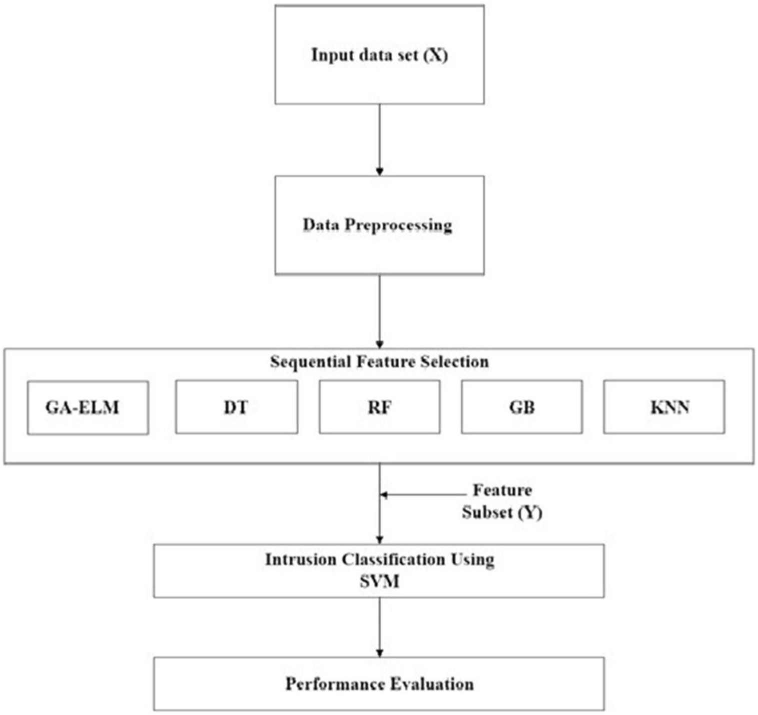

Fig. 1 Proposed IDS Model

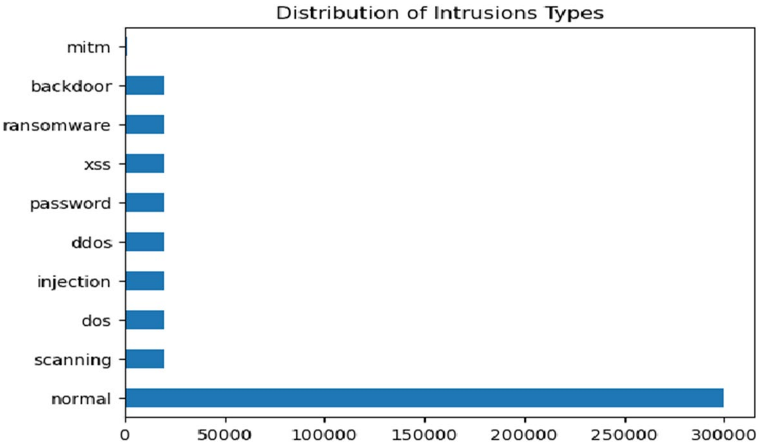

Fig. 2 IoT_ToN training–testing attack distributions

Methodology

The proposed IoT intrusion detection system consists of three stages: data preprocessing/cleaning, sequential forward selection, and classification, as shown in Fig. 1.

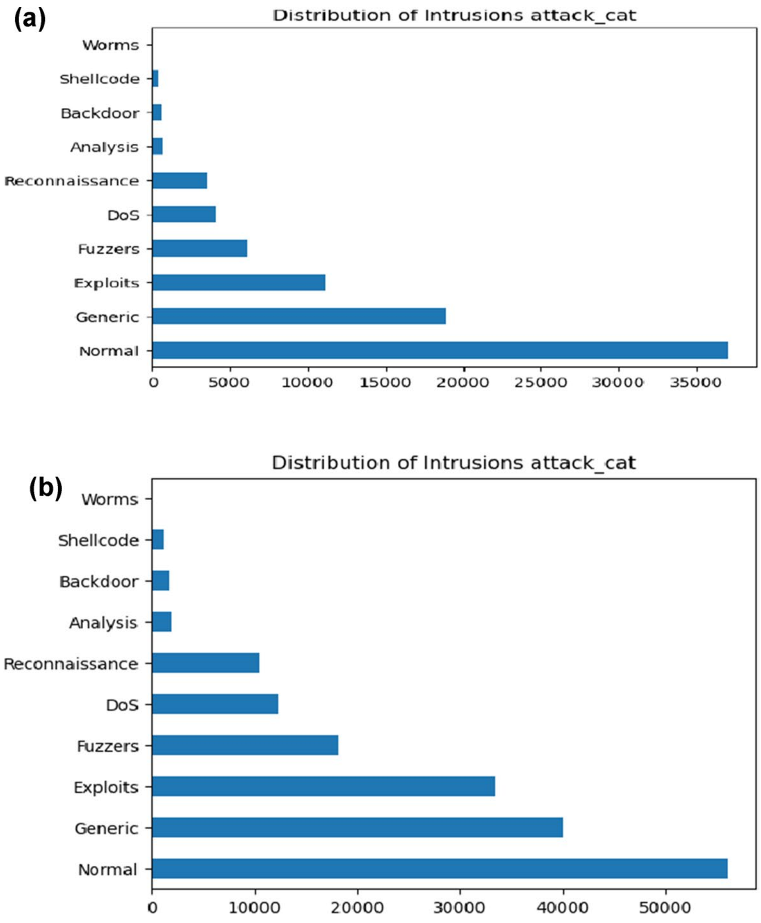

Fig. 3 **a** UNSWNB15 training dataset attack distributions. **b** UNSWNB15 testing dataset attack distributions

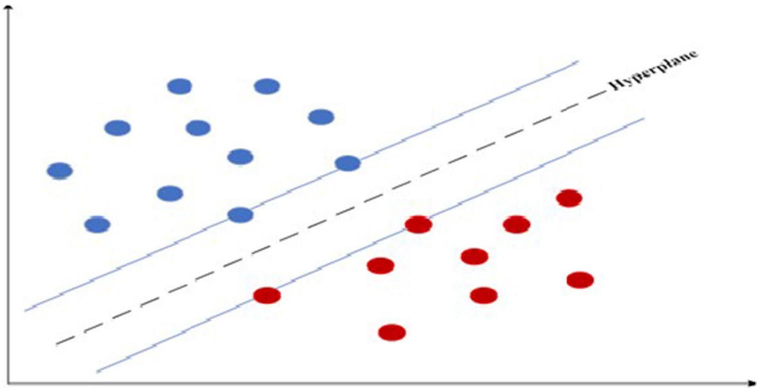

Fig. 4 Support Vector Machine

Datasets

To evaluate the proposed model, the researchers selected two public datasets namely the IoT_ToN dataset and the UNSWNB15 dataset. The aim was to eliminate any bias in the evaluation of the model.

IoT_ToN network dataset

The IoT_ToN network dataset was used in this study because it has many advantages compared to the existing datasets. This data set is heterogeneous capturing data from four sources namely network traffic, windows, Linux operating systems, and IoT/IIoT services. In addition, this data set captures current IoT networks and complex cyber malicious activities making it suitable for the evaluation of new intrusion detection systems. The data set consists of nine attack distributions as shown in Fig. 2.

UNSWNB15 dataset

This dataset has been widely applied in the evaluation of intrusion detection systems in many studies. It was first published in 2015 [27]. The dataset consists of nine types of attacks and several normal traffic as shown in Figs. 3, 4 representing class distribution for both training and testing datasets, respectively. In addition, each attack consists of 44 features and the class label.

Data preprocessing phase

Data preprocessing is an important task in machine learning. The raw data is transformed during this phase into a form suitable for a particular prediction model. In most cases, raw data cannot be used directly. The major tasks to be performed during this stage are data cleaning and data transformation. Data cleaning is the process of finding and correcting any errors within the dataset. During data cleaning, columns that have the same value or no variance and duplicate rows of data are removed. In addition, missing values are marked and replaced using statistics or a learned model. Data transformation is the process of changing the scale or distribution of variables in the raw data. Data may have one of a few types, such as numeric or categorical, with subtypes for each, such as integer and real-valued floating-point values for numeric, and nominal, ordinal, and boolean for categorical. We may wish to convert a numeric variable to an ordinal variable in a process called discretization. Alternatively, we may encode a categorical variable as integers or Boolean variables, which are needed on most classification tasks.

Sequential forward selection (SFS)

In the second phase, the sequential forward selection is applied using different models for the selection of the optimal feature subset. Sequential feature selection highly depends on the base classifier's performance to add or remove features [21, 22]. Some well-known classifiers include support vector machine (SVM), decision tree (DT), random forest (RF), and K-nearest neighbor (KNN) classifier. This study proposes a hybrid classifier based on GA-ELM be adopted as a base estimator in SFS. SFS to select optimal features starts with an empty feature subset and adds a feature on each iteration. After the addition of the feature, the algorithm evaluates the accuracy of the classifier based on the selected feature subset to decide which feature should be selected. The feature with the best accuracy is selected to be part of the relevant feature subset, The selection process runs until the terminal condition is met. The selected feature subset is a dimensional optimal feature subset whose classification performance is the first highest.

Let complete dataset: $M = \{P_1, P_2 \ldots . Pk\}$

Let the new subset: $D = \{\}$

```
For y iteration do
$d_{add =}$ best F (S + d), where d ∈ M–D
D = D + $d_{add}$
y = y + 1
```

As mentioned previously the effectiveness of SFS is determined by the base classifier. This study will investigate the performance of the GA-ELM classifier for optimal feature selection using the SFS-wrapper selection technique.

ELM

When ELM was first proposed by [28], the aim was to improve the performance of feedforward neural networks, which are slow by nature. ELM is reported to have a fast-learning speed and better generalization performance. The main limitation of this algorithm is its randomization in initial parameters (weights and biases) selection. One way to overcome this challenge is to optimize the input parameters selection using the metaheuristic algorithms as proposed by [26]. ELM can be explained as follows, according to [28]: Given N distinct training set $(\mathbf{X}_i, \mathbf{t}_i)$, where $\mathbf{X}_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T \in \mathbf{R}^n and \mathbf{t}_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T \in \mathbf{R}^m$. Z and $g(x)$ represents number of hidden nodes and activation function, respectively. ELM can be implemented by randomly assigning the parameters of the hidden nodes (ω, b), computing the hidden layer output matrix ($\mathbf{H}$) and the output weights ($\boldsymbol{\beta}$). Using N samples, our target output $\mathbf{T}$ can be obtained using the equation below:

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T} \tag{1}$$

where

$$\left(\omega_1, \cdots, \omega_Z,\ b_1, \cdots, b_Z,\ x_1, \cdots, x_N \right) \tag{2}$$

$$\begin{bmatrix} g(\omega_1, x_1 + b_1) & \cdots & g(\omega_Z, x_1 + b_Z) \\ \vdots & \cdots & \vdots \\ g(\omega_1, x_N + b_1) & \cdots & g(\omega_Z, x_N, b_N) \end{bmatrix}_{NZ} \tag{3}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1^T \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_Z^T \end{bmatrix}_{Zm} \tag{4}$$

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T \\ \mathbf{t}_2 \\ \vdots \\ \mathbf{t}_N^T \end{bmatrix}_{Nm} \tag{5}$$

To compute the weights connecting the hidden layer and the output layer represented by $\boldsymbol{\beta}$, the least-squares technique is applied to minimize the error between the target and the output.

$$\hat{\beta} = \mathbf{H}^{\dagger}\mathbf{T} \tag{6}$$

where $\mathbf{H}^{\dagger}$ is the Moore–Penrose generalized inverse of matrix **H,** and **T** is the target.

Genetic algorithm (GA)

GA is a well-known evolutionary algorithm that has been applied in many fields as a search space and optimization algorithm [29, 30]. As part of the evolutionary algorithm, GA solves complex problems through the evolutionary mechanism. Most of the existing optimization techniques suffer from the problem of local minima and lack the capabilities of finding global solutions. However, GA, due to its randomness, can find global solutions. Over time GA has proven to be effective and efficient in finding global optimum solutions. Due to the mentioned advantages, GA was adopted to optimize ELM in this study. The standard GA procedure is described below:

i. The initial population is randomly generated. The population is made up of individuals with weights and biases.

ii. The fitness value of each individual is calculated using Eq. (7)

$$\sum \left(\text{population\_i} \, * \, \text{equation\_inputs\_j}\right) \tag{7}$$

iii. Research by [28] compared three GA selection criteria namely roulette, K-tournament and random. The aim was to study their effects on the overall performance of the OGA–ELM model. According to the study, there was no significant difference in classification accuracy. K-tournament achieved the highest accuracy of 100%, while random criteria achieved the lowest accuracy of 99.38%. With these results, this study used random criteria, which is simple and easy to implement. Using random selection, 2 parents are randomly selected from the initial population.

iv. One-point crossover method is applied on the selected parents. The cut point is the center of the two genes. The tail of the two genes is switched to form new offsprings.

v. The chromosome to be subjected to mutation is randomly selected. The aim is to alter the genetic composition of the chromosome, hence creating diversity within the population, which improves GA performance. This work uses uniform mutation. The uniform mutation works to substitute the selected gene's value with a uniform random value chosen from the gene's user-specified upper and lower bounds (for the input-hidden layer weights [− 1, 1] while for the hidden layer biases [0, 1].

Following the selection, crossover, and mutation processes, the generation of a new population is achieved. This new population is used in the subsequent iteration, following which the process is repeated until the maximum number of generations (50) is reached. The GA approach will seek to enhance the ELM, as discussed in the next subsection.

GA-ELM

In this study, GA was used to generate the optimal weights to improve the performance of ELM. The optimized ELM algorithm follows the steps below:

1. Start
2. Split the dataset into training and test datasets
3. Set the number of hidden neurons and the activation function
4. Randomly initialize the input weights and biases
5. Train the ELM and extract the output values
6. ELM output value forms the initial input value of GA
7. Initialize the fitness score
8. Select the weight with best fitness score
9. Crossover weights
10. Mutate weights
11. Test if the termination condition has been achieved
12. If "NO" repeat step 7–10
13. If "YES" pick the optimal weights and train the ELM
14. Test the accuracy of the model.
15. END

The optimized ELM (GA-ELM) will be adopted as the base classifier in this study to select the optimal features.

Classification

To test the classification accuracy of the selected optimal feature, this study adopted an SVM classifier. A support vector machine is a type of supervised machine learning algorithm that is used to perform classification and regression [31, 32]. SVM was first proposed by Cortes and Vapnik [33] in 1995. By using a hyperplane, they perform classification. Hyperplane maximizes the edge between two classes. The vectors that characterize the hyperplane are called support vectors. In this algorithm, each piece of data is considered a point in n-dimensional space. The value of each piece of data is taken as a coordinate on the plane, and then classification is performed by finding an optimal hyperplane that divides the two classes. The main advantage of using SVMs is that they are good at generalization and can overcome the curse of dimensionality [31]. Figure 4 shows the support vector and optimal hyperplane that perform classification.

With the training sample of (x_i, y_i), $i=\{1, 2, 3\ldots M\}$, where M is the total number of predictors, y_i defines the class of the training data as either -1 or $+1$ $(=\{y_i=\{-1,+1\})$, $x \in R.^n$ where n is the number of features in each sample. The standard equation of the hyperplane in SVM is shown in equation [8]

$$W^T x + b = 0 \tag{8}$$

To predict the class of new observation (Y) use the equations below.

$$+1 \; if \; W^T x + b \geq 0 \tag{9}$$

$$-1\, if\ W^{T}x + b \leq 0 \tag{10}$$

The goal of SVM in training is to maximize the width or distance of the margins between the two hyperplanes i.e. equation [9] and [10]. The distance (d) between the two line is given by:

$$d = \frac{2}{||w||} \tag{11}$$

To achieve the max distance, we can minimize the denominator (||w||)

$$\max\frac{1}{||w||} \leftrightarrow \min||w|| \leftrightarrow min\frac{1}{2}||w||^{2} \tag{12}$$

To calculate the extremes of a function within a given constraints, langrage multipliers are applied.

$$L = \frac{1}{2}||w||^{2} - \sum \alpha_{i}[y_{i}(\overline{x}_{i}\overline{w} + b) - 1] \tag{13}$$

With the partial derivatives of *w* and *b* being zero, Quadratic programming (QP) problem is transformed into:

$$L = \sum \alpha_{i} - \frac{1}{2}\sum_{i}\sum_{j}\alpha_{i}\alpha_{j}y_{i}y_{j}(x_{i.}x_{j}) \tag{14}$$

With the above equation [14], the optimization depends on the dot product of the samples x_i and x_j.

For non-liner classification tasks, the vectors can be transformed into a new sample space for classification.

$$K\left(x_{i}x_{j}\right) = \varnothing(\overline{x}_{i}).\varnothing(\overline{x}_{j}) \tag{15}$$

where K is a kernel function which provides the dot products of the vectors in another space.

Experiments

This section presents experimental implementation and significant results evaluations and discussions.

Dataset

To evaluate the model, the study used the publicly available datasets referred to as the TON_IoT network dataset [34] and UNSWNB15 dataset [27]. According to the researchers, TON_IoT network dataset was developed for the evaluation of AI-based security solutions. This data was adopted because it reflects actual IoT cyber activities, which makes the data reliable when investigating the performance of new IoT IDS. The initial inspection of the dataset proved that the dataset had missing data for some feature columns, as shown in Fig. 5, and categorical data. The researchers applied the Ordinal Encoder scheme to convert the categorical dataset to a numerical dataset. Secondly, the

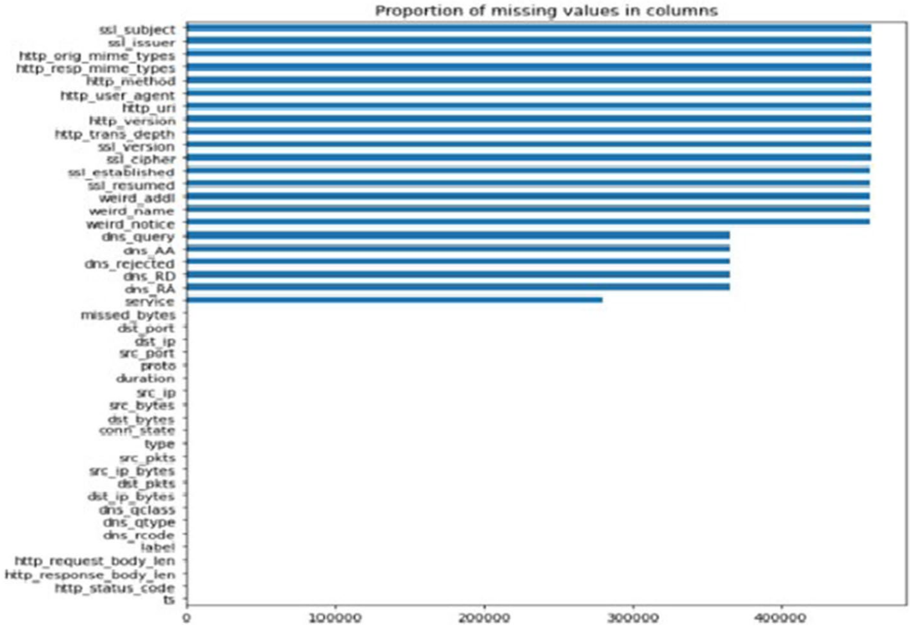

Fig. 5 The proportion of missing values in the column

researchers replaced all the missing values with a constant figure. Research done by [35] recommended that both IP (Internet Protocol) addresses and ports be dropped in the construction of the new machine-learning models. In addition, the *type* attribute was dropped because it is only applicable in a multi-classification model [35]. The researchers performed data splitting into training and testing datasets before normalizing the data set to avoid data leakage, as suggested [4]. UNSWNB15 dataset was subjected on the same process of preprocessing to clean and prepare the data for evaluation. However, the major difference in the preprocess stage is that there was no need for data splitting with UNSWNB15 dataset because the authors had already separated the dataset into training and testing sets.

Parameter settings

For selection of optimal features, the SSF wrapper-based method GA-ELM was used. GA-ELM acted as the base classifier in this study. The GA-ELM's parameters were set by these values (hidden_units = 100, activation_function = 'relu', x = X_train_minmax, y = y_train, weight_type = "GA_weight", C = 0.1, alg_type = "clf"). The parameters of the proposed approach are listed below in summary:

- hidden_units = 100: Number of hidden neurons.
- activation_function = 'relu': The activation function to be used. The rectified linear unit (ReLU) function is used in this case.
- x = X_train_minmax: The input features to be used. Each feature is scaled to the range [0,1].
- y = y_train: The output targets to be used.
- weight_type = "GA_weight": Optimzed input weights from GA.
- C = 0.1: Regularization parameter.

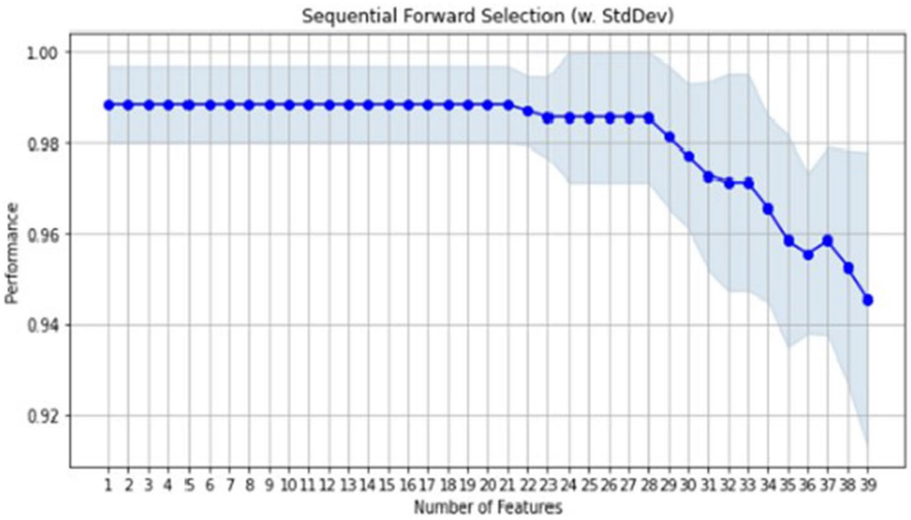

Fig. 6 Optimum features for GA-ELM

Table 2 Selected features

ID	Name	Description
0	duration	Packet connection time
9	dns_qtype	Value which specifies the DNS (Domain Name System) query types
11	http_request_body_len	The original size of the HTTP data from the client
12	http_response_body_len	The original size of the HTTP data from the server
13	http_status_code	HTTP server status
14	proto	Transport layer protocols of flow connections
15	service	Dynamically detected protocols, such as DNS, HTTP and SSL (Secure Socket Layer)
18	dns_AA	Authoritative answers of DNS, where T denotes server is authoritative for query
23	ssl_cipher	SSL cipher suite which the server chose
24	ssl_resumed	SSL flag shows the session that can be used to start new connections, where T refers to the SSL connection is initiated
25	ssl_established	SSL flag indicates establishing connections between two parties, where T refers to establishing the connection
26	ssl_subject	Subject of the X.509 cert offered by the server
27	ssl_issuer	Trusted owner/originator of SLL and digital certificate (certificate authority)
28	http_trans_depth	Pipelined depth into the HTTP connection
29	http_method	HTTP request methods such as GET, POST and HEAD
30	http_uri	URIs used in the HTTP request
31	http_version	The HTTP versions utilized such as V1.1
32	http_user_agent	Values of the User- Agent header in the HTTP protocol
33	http_orig_mime_types	Ordered vectors of mime types from source system in the HTTP protocol
34	http_resp_mime_types	Ordered vectors of mime types from destination system in the HTTP protocol
35	weird_name	Names of anomalies/violations related to protocols that happened

- alg_type = "clf": "clf" refers to a classification problem.

After setting the parameters the optimized ELM was used in the sequential forward selection. To perform SFS, the study used *SequentialFeatureSelector* function from the library of *mlxtend.* The optimized ELM classifier is employed to select the optimal parameters as an estimator for the SequentialFeatureSelector function. The researchers used k-fold cross-validation on the feature selector to avoid the issue of overfitting. The basic principle of operation in this approach is to subdivide the training data into small K subsets. The model is trained using these K subsets. The performance of the model is the average value of the scores obtained from the subsets. In this work the researchers used $k=10$, to subdivide the training data into tenfold of the same size. Finally, a fit function can pass all training and testing datasets.

The estimator requires 21 features to register optimum classification as shown in Fig. 6 below. The generated features (K) were used for testing the classification accuracy of the model. Table 2 is the summary of the selected feature subset.

Feature classification

As earlier mentioned, kernel functions can be used in SVM to transform input vectors with complex boundaries into a new space dimension for classification, this study opted to apply Gaussian radial basis function (RBF). According to a study done by [36], RBF are effective in separation of samples with sophisticated patterns. The RBF kernel used in this study, is shown below:

$$K\left(x_i x_j\right) = e^{-\gamma ||x_i - x_j||^2}, \gamma > 0 \tag{16}$$

The SVM classifier was trained using the training data, which was generated using the informative features selected by the optimized GA-ELM. The trained SVM was then assessed with the testing data. The metrics used to test the performance of classifier were accuracy, precision, and recall. These three metrics are derived from five parameters namely: the true positive (TP), false positive (FP), false negative (FN), and true negative (TN) rates:

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

$$Precision = \frac{(TP)}{(TP + FP)}$$

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

Table 3 The selected set of features from The IoT_ToN network dataset

Method	Selected features	Feature indexes
GA-ELM	21	(0,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28)
DT	35	(0,1,2,3,4,5,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,27,28,29,30,32,34,35,36,37,38)
RF	25	(0,4,9,10,11,12,13,14,15,19,20,22,23,24,25,26,27,28,29,30,31,32,33,34,35)
GB	34	(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,32,33,34,35,36,37,38)
KNN	21	(0,4,11,12,13,14,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36)

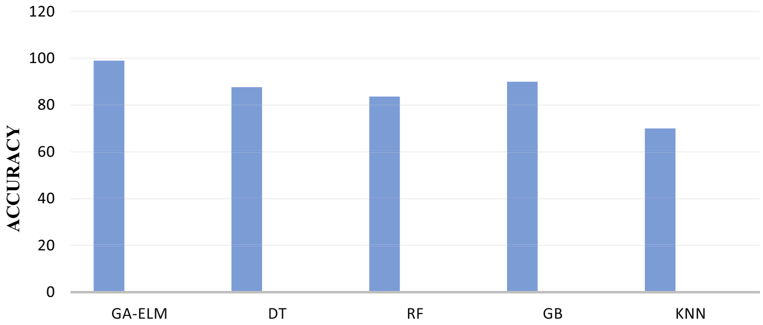

Fig. 7 Accuracy results of the algorithms

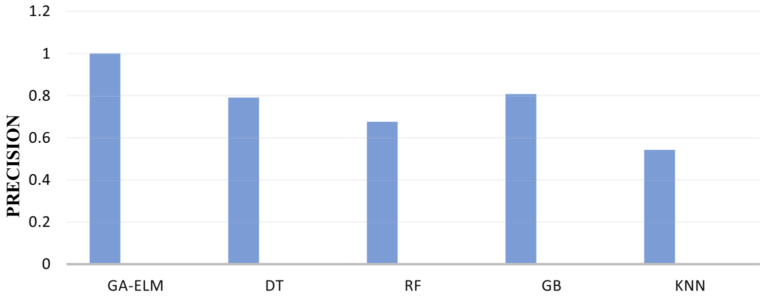

Fig. 8 Precision results of the algorithms

Here,

TP is the correctly detected real positive data;

TN is the correctly detected real negative data;

FP is the data for positives wrongly detected as negatives;

and

FN is the data for negatives wrongly detected as positives.

Results and discussion

The researchers evaluated the proposed model in terms of accuracy, precision, and recall. SVM was adopted as the base classifier in this study. The performance of the model was compared with the other four state-of-the-art algorithms.

The IoT_ToN network dataset evaluation

Table 3 presents all the state-of-the-art algorithms used to compare with the proposed algorithm and the number of features selected for the IoT_ToN network data set. The proposed model together with the KNN reduced the number of features from 45 to 21, which was a superior performance compared to DT, RF and GB. This was followed closely by RF which selected the 25 best features. The third in ranking in terms of feature reduction was GB which managed to select 34 best feature subsets. However, DT performed poorly on feature reduction, with 35 sub-features.

Figure 7 illustrates the result of the accuracy of the evaluated algorithms. To test the accuracy of the models, the above selected sub-features formed the input to the base classier. Each bar in Fig. 4, stands for the score of each algorithm. The results show that

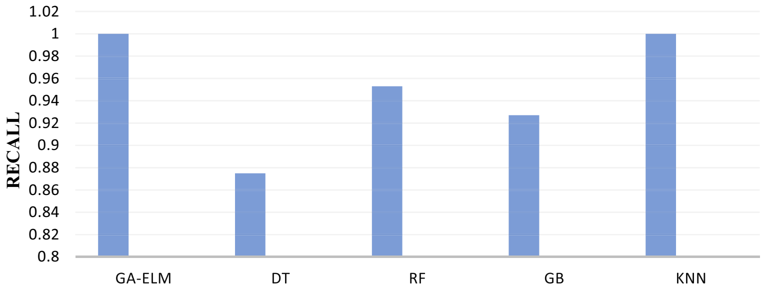

Fig. 9 Recall results of the algorithms

Table 4 The selected set of features from the UNSWNB15 dataset

Method	Selected features	Feature indexes
GA-ELM	21	(6, 7, 11, 16, 19, 20, 21, 22, 25, 28, 29, 30, 31, 33, 34, 35, 38, 39, 40, 41, 42)
DT	20	(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 39)
RF	13	(6, 7, 16, 19, 20, 21, 22, 25, 31, 33, 34, 38, 39)
GB	30	(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 32, 33, 34, 35, 36, 39)
KNN	11	(6, 7, 20, 21, 22, 25, 33, 34, 35, 38, 39)

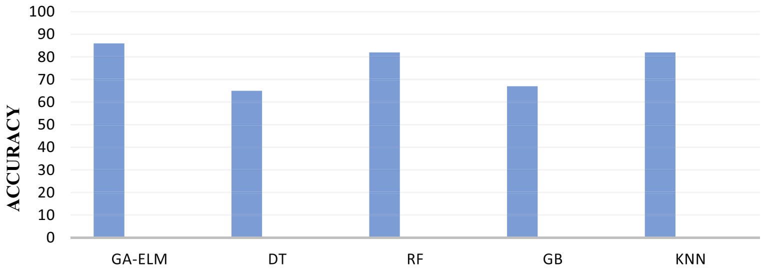

Fig. 10 Accuracy results of the algorithms

the proposed algorithm scored an accuracy of 99% which was the highest compared to all other evaluated algorithms. GB achieved the second-highest accuracy score of 90%. DT, RF and KNN scored accuracy of 87.67%, 83.67 and 70% respectively. Figure 8 illustrates the result of the precision for the evaluated algorithms. Like the earlier results on accuracy, both GA-ELM and GB achieved the highest scores in precision compared to the other tested algorithms, 1 and 0.808, respectively. DT, RF and KNN registered precision scores of 0.791, 0.676 and 0.543, respectively. Figure 9 illustrates the result of the recall for the evaluated algorithms. On this measure, GA-ELM and KNN achieved the highest score compared to the other evaluated algorithms. Both models registered recall score of 1. RF, GB, and DT registered recall scores of 0.953, 0f.927 and 0.874, respectiv ely.

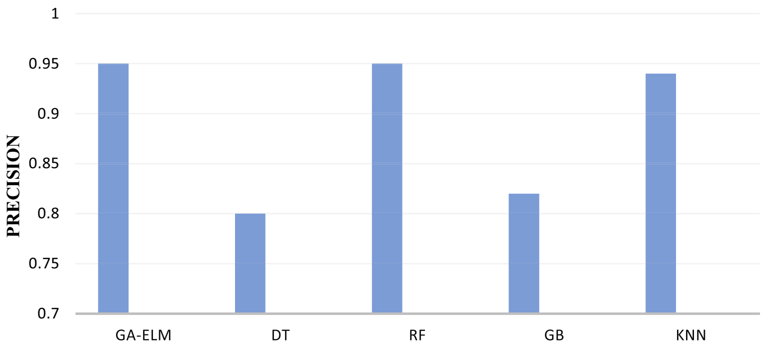

Fig. 11 Precision results of the algorithms

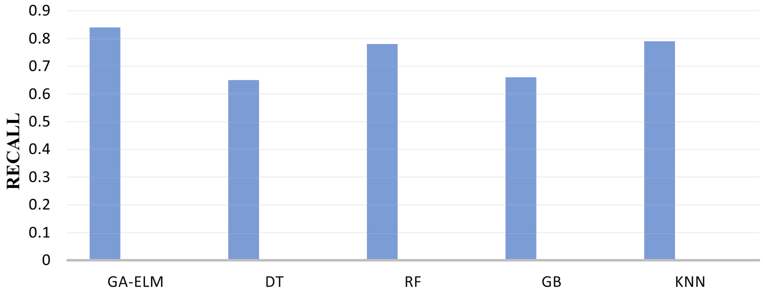

Fig. 12 Recall results of the algorithms

The UNSWNB15 evaluation

Table 4 presents all the selected features subsets from the UNSWNB15 dataset using the proposed model and the other state-of-the-art algorithms. The proposed model reduced the number of features from 44 to 21, while DT, RF GB and KNN reduced the features to 20, 13, 30 and 11, respectively. However, GB performed poorly on feature reduction on this dataset.

Figure 10 illustrates the result of the accuracy of the evaluated algorithms. To test the accuracy of the models, the above selected sub-features formed the input to the base classier. Each bar in Fig. 4, stands for the score of each algorithm based on the selected features subset. The results show that the proposed algorithm scored an accuracy of 86% which was the highest compared to all other evaluated algorithms. KNN and RF achieved the second-highest accuracy score of 82%. DT and GB scored accuracy of 65, and 67% respectively. Figure 11 illustrates the result of the precision for the evaluated algorithms. Both GA-ELM and RF achieved the highest precision score of 0.95. This was followed closely by KNN that registered precision scores of 0.94. However, DT and GB had a precision score of 0.80 and 0.82, respectively. Figure 12 illustrates the result of the recall for the evaluated algorithms. On this measure, GA-ELM achieved the highest recall score of 0.84 compared to the other evaluated algorithms. KNN, RF, GB, and DT registered recall scores of 0.79, 0.78, 0.66 and 0.65, respectively.

Limitation of the proposed approach

The main disadvantage of the Sequential forward selection (SFS) is that once a feature has been selected there is no chance of removing the feature even if it does not add value to the model with the addition of new features. In addition, selecting a feature per its performance does not grantee better model performance.

Threats to validity

To evaluate the models, this work used IoT_ToN network and UNSWNB15 datasets which are public datasets and highly recommended for testing IDS due to their reflection of the current cyber threats. The results obtained in this study may not be replicated in a real-world environment due to open-source assessment and classification tools.

Conclusion

In this paper, a hybrid wrapper features selection method based on a genetic algorithm and an extreme learning machine for intrusion detection in an IoT environment was proposed. The aim was to reduce the number of features in the IoT data set while improving the performance of IoT intrusion detection systems. The model was evaluated using the IoT_ToN network data set, which captures most of the attacks in the IoT environment. To avoid any biase the model was further evaluated using UNSWNB15 dataset. The researchers used three metrics for the evaluation of the model: precision, accuracy, and recall. The results of the model were compared to other state-of-the-art classifiers, namely the random forest classifier, the decision tree classifier, the gradient boosting classifier, and the k-nearest neighbors classifier. SVM was adopted as the base classifier to classify the selected feature subset.

The proposed GA-ELM feature selection algorithm reduced the number of features in the IoT_ToN datasets from 49 to only 21. Using the UNSWNB15 dataset, the model had a relatively reliable performance, on this dataset it reduced the 44 features to 21 feature subsets. However, KNN managed to reduce the features from 44 to 11 which was an outstanding performance compared to the reset. The model achieved high accuracy, precision, and recall compared to the other evaluated algorithms. This performance showed that feature reduction should not only focus on the elimination of features but also on the quality of the features selected.

In the future, we propose that the performance of the proposed model be evaluated with other types of data sets. Also, the model can be deployed in a real IoT environment for intrusion detection.

Acknowledgements
Not applicable.

Author contributions
EMM and ZW proposed the adopted algorithms and obtained the datasets for the research. ZW's rich experience was instrumental in improving our work. EMM did the literature survey of the paper and wrote the main manuscript. All authors contributed to the editing and proofreading. All authors read and approved the final manuscript.

Funding
This research is partially supported by the South African National Research Foundation (Grant Nos. 132797 and 137951), the South African National Research Foundation incentive grant (No. 114911), and the South African Eskom Tertiary Education Support Programme.

Availability of data and materials

The datasets generated during and/or analyzed during the current study are available in the Ton_iot datasets (2020) repository, https://www.unsw.adfa.edu.au/unsw-canberra-cyber/cyberse curity/ADFA-ton-iot-Datasets/ and https://research.unsw.edu.au/projects/unsw-nb15-dataset.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

The authors give their consent for publication.

Competing interests

The author declares no competing interests.

Received: 27 April 2023 Accepted: 19 January 2024
Published online: 01 February 2024

References

1. Kawamura A, Chakraborty B. A hybrid approach for optimal feature subset selection with evolutionary algorithms. Proceedings-2017 IEEE 8th International conference on awareness science and technology, ICAST 2017, 2018-Janua(iCAST), 2017. https://doi.org/10.1109/ICAwST.2017.8256521
2. Alazzam H, Sharieh A, Sabri KE. A feature selection algorithm for intrusion detection system based on Pigeon Inspired Optimizer. Expert Syst Applicat. 2020. https://doi.org/10.1016/j.eswa.2020.113249.
3. Wu Q, Ma Z, Fan J, Xu G, Shen Y. A feature selection method based on hybrid improved binary quantum particle swarm optimization. IEEE Access. 2019;7:80588–601. https://doi.org/10.1109/ACCESS.2019.2919956.
4. Brownlee J. Data preparation for machine learning (and J. H. Sarah Martin and my technical editors Michael Sanderson and Arun Koshy, Andrei Cheremskoy (ed.); v1.1). 2020.
5. Al-Yaseen WL, Idrees AK, Almasoudy FH. Wrapper feature selection method based differential evolution and extreme learning machine for intrusion detection system. Pattern Recogn. 2022;132: 108912. https://doi.org/10.1016/j.patcog.2022.108912.
6. Liu Y, Xu Z, Yang J, Wang L, Song C, Chen K. A novel meta-heuristic-based sequential forward feature selection approach for anomaly detection systems. Proceedings-2016 International conference on network and information systems for computers, ICNISC 2016, . 2017. https://doi.org/10.1109/ICNISC.2016.20
7. Onah JO, Abdulhamid SM, Abdullahi M, Hassan IH, Al-Ghusham A. Genetic algorithm based feature selection and Naïve Bayes for anomaly detection in fog computing environment. Mach Learn Applicat. 2021;6(September): 100156. https://doi.org/10.1016/j.mlwa.2021.100156.
8. Elemam T, Elshrkawey M. A highly discriminative hybrid feature selection algorithm for cancer diagnosis. Sci World J. 2022. https://doi.org/10.1155/2022/1056490.
9. Singh N, Singh P. A hybrid ensemble-filter wrapper feature selection approach for medical data classification. Chemom Intell Lab Syst. 2021;217(July): 104396. https://doi.org/10.1016/j.chemolab.2021.104396.
10. Rahman MA, Asyhari AT, Wen OW, Ajra H, Ahmed Y, Anwar F. Effective combining of feature selection techniques for machine learning enabled IoT intrusion detection. Multimed Tools Appl. 2021;80(20):31381–99. https://doi.org/10.1007/s11042-021-10567-y.
11. Rahman MA, Asyhari AT, Leong LS, Satrya GB, Hai Tao M, Zolkipli MF. Scalable machine learning-based intrusion detection system for IoT-enabled smart cities. Sustain Cities Soc. 2020;61(January): 102324. https://doi.org/10.1016/j.scs.2020.102324.
12. Shafiq M, Tian Z, Kashif A, Du X, Guizani M. IoT malicious traffic identification using wrapper-based feature selection mechanisms. Comput Secur. 2020. https://doi.org/10.1016/j.cose.2020.101863.
13. Vijayanand R, Devaraj D. A novel feature selection method using whale optimization algorithm and genetic operators for intrusion detection system in wireless mesh network. IEEE Access. 2020;8:56847–54. https://doi.org/10.1109/ACCESS.2020.2978035.
14. Ghanem WALIHM, Abduljabbar S, Ghaleb A, Jantan A, Nasser AB, Abdulla S, Saleh M, Saad AHY, Member S, Omolara AE. Cyber intrusion detection system based on a multiobjective binary bat algorithm for feature selection and enhanced bat algorithm for parameter optimization in neural networks. IEEE Access. 2022;10(July):76318–39. https://doi.org/10.1109/ACCESS.2022.3192472.
15. Cui X, Li Y, Fan J, Wang T, Zheng Y. A hybrid improved dragonfly algorithm for feature selection. IEEE Access. 2020;8:155619–29. https://doi.org/10.1109/ACCESS.2020.3012838.
16. Moslehi F, Haeri A. A novel hybrid wrapper–filter approach based on genetic algorithm, particle swarm optimization for feature subset selection. J Ambient Intell Humaniz Comput. 2020;11(3):1105–27. https://doi.org/10.1007/s12652-019-01364-5.
17. Al-Tashi Q, Abdul Kadir SJ, Rais HM, Mirjalili S, Alhussian H. Binary optimization using hybrid grey wolf optimization for feature selection. IEEE Access. 2019;7:39496–508. https://doi.org/10.1109/ACCESS.2019.2906757.
18. Rasool A, Tao R, Kamyab M, Hayat S. GAWA-A feature selection method for hybrid sentiment classification. IEEE Access. 2020;8:191850–61. https://doi.org/10.1109/ACCESS.2020.3030642.
19. Sowmya, Anita TM. An intelligent hybrid GA-PI feature selection technique for network intrusion detection systems. Int J Intell Syst Appl Eng. 2023: 11(7s); 718–731. https://www.ijisae.org/index.php/IJISAE/article/view/3010.

20. Santhi V, Priyadharshini J, Swetha M, Dhanavandhana K. A Hybrid feature extraction method with machine learning for detecting the presence of network attacks. 2023 International conference on intelligent systems for communication, IoT and security (ICISCoIS), coimbatore, India. 2023. https://doi.org/10.1109/ICISCoIS56541.2023.10100339.
21. Abiodun EO, Alabdulatif A, Abiodun OI, Alawida M, Alabdulatif A, Alkhawaldeh RS. A systematic review of emerging feature selection optimization methods for optimal text classification: the present state and prospective opportunities. Neural Comput Appl. 2021;33(22):15119. https://doi.org/10.1007/s00521-021-06561-y.
22. Ben Brahim A, Limam M. A hybrid feature selection method based on instance learning and cooperative subset search. Pattern Recogn Lett. 2016;69:28–34. https://doi.org/10.1016/J.PATREC.2015.10.005.
23. Raschka S. Sequential Feature Selector: The popular forward and backward feature selection approaches (including floating variants)–mlxtend. 2022. http://rasbt.github.io/mlxtend/user_guide/feature_selection/SequentialFeatureSelector/#:~:text=RFE%20is%20computationally%20less%20complex,defined%20classifier%2Fregression%20performance%20metric.
24. Fahmiin MA, Lim TH. Evaluating the effectiveness of wrapper feature selection methods with Artificial neural network Classifier for diabetes prediction. 2020. https://doi.org/10.1007/978-3-030-43215-7.
25. Kanna PR, Santhi P. Unified deep learning approach for efficient intrusion detection system using integrated spatial-temporal features. Knowl-Based Syst. 2021;226: 107132. https://doi.org/10.1016/j.knosys.2021.107132.
26. Kanna PR, Santhi P. Hybrid intrusion detection using mapreduce based black widow optimized convolutional long short-term memory neural networks. Expert Syst Appl. 2022;194: 116545. https://doi.org/10.1016/j.eswa.2022.116545.
27. Moustafa N, Slay J. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set, 2015 Military Communications and Information Systems Conference (MilCIS), Canberra, ACT, Australia. 2015. https://doi.org/10.1109/MilCIS.2015.7348942.
28. Huang GB, Zhu QY, Siew CK. Extreme learning machine: theory and applications. Neurocomputing. 2006;70(1–3):489–501. https://doi.org/10.1016/j.neucom.2005.12.126.
29. Eshtay M, Faris H, Obeid N. Metaheuristic-based extreme learning machines: a review of design formulations and applications. Int J Mach Learn Cybern. 2019;10(6):1543–61. https://doi.org/10.1007/s13042-018-0833-6.
30. Albadr MAA, Tiun S, Ayob M, AL-Dhief FT. Spoken language identification based on optimised genetic algorithm–extreme learning machine approach. Int J Speech Technol. 2019;22(3):711–27. https://doi.org/10.1007/s10772-019-09621-w.
31. Kumari A, Mehta AK. A hybrid intrusion detection system based on decision tree and support vector machine. 2020 IEEE 5th International Conference on Computing Communication and Automation, ICCCA 2020. https://doi.org/10.1109/ICCCA49541.2020.9250753
32. Chen C, Song L, Bo C, Shuo W. A support vector machine with particle swarm optimization grey wolf optimizer for network intrusion detection. proceedings-2021 International Conference on Big Data Analysis and Computer Science, BDACS. 2021. https://doi.org/10.1109/BDACS53596.2021.00051
33. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995. https://doi.org/10.1007/BF00994018.
34. Moustafa N. The TON_IoT Datasets. 2020. https://www.unsw.adfa.edu.au/unsw-canberra-cyber/cyberse%0Acurity/ADFA-ton-iot-Datasets/
35. Moustafa N. A new distributed architecture for evaluating AI-based security systems at the edge: network TON_IoT datasets. Sustain Cities Soc. 2021;72(April): 102994. https://doi.org/10.1016/j.scs.2021.102994.
36. Ahmad I, Basheri M, Iqbal MJ, Rahim A. Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection. IEEE Access. 2018;6:33789–95. https://doi.org/10.1109/ACCESS.2018.2841987.

Publisher's Note

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