T-SNERF: A novel high accuracy machine learning approach for Intrusion Detection Systems

Mohamed Hammad | Nabil Hewahi | Wael Elmedany

CORRESPONDENCE
Mohamed Hammad, College of Information Technology, University of Bahrain, Sakhir, Manama, Bahrain
Email: mohammad90@gmail.com

Abstract
In the last few decades, Intrusion Detection System (IDS), in particular, machine learning-based anomaly detection, has gained importance over Signature Detection Systems (SDSs) in the novel attacks detection. Herein, a novel approach called T-Distributed Stochastic Neighbour Embedding and Random Forest Algorithm (T-SNERF) is presented for the classification of cyber-attacks. The approach consists of three different steps. First, the examination of feature correlations is provided. Second, the T-Distributed Stochastic Neighbour Embedding (T-SNE) data dimensional reduction technique is used. Third, Random Forest (RF) technique is utilised to evaluate the complications in the accuracy and False-Positive Rate (FPR). The proposed approach has been tested on various well-known datasets, namely, UNSW-NB15, CICIDS-2017, and phishing datasets. The proposed novel approach achieved significant results compared with existing approaches, achieving 100% accuracy, and 0% FPR for the UNSW-NB15 dataset, and achieving high accuracy rates, up to 99.7878%, and 99.7044%, for CICIDS-2017 and Phishing datasets respectively.

1 | INTRODUCTION
Network security is gaining importance with computer networks for the protection against network attacks. Traditional security tools are used in many organisations such as firewalls, techniques of anti-spam, anti-viruses' software, etc. However, more complex attacks cannot be identified by these tools. Network Intrusion Detection Systems (NIDSs) is the second line of defence that emerged for tracking the activity of the network and detecting disruptive events. NIDSs are now seen as effective defensive mechanisms that can defend against most disruptive threats and attacks [1].

This research aimed to study network anomaly detection systems using multiple machine learning techniques. The two main problems with existing state-of-the-art techniques are the reduction of false-negative and false-positive rates, and real-time processing to detect network intrusion. Some studies focussed on rule-based identification for complex network attacks where pre-designed processing rules are used for attack detection [2]; however, the rule-based expert approaches are less effective for very large datasets. Also, a recent post-Internet network architecture called wireless sensor networks (WSN) is increasingly utilised. The technical requirements of intrusion detection become more complicated by its addition to conventional and mobile networks with the diversification of intrusion methods available for malicious agents that should be taken into consideration [3]. WSNs are exposed to many security attacks easier than wired networks due to its limited distribution, and the multi-hop communication, bandwidth, and battery power used therein. Therefore, it is very important to design an efficient IDS for WSNs. The development of new attacks (e.g., black hole, sink hole, abnormal transmission, and packet dropping attacks) has taken place due to this new environment. By mechanism, attack detection is classified into traditional signature-based detection, an anomaly in behaviour detection and hybrid detection mechanisms. WSN protection is mostly done using anomaly in behaviour-based detection [4].

However, the detection of intrusion is inadequate for the protection against any attack for many reasons. First of all, its learning capacity is limited, it is based on the summation of features from raw data and then transform them into vectors to be inputs for the classifier. When the complexity of network structure increases, learning performance declines. Second, using an intrusion detection method with only one or two levels of information is not enough for the recognition of any additional attack types. Third, there is a similarity between network intrusion

© 2021 The Authors. IET Information Security published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.
in real network datasets and those in normal datasets which limits classifiers to be able to classify them. Next, the unpredictability of behaviours for intrusion actions causes costly mistakes in IDS for intrusion detection. Therefore, finding an efficient detection method for intrusion becomes a necessity [5]. Finally, traditional approaches for intrusion detection become inconvenient because of large-scale data with high-dimensional structures generated by the variety of network types [6].

For the determination of analysis aspects, a part of the UNSW-NB15 dataset is split into training and testing sets. The evaluation of the complexity of these sets is the goal of the three aspects. First, the Kolmogorov-Smirnov Test [7] is applied to define and compare the distribution of the training and testing sets; the asymmetry of the features is measured using skewness [8]; and the features flatness is estimated using kurtosis [8]. In case an approximate similarity exists between these statistics and the features of the training and testing sets, results’ reliability can be reached. Second, correlations of the features are calculated in two ways: (1) for the first perspective, Pearson’s Correlation Coefficient (PCC) [9] is calculated. For the second perspective, the Gain Ratio (GR) method [10] is utilised. Third, for the measurement of accuracy and False Alarm Rates (FARs), five techniques exist, namely, Naive Bayes (NB) [11], Decision Tree (DT) [12], Artificial Neural Network (ANN) [12,13], Logistic Regression (LR) [13], and Expectation-Maximisation (EM) Clustering [14] are applied for the training and testing sets. Furthermore, a comparison between the results of the UNSWNB15 dataset and the KDD99 dataset [15] takes place to recognise UNSWNB15 dataset capability in the evaluation of old and recent classifiers.

1.1 | IDS and Limitations of traditional antivirus

The commonly used signature-based firewalls and antivirus are considered reactive and cannot recognise the first time seen attacks. The signature, usually a hash, is generated manually by malware analysts for the detection of a specific part of malware. The signature is saved in the malware database. During any new scan, antivirus software accesses the database of signature. In the beginning, malware analysts were efficient because of the simplicity of attacks (viruses, trojans, and worms) but with the appearance of automated malware polymorphism and obfuscation, this method becomes incompetent. Zero-day attacks (malicious files targeting vulnerabilities that are previously undisclosed) could not be recognized by Signature-based detection. The development of alternatives to supplement traditional signature-based detection becomes a necessity for the creation of a robust antivirus product.

IDS is the active process of detecting unauthorized activity in system and network. It can be any hardware, software, or the combination between them for surveillance system or network of systems in case of any suspicious activity. Catching perpetrators in the act before any deterioration of resources. An IDS is the guarantee to protect the system from any attack. The main responsibilities of network activity are: Monitoring network activity, audit network, and system configurations for vulnerabilities, analysing data are the main responsibilities for IDS. It is an indispensable component in the security toolbox providing three functions: monitoring hosts, detecting the behaviours, and generating an alert. IDS can be considered as firewall functionality even though there is a difference between both. For the protection of the information flow and the prevention of intrusions, a firewall is safeguarded, whereas detecting the status of the network (under attack or not) or determining if the firewall security worked well are the main responsibilities for IDS. The combination of firewalls and IDS improves the security of the network.

The act of ensuring that no malicious activities occurred in a network system is called ‘Network security’. It can treat different types of network attacks: spoofing, denial of service, intrusion detection. It is also called ‘communication security’, also deals with the protection of the transmission of information [16]. Some tools are available for network security like Wireshark [17], Snort [18], and TippingPoint [19]. Hosts participating in a network of computers is highly preferable for the possession of host security. In this case, the applications must be equipped with application security.

Most existing security systems require a network administrator for the surveillance of one or more tools listed above. In case of any security breach, the network administrator with the help of other analytical tools starts taking all the required measures. Rule-based network security tools include Snort, TippingPoint, and their variations. These tools come with security rules already developed in them. Adding more rules is the responsibility of the network administrator, and the manufacturers can take this responsibility via updates.

1.2 | Machine learning-based IDSs

Machine learning is a data-dependent process, where the first step is to understand the data. In this part, several ways of machine learning application to IDS type were introduced using multiple datasets. This variety represents different types of attacks including the behaviours of host and network. System logs reflect the behaviours of the host, and network traffic reflects network behaviours. There are several types of attacks, each one has a particular pattern. Thus, it is important to select suitable data sources for the detection of various attacks according to the characteristics of the attack. Sending many packets in a quite brief time is one of the key features of a Deny Of Service (DOS) attack; therefore the data flow is convenient for the detection of a DOS attack.

1.3 | Flow-based attack detection

The most common source of data for IDSs is packets grouped in a period that exists inflow data. The detection of flow attacks has two advantages: (1) Flow reflects the entire environment of a network and allows the detection of most attacks, in particular DOS and Probe. (2) Pre-processing the flow without
parsing for packets or restructuring of the session is easy. However, the packet content is ignored by flow; that is why it gives an unsatisfactory detection result for U2R and R2L. When flow features are extracted, cached packets are important; thus, it involves some hysteresis. Feature engineering and deep learning methods are included in a flow-based attack. However, a strong heterogeneous flow can cause weak detection effects. Usually, a grouping of traffic is used to solve this issue.

1.4 Contribution to the field

The contribution herein was to develop high accuracy machine learning model for intrusion detection using latest network traffic datasets, implement different machine learning techniques to detect most advanced and rare network attacks, and implement data dimensionality reduction machine learning technique to increase the individual classification accuracy of the network attacks. The proposed approach using only relative features in the datasets and maintaining at the same time low dimensionality which can in its turn reduce the training time. Also, to study network anomaly detection systems using multiple machine learning techniques, summarised by the following steps: the use of a combination of T-SNE algorithm, the selection of the most important subset of features by using Correlation-based Feature Selection (CFS) and finally using Random Forest as a classification technique. Weka [20] and R are the programming language for this work.

1.5 Paper structure

Herein, Section 2 is a presentation of related works; Section 3 provides a concise overview about all of the three used datasets; Section 4 describes the methodology background; Section 5 presents a clear overview of the proposed process and gives a description about the workings of the algorithm; Section 6 is for the evaluation criteria, results of the experiment, and results and discussion; Section 7 lists some of the challenges of using machine learning algorithms in IDS; Section 8 concludes the work and presents the future work.

2 RELATED WORK

Our work is addressing the accuracy of the current intrusion detection techniques and proposing a novel technique using multiple machine learning techniques to enhance the accuracy and detection of rare attacks, while minimising the consumption of the resources by reducing the feature set and implementing data dimensionality reduction technique to minimise the training and testing time of the machine learning model. The proposed technique is clearly scoring higher detection rates than any available models.

In [21], a fusion model that integrates rank based chi-square feature selection with multi class SVM optimised by kernel scale achieved an accuracy of 97.44% using NSL-KDD dataset. Other novel method of fusion of PCA and optimised SVM in [22] has been proposed, the obtained accuracy was 99.78% using KDDCup99 dataset.

Several models and approaches for network intrusion detection have been proposed over the past few years, based on traditional machine learning. Examples include the Support Vector Machine (SVM) [23,24], RF [23–25], Neurotree [26], KNN, Random Effects Logistic Regression (RELR) [35], Principal Component Analysis (PCA) filtering [36], and others reported in [37–39]. In [40], multi-class classification has been used, the results on the UNSW-NB15 dataset are as follow: accuracy of 83.47% and FAR of 6.57%. When the RF classifier is applied on the UNSW-NB15 dataset by Jianarthan and Zargari [41] for the detection and the classification of intrusion attacks using five selected features, the results are as follows: an accuracy of 81.6175% and FAR of 4.4%.

For the classification of the UNSW-NB15 dataset [42], multiple machine learning techniques are used herein. The development of the UNSW-NB15 dataset using IXIA Perfect Storm is a robust network dataset that represents recent network traffic scenarios and several low footprint intrusions [43]. New studies have proven that traditional datasets should be replaced by new benchmark datasets, which means that recent day network traffic is not represented anymore by these datasets [43,44]. Some of the old datasets are KDD98 [45], KDDCUP99 [15], and NSL-KDD [46]; hence UNSW-NB15, CICIDS-2017 [47], and Phishing datasets are used herein.

Deep learning is a machine learning branch, recently utilised to detect network intrusion. In the previous research in IDS, many deep learning algorithms were used for the unsupervised feature such as Deep Belief Networks (DBNs), restricted Boltzmann machines (RBMs), auto-encoder, and deep neural networks (DNNs). For example, a new method proposed by Erfani et al. [48] is DBNs with a linear uni-class SVM for the detection of intrusion. Its application is on multiple benchmark datasets. Likewise, for learning compressed features from a specific set of features that don’t exist in the payloads of packets, a discriminative RBM (DRBM) method is introduced by Fiore et al. [49]. For classifying behaviours, compressed features are the input of a soft-max classifier. DNNs are a deep learning method introduced by Javadi et al. [50] for anomaly detection. The results proved the efficiency of a deep learning model for detecting a flow-based anomaly in software-defined networks (SDNs). For the NSL-KDD dataset, a deep learning model is proposed by Tang et al. [51], which is using Self Taught Learning (STL) to build a network IDS. The results demonstrated that deep learning outperforms past research studies by its performance and accuracy. Wang [51] introduced a deep learning method for network traffic detection from raw data based on stacked auto-encoder. It achieved remarkably high performance. Also, a deep learning method built on recurrent neural networks (RNNs) for detecting the intrusion is proposed by Yin et al. [52]. The application of RNNs was on the NSL-KDD dataset. It proved
that deep learning methods are more efficient than traditional machine learning classification algorithms for IDS. The four hidden layers deep learning method based on RBM and DBN to reduce the number of features is proposed by Alrawashdeh and Purdy [53]. The update of the DBN weight is in a fine-tuning phase while Logistic Regression is used for classification. The accuracy of the model after its application on the KDD99 dataset is 97.9% and a false alarm rate of 0.5%. This accuracy is insufficient to build a robust model for detecting network intrusion [54].

A deep learning approach based on a non-symmetric deep auto-encoder (NDAE) is presented by Shone et al. [55] for intrusion detection. The KDD99 dataset was used for the application using RF for classification. The accuracy of the results is 97.85%. However, this method is inefficient for detecting complicated attacks because of the high value of false alarm rate 2.15%. More recently, a model using PCA and a Gaussian-binary restricted Boltzmann machine (GRBM) is proposed by Nguyen et al. [56] for the detection of cyber-attacks in a mobile cloud environment. However, the unclear testing process of this method does not allow comparative benchmarking.

Few works exist that study the application of classification techniques on UNSW-NB15. The statistical analysis presented by Moustafa and Slay [28] is for the observations and attributes in UNSW-NB15 [42]. For the calculation of accuracy and FARs, five different classifiers are used. A group of techniques for detecting intrusion, AdaBoost is proposed by Moustafa et al. [57]. DT, NB, and ANN machine learning techniques are used on UNSW-NB15 and NIMS botnet datasets. A high accuracy of 99.54% and low false-positive rates of 1.38% are provided by the ensemble technique. The techniques of machine learning are used on UNSW-NB15 using identifiers of flow [42] for the efficient detection of botnets and their tracks.

The UNSW-NB15 dataset is used for NIDSs by Mogal et al. [58]. Herein, Central Points of attribute values with the a priori algorithm is used for pre-processing. NB and logistic regression are machine learning classifiers. The results are improved after pre-processing. The research of Moustafa and Slay [28] concentrated on the classification of the different types of attacks that were captured in the UNSW-NB15 dataset [42]. Herein, it is focused on the identification of the important features used in UNSW-NB15 datasets by using multiple machine learning techniques such as NB, EM, and association rule mining. However, the accuracy value for these techniques was not so high for the rare attacks (e.g., 20% for BackDoor).

Some works are made by Cannady [59,60] on network classification. He proved that neural networks are suitable solutions for a specific problem when they are trained using selective sets of the training dataset. Since the model does not have the ability to work with continuous data, the protection of the system is a necessity to take off-line the data whenever training the model is needed and run it to the updated set of selective data.

A classification method is proposed by Hansman and Hunt [61]. It comprises four distinct dimensions. Their classification scheme as a whole contains different types of breaches, helping the protection by maintaining clarity in the language that defines the various types of attacks. The system is improved by a robust style where the distinct attack types are described in detail. The first dimension is an assistant for the administrator in the categorisation of the breach, The second aspect concentrate on the description of the target of the breach and the third one describes the mechanism reflecting the various stages of vulnerability. The possible impacts are defined by the final aspect to be obtained prior to the final act.

Mayhew et al. [62] proposed a packet detection system, based on SVM and K-means. They collected packets from a network of real businesses and parsed them with Bro. Firstly, the packets are grouped by type of protocol. Then, they clustered data for the different protocol datasets with the K-means++ algorithm. Thus, grouping the original dataset into several clusters takes place. The data are homologous in any cluster. Next, the features are extracted from the packets, and the SVM models are trained on each cluster. The detection accuracy of E-mail, Wiki, TCP, Twitter, and HTTP were 93%, 99%, 92.9%, 96%, and 96.6%, respectively.

Goeschel et al. [63] proposed A hybrid algorithm. The first step is done using NB, SVM, and decision tree algorithms; training on the SVM model divides the data into normal or abnormal samples. A decision tree model is used for the evaluation of attack types for the abnormal samples. Known attacks can be identified using the decision tree algorithm. Accordingly, unknown attack types can be identified using the NB algorithm. Using the three mentioned algorithms, a hybrid technique accomplished a good detection rate on the KDD99 dataset of 99.62% and FPR of 1.57%.

A spectral clustering-based and DNN classification method is introduced by Ma et al. [64]. Low accuracy is the cause of a heterogeneous flow. Hence, the original dataset was first divided into six homogenous subsets. Then, each subset is trained separately using the DNN algorithm. The accuracy results obtained for the NSL-KDD dataset is 92.1%.

Processing raw data directly in deep learning methods allows learning features and achieving classification in parallel. Poturi et al. [65] have suggested a detection method based on CNN. Experiments are conducted on the UNSW-NB15 and the NSL-KDD datasets. In such datasets, the type of data is a feature vector. The conversion of feature vectors into images takes place because processing 2-dimensional (2D) data on Convolutional Neural Network (CNNs) is good. Nominal features were one-hot coded, the dimensionality of features increases from 41 to 464. Then, each pixel is represented by 8-byte. Zeros are used as padding for blank pixels. The output was images of 8*8 pixels by transforming the vectors of feature into it. Finally, for the classification of attacks, a three-layer CNN is constructed. The performance of the proposed CNN is better than others with an accuracy of 94.9% on the UNSW-NB 15% and 91.14% on the NSL-KDD, compared to other DNN (GoogLeNet and ResNet 50).

The literature reviewed shows a clear lack of very high accuracy machine learning models to detect rare attacks using only relative features in the datasets and maintaining at the
same time low dimensionality which can in its turn reduce the training time. We therefore find the need for using CFS, data dimensionality reduction techniques with very powerful RF classifier to build our model.

3 | BENCHMARK DATASETS IN IDS

Machine learning has the function of extracting useful information from the data; hence, machine learning success relies on input data quality. The machine learning methodology is focussed on understanding the data. For IDSs, network and host behaviours should be captured correctly as well as the data should be reachable. In IDSs the source of data types includes packets, sessions, flows, and logs. Creating a dataset is a difficult, and time-consuming process. It can be re-used repeatedly by several researchers after a benchmark dataset is created. In addition to convenience, the use of benchmark datasets offers two other benefits. (1) Benchmark datasets are authoritative, and the findings of studies are more compelling. (2) Many publications have been used common benchmark datasets which allow the comparison of new study results with previous studies. Benchmark datasets are used herein: UNSW-NB15, and Phishing to implement various machine learning algorithms.

3.1 | UNSW-NB15

Three virtual machines were configured to capture the traffic of the network and the extraction of 47 features and two class labels. The setup of the environment was done by the University of South Wales [43]. UNSW-NB15 dataset is an evaluation of IDS datasets which is more complex compared to other benchmark datasets such as NSL-KDD and KDD99 because of different reasons: (1) traditional fewer types of attacks, (2) the absence of normal traffic situations. Recently, the UNSW-NB15 dataset was created for these issues. This dataset is composed of nine modern attack types and new normal traffic patterns, and 49 features that include the data flow between hosts and the examination of network packets to distinguish between normal and attack types of observations.

3.1.1 | Attack types

There are nine types of attached captured in the UNSW-NB15 dataset [43], mentioned in Moustafa and Slay [28,66] as follows:

1. **Fuzzers:** An attack using large quantities of random data called “Fuzz” to cause a network outage or crash servers across the network.
2. **Analysis:** attacks formed from spam files, footprinting, vulnerability scans, and port scans are included in this class. It is often referred to as Active Reconnaissance, in which scanning the network takes place without being exploited.
3. **Backdoors:** This family uses a technique by which a legitimate portal of the system is utilised by attackers for gaining unlawful access. Malicious software is used as part of an exploit to insert themselves in a device and give remote access for cyber-attackers.
4. **Denial of service (DoS):** A popular cyber-attack in which the attacker tries to expose a computer with many unauthorized demands for communication to make the network resources temporarily or permanently inaccessible to its expected users. These can be hard to differentiate from a legitimate network activity; however, some indicators exist for detecting these ongoing disruptive activities.
5. **Exploits:** In general, exploit attacks are accomplished by taking known vulnerabilities in operating systems as targets to exploit. For the automation of such attacks, exploit tools are used upon the discovery of a possible weakness in a network.
6. **Generic:** Is a cipher based attack; it is a type of collision attack on a generated secret key. The application of this type of attack is mainly focussed on message authentication, block, and stream block cipher. It relies on the greater likelihood of collisions between attempted random attacks.
7. **Reconnaissance:** Details about any public network or target host are gathered and are then used by manipulating techniques to use the collected information about the targeted networks or individual hosts. Free public information is used in this class ‘Whois’ service, Shodan, and ARIN records. Searches in social media help in such type of attacks. It can be called passive reconnaissance.
8. **Shellcode:** It is can be considered a sub-type of the of exploit attacks. This attack uses a tiny coding part as an exploit ‘s payload. To have remote access to a device, the injection of malicious code into an active application takes place. The attacker can control the compromised machine through a command shell.
9. **Worms:** Network propagation allows a malicious attack called a ‘worm’ to propagate through network propagation. The infection of a large network is fast. When a worm infects computers, it transforms them into zombies or bots to use them in distributed attacks.

Table 1 shows two types of sets in the UNSW-NB15 dataset: training and testing sets; several records have been split into training and testing sets with an approximate 60%:40% ratio respectively.

3.1.2 | Features

UNSW-NB15 dataset has 47 features and two labels [28,43,66]. The identification of features is included in flow features. It defines the flow of network traffic between connected hosts’ packets; basic features are about features that reflect basic protocol of header information; TCP header information attributes relating to the size of content are included in content features, packet flow, and source IP address; time features include features related to the time, for instance, the time
sources. There are 1353 different websites gathered in the
ated with phishing and trusted websites from many websites’
transitional payment systems such as online transactions, elec-
Institute for Cybersecurity. It has 5 days of network traffic
3.2 CICIDS-2017
The CICIDS-2017 dataset has been developed by the Canadian
Institute for Cybersecurity. It has 5 days of network traffic activities. CICIDS-2017 shows seven types of network attacks, Infiltration Attack, Brute Force Attack, DoS Attack, DDoS Attack, Botnet, Heart Bleed Attack, and Web Attack. Table 2 shows the types of sets in the CICIDS-2017 dataset. CICIDS-2017 dataset has 79 features and one label, it has been recorded in CSV format so it is very easy to be used to implement machine learning algorithms [47].

3.3 Phishing
There are 10 features in the Phishing dataset mainly related to transitional payment systems such as online transactions, electronic payments, e-commerce [67]. Other attributes are associated with phishing and trusted websites from many websites’ sources. There are 1353 different websites gathered in the Phishing dataset out of which 805 are identified as a phishing attack and others are identified as legitimate websites.

4 METHODOLOGY BACKGROUND
This method is about five parts. The beginning is to collect data from the network. During this phase, network data were collected using the benchmark UNSW-NB15 dataset. The second stage is the selection of attributes. In this step, for classification purposes, feature reduction, and ranking of the attributes, the (CFS) technique is applied. The third stage relates to data dimensional reduction. In this stage, the T-SNE algorithm is used to reduce the dimensionality of the data by representing high-dimensional data in two dimensions to allow visualization using scatter plots. The fourth step is the classification. In this step, the Random Forest (RF) technique is utilised to construct the model of classification. The last step is Accuracy. For the evaluation of detection models efficiency, accuracy, recall, Precision, F-Measure, and (FPR) are applied. When the data is trained and the model is developed, accuracy factors are reported.

4.1 Feature selection
Machine learning processing can run faster and more accurately by using specific relevant features [68]. There are many feature selection methods and algorithms that have been created such as the Gini index, Information Gain, correlation coefficients, and uncertainty [69]. One of the fastest feature selection algorithms is CFS, which uses heuristic evaluation function [70] for correlation-based features ranking. The previous process is called a scheme-independent attribute subset evaluator which is used by the CFS, which takes into account every attribute prediction value and inter-redundancy degree. The two main criteria when selecting the subset of the attributes are (a) strong relation to the class attribute and (b) without strong correlation. Attributes with a high correlation to each other are not selected to reduce the redundancy. Also, irrelevant attributes are not selected because it has no impact on the class attribute. The evaluation function of CFS’s feature is [70]:

\[
\text{Merit}_S = \frac{k r_{sf}}{\sqrt{k + k(k - 1) r_{ff}}}
\]  

(1)

\(\text{Merit}_S\) is the heuristic merit of a \(S\) subset of features that contains \(k\) features, the mean feature to class correlation is \(r_{sf}\) where \(f \in S\). While the average inter-correlation feature is \(r_{ff}\). The CFS is defined as follows:

\[
\max_k \left[ \frac{r_{f_1} + r_{f_2} + \cdots + r_{f_k}}{\sqrt{k + 2(r_{f_1} + \cdots + r_{f_k} + \cdots + r_{f_k})}} \right]
\]  

(2)
The correlations are represented by $r_{df}$ and $r_{vf}$ variables. Herein, the aim is to use machine learning algorithms and to feature a selection method to analyse the features included in all of the three datasets in which the main target of increasing the accuracy of the IDS in any network system can be achieved. This is why duplicated and irrelevant features are discarded from each dataset for much faster training and testing of the model with lesser resource utilization and at the same time higher attack detection rates.

4.2 | T-SNE

The T-SNE algorithm was developed by Laurens van der Maaten and Geoffrey Hinton [71]. It is a nonlinear dimensionality reduction algorithm, popular for representing high-dimensional data in two or three dimensions to allow visualization using scatter plots.

One of the main components in data analysis is a visual representation, as it enables hypotheses and intuitions to be formed for the processes producing the data. Visual analytic offers and builds methods from complex data to achieve such an understanding. This seeks to develop methods for analysts to analyse the underlying mechanisms of the data [72]. There are many data dimensionality non-parametric visualization algorithms used to visualise the datasets such as Classical scaling [73], which is closely related to Sammon mapping [74], PCA [75, 76], Locally Linear Embedding [77], and Isomap [77].

T-SNE is one of the most commonly used methods for data visualization to reduce dimensionality [78]. T-SNE is much easier to refine and provides substantially improved visualizations by reducing the tendency in the centre of the map to crowd points together. T-SNE is the best option while creating a single map visual representation at many scales [71].

4.3 | Random forest

Random decision forests classification algorithm is an ensemble machine learning method consisting of many decision trees in the training process and producing the output class using classification or prediction using regression of multiple trees. Random decision forests are right in outfitting the behaviour of decision trees to their training set.

T-SNE and Random forest are combined to get a good model for classifying UNSW-NB 15, CICIDS-2017, and phishing datasets. The coming part describes our approach and shows the algorithm and its detailed diagram.

5 | THE PROPOSED APPROACH OF T-SNERF

Many classification models such as SVM, RF, NB, BPNN, J48, ANN, and DNN have a good performance to solve efficiently complex problems in particular detection of intrusion problems. Their performance becomes poor when facing some issues related to intrusion detection fields such as data randomness. Therefore, within the next section, an approach is proposed to solve this issue.

5.1 | T-SNERF

In our algorithm, we used the train and test datasets of UNSW-NB 15, CICIDS-2017, and phishing datasets, then we used correlations feature-class to extract the important features and then R language code is used to implement T-SNE algorithm where data dimensionality can be reduced, subsequently, Weka machine learning application is used to implement random forest algorithm, that starts with the random sample selection from a given dataset. Next, the algorithm must create a decision tree for every sample. Then, all decision trees must produce prediction results. The proposed T-SNERF novel approach is shown in Figure 1. The figure is proposing a novel approach using multiple machine learning techniques to enhance the detection rate of rare attacks using only relative features in the datasets and maintaining at the same time low dimensionality which can in its turn reduce the training time.

5.2 | The T-SNERF algorithm

T-SNERF is combined with two main algorithms (T-Distributed Stochastic Neighbour Embedding and random forest algorithm), and these algorithms are used for predicting/classifying the categories network attacks type. The T-SNERF algorithm is illustrated in Algorithm 1.

Algorithm 1 Visualizations obtained by T-SNE for CICIDS-2017 dataset
EXPERIMENTAL RESULTS

This experiment examines and compares T-SNERF with other IDS machine learning models: SVM, J48, RF, NB, ANN, K-NN, and other hybrid techniques. All datasets’ classification results are compared against other machine learning algorithms and approaches using the accuracy of attack detection and FPR.

6.1 Performance metrics

For the evaluation of the performance of detection models: accuracy, precision, recall, F-measure, FPR and specificity are utilised in this research. The calculation of these parameters are as follows [79]:

\[
\text{Accuracy(ACC)} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)
\]

\[
\text{Precision(P)} = \frac{TP}{TP + FP} \quad (4)
\]

\[
\text{Recall(R)} = \frac{TP}{TP + FN} \quad (5)
\]

\[
F - \text{measure} = \frac{2(P \times R)}{P + R} \quad (6)
\]

\[
\text{FalsePositiveRate(FPR)} = \frac{FP}{FP + TN} \quad (7)
\]

\[
\text{Specificity} = \frac{TN}{TN + FP} \quad (8)
\]

When an attack is identified correctly, this represents the true positive (TP); a true negative (TN) when the model knows the normal network properly; a false negative (FN) when an attack is not identified correctly; and false positive (FP) when a normal network is identified as an attack. The accuracy rate calculates the percentage of correct attacks detection, recall.
shows the degree of correct attack detection of all output predicted as attack and specificity is the properly classified percentage of a normal network. Higher accuracy and recall with low FPR are indicators for a good classifier.

### 6.2 Results, comparisons of the accuracy, and discussion

In our approach, we used R programing language to achieve the requirement of the T-SNE algorithm, which is to find the similarity between the data points in the low-dimensional space to reduce the dimensionality of the data. The algorithm added two new columns to the original datasets. The two new columns are the map coordinates produced by T-SNE which represents the x and y-axis, and then the Random Forest algorithm is used to train these enriched datasets. Figure 2–4 present the datasets’ declaration of target distribution for the attack records for UNSW-NB1, CICIDS-2017, and Phishing datasets, respectively. There are two classes that defined in the datasets that is normal or attack classes, which are represented by the grey-scale of the points. The separation between the points is not clear due to a high number of instances. However, it is clear that data points are mostly close to each other in all of the three datasets. Some of the data points were not correctly classified by the RF in the final stage of the T-SNERF because they are difficult to identify.
Naive Bays, Random Forest, and J48 classifiers have been implemented after using CFS only without implementing the T-SNE algorithm using the UNSW-NB15 dataset. The result shows that RF and J48 algorithms performed best results with 97.59% and 93.78%, respectively, that is why this Random Forest algorithm has been selected in the T-SNERF approach. Table 3 shows the obtained accuracy and FPR for implemented classification techniques.

Three combined machine learning algorithms: CFS, T-SNE, and RF have been implemented together to produce these remarkable results. Table 4 shows the results for T-SNERF using multiple benchmark datasets used in machine learning-based IDS.

**Experiment 1: UNSW-NB15 data set.** Table 5 presents the accuracy obtained by implementing the proposed T-SNERF algorithm. Random forest initially obtained 97.60% in Table 3, which increased by 7.46% after using CFS technique, compared to 90.14% in [80]. Also, random forest accuracy has been increased by 2.4% by implementing T-SNE algorithm prior to the classification to achieve 100% accuracy and 0% FPR using nine features selected using CFS.

**Experiment 2: CICIDS-2017 dataset.** Table 6 compares the T-SNERF model with recent work on the CICIDS-2017 dataset. The comparison is based on the classifier used in each study, the results of the accuracy, and FAR. The proposed novel approach achieved significant results compared to existing approaches, achieving 99.7878% accuracy, and 0.003% FPR for the UNSW-NB15 dataset using nine features selected using CFS.

**Experiment 3: Phishing dataset.** Table 7 shows the proposed novel approach which achieved significant results compared to existing approaches, achieving 99.7044% accuracy and 0.003% FPR for the Phishing dataset using five features selected using CFS.

According to the work done in [81], feature selection is used with an NB classifier that achieved high accuracy. The results showed very high accuracy for rare types of attacks. As well as good enhancement in the FPR using NB and feature selection algorithms.

The feature selection process in [81], however, has almost no effect on the J48 algorithm, in which the accuracy performance of J48 classification was not impacted by using feature selection. We conclude from this that the UNSW-NB15 dataset has many redundant features.

Receiver operating curves (ROC): The ROC function is commonly used to demonstrate the discriminative potential of any machine learning algorithm. As shown in Figure 5, the performance of proposed T-SNERF algorithm is presented using the grey-dotted curve using recall function from Equation 5 on the y-coordinate versus the FPR from Equation 7 on the x-coordinate is used to generate the ROC. The area under the curve (AUC) is presented by the black point on the curve using Equation 8. As shown in the curve the value of AUC is equal to 1, which declares the power of the high accuracy for the T-SNERF algorithm which confirms that the model produces the best results.

**7 | CHALLENGES**

Although machine learning methods have good achievements for intrusion detection, they still face some challenges.

1. **Difficulty in getting datasets that simulate latest types of attacks.** Currently, the most common dataset is KDD99 that has several issues and needs recent datasets. The building of recent datasets, however, relies on the knowledge of expert knowledge and the price of labour time. However, the lack of datasets is intensified by the variability of the Internet environment. There are many new forms of attacks that are not reflected in current datasets. Also, being representative, balanced, and less redundant are the main conditions for available datasets. The systematic construction of datasets and learning may be solutions to this problem.

2. **Lower accuracy of detection for real conditions.** Machine learning approaches possess a potential for detecting intrusions, but on completely unknown data they still struggle to perform well. Most of the recent research used labelled datasets. As a result, without covering all samples in the real world, there is no guarantee of good results in actual environments—even if its accuracy is high on test sets.

3. **Low efficiency.** Most research gives importance for the results of the detection; thus, complicated models and extensive methods of pre-processing data that are typically employed, lead to low performance. However, the detection of attacks in real-time is a necessity for IDS to mitigate the harm as much as possible. a trade-off exists between effect and efficiency.

**8 | CONCLUSION**

Herein, a novel high accuracy machine learning algorithm T-SNERF has been introduced. It is used to solve the problem of network intrusion detection. Experiments have been applied on public datasets, in particular, the benchmark UNSW-NB15, CICIDS-2017, and Phishing datasets. T-SNERF accomplished...
excellent results with an accuracy of 100%, with a zero FPR for the UNSW-NB15 dataset, and a very high accuracy of 99.7878% and 99.7044% for CICIDS-2017 and Phishing datasets respectively. T-SNERF data dimensional reduction was exploited for the generation of Random Forests to improve the accuracy of the classifier. Also, its low running makes it suitable to deploy in the future for detecting intrusion tasks in real-time.

Our work is extended in the following directions. First, we have a plan for integrating our machine learning algorithm with recent machine learning reinforcement algorithms [87], for the optimization of our system for detecting network intrusion. Additionally, an intrusion recovery programme can be introduced to repair misuses and anomalies of intrusions that can occur in a device, so that after an attack or abuse or an anomaly has been detected, a procedure known as patching has to fix the programme or software or operating system related to the detected anomaly.

REFERENCES

1. Bijone, M.: A survey on secure network: intrusion detection & prevention approaches. Am. J. Inf. Syst. 4(3), 69–88 (2016)
2. Manikopoulos, C., Papavassiliou, S.: Network intrusion and fault detection: a statistical anomaly approach. IEEE Commun. Mag. 40(10), 76–82 (2002)
3. Qiu, M. et al.: Energy efficient security algorithm for power grid wide area monitoring system. IEEE Trans. Smart Grid 2(4), 715–723 (2011)
4. Roman, R., Zhou, J., Lopez, J.: Applying intrusion detection systems to wireless sensor networks. In: IEEE Consumer Communications & Networking Conference, pp. 640–644. CCNC, Las Vegas (2006). https://doi.org/10.1109/CCNC.2006.1593102
5. Sommer, R., Passon, V.: Outside the closed world: on using machine learning for network intrusion detection. In: 2010 IEEE symposium on security and privacy, pp. 305–316. IEEE, Oakland (2010)
6. Zhang, Y., Lee, W., Huang, Y.-A.: Intrusion detection techniques for mobile wireless networks. Wireless Netw. 9(5), 545–556 (2003)
7. Justel, A.: Daniel pe na, and rubén zamar. 1997. Stat. Probab. Lett. 35(3), 251–259 (1997)
8. Mardia, K.V.: Measures of multivariate skewness and kurtosis with applications. Biometrika. 57(3), 519–530 (1970)
9. Bland, J.M., Altman D.G.: (1988) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1(8476):307–310.
10. Hall, M.A., Smith, L.A.: Practical feature subset selection for machine learning. In: C. McDonald (ed.), Computer Science ’98 Proceedings of the 21st Australasian Computer Science Conference ACSC’98, Perth, 4-6 February 1998 pp. 181–191. Springer, Berlin (1998)
11. Panda, M., Patra, M.R.: Network intrusion detection using naive Bayes. Int. J. Comput. Sci. Netw. Secur. 7(12), 258–263 (2007)
12. Bouzida, Y., Cuppens, F.: Neural networks vs. decision trees for intrusion detection. In: IEEE/IST Workshop on Monitoring, Attack Detection and Mitigation, MonAM, vol. 28, p. 29. (2006)
13. Srinivas, M., Andrew, H., Aijit, A.: Intrusion detection using an ensemble of intelligent paradigms. J. Netw. Comput. Appl. 28, 167–182 (2005)
14. Syarif, I., Prugel-Bennett, A., Wills, G.: Unsupervised clustering approach for network anomaly detection. In: International Conference on Networked Digital Technologies, pp. 135–145. Springer, Heidelberg (2012)
15. KDD Cup 1999 data, 2020, http://engineering.purdue.edu/simmark/purcish. Accessed 15 April 2020. [Online]
16. Russell, D. et al.: Computer security basics. O’Reilly Media, Inc. Sebastopol (1991)
42. UNSW-NB15, 2020. https://www.unsw.adfa.edu.au/australian-centre-for-cyber-security/cybersecurity/ADFA-NB15-Datasets/. Accessed 16 April 2020. [Online].

43. Moustafa, N., Slay, J.: UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In: 2015 Military Communications and Information Systems Conference (MICIS), pp. 1–6. IEEE, Canberra (2015). https://doi.org/10.1109/MICIS.2015.7348942

44. Creech, G., Hu, J.: Generation of a new ids test dataset: time to retire the kdd collection,. In: 2013 IEEE Wireless Communications and Networking Conference (WCNC), pp. 4487–4492. IEEE, Shanghai (2013). https://doi.org/10.1109/WCNC.2013.6553301

45. McHugh, J.: Testing intrusion detection systems: a critique of the 1998 and 1999 darpa intrusion detection system evaluations as performed by lincoln laboratory. ACM Trans. Inf. Syst. Secur. 3(4), 262–294 (2000)

46. NSL-KDD dataset for network-based intrusion detection systems, 2020. http://nsl.cs.unb.ca/NSL-KDD/. Accessed 16 April 2020. [Online].

47. Sharafaldin, I., Lashkari, A.H., Ghorbani, A.A.: Toward generating a new intrusion detection dataset and intrusion traffic characterization. In: ICISSP, pp. 108–116. Funchal (2018)

48. Erfani, S.M. et al.: High-dimensional and large-scale anomaly detection using a linear one-class svm with deep learning. Pattern Recogn. 58, 121–134 (2016)

49. Fiore, U. et al.: Network anomaly detection with the restricted Boltzmann machine. Neurocomputing, 122, 13–23 (2013)

50. Javaid, A. et al.: A deep learning approach for network intrusion detection system. In: Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies, pp. 21–26. formerly BIONETICS (2016). https://doi.org/10.4108/eai.3-12.2015.2262516

51. Tang, T.A. et al.: Deep learning approach for network intrusion detection in software defined networking. In: 2016 International Conference on Wireless Networks and Mobile Communications (WINCOM), pp. 258–263. IEEE, Fez, Morocco (2016). https://doi.org/10.1109/WINCOM.2016.7777224

52. Yin, C. et al.: A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access. 5, 21 954–21 961 (2017)

53. Arefawashdeh, K., Purdy, C.: Toward an online anomaly intrusion detection system based on deep learning. In: 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 195–200. IEEE (2016)

54. Khan, F.A. et al.: A novel two-stage deep learning model for efficient network intrusion detection. IEEE Access. 7, 30 373–30 385 (2019)

55. Shone, N. et al.: A deep learning approach to network intrusion detection. IEEE Trans. Emerg. Top. Comput. Intell. 2(1), 41–50 (2019)

56. Nguyen, K. K. et al.: Cyberattack detection in mobile cloud computing: a deep learning approach. In: 2018 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–6. IEEE, Barcelona (2018). https://doi.org/10.1109/WCNC.2018.8376973

57. Moustafa, N., Turnbull, B., Choo, K-K. R.: An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of internet of things. IEEE Internet Things J. 6(3), 4815–4830 (2018)

58. Mogal, D.G., Ghungrad, S.R., Bhussare, B.B.: Nids using machine learning classifiers on unswe-nb15 and kddcup99 datasets. Int. J. Adv. Res. Comput. Commun. Eng. 6(4), 533–537 (2017)

59. Cannady, J.: Distributed detection of attacks in mobile ad hoc networks using learning vector quantization. In: 2009 Third International Conference on Network and System Security, pp. 571–574. IEEE (2009)

60. Nzhiga, J-P., Cannady, J.: Minimal dataset for network intrusion detection systems via mid-pca: a hybrid approach. In: 2012 6th IEEE International Conference Intelligent Systems, pp. 453–460. IEEE (2012)

61. Hansman, S., Hunt, R.: A taxonomy of network and computer attacks. Comput. Secur. 24(1), 31–43 (2005)

62. Mayhew, M. et al.: Use of machine learning in big data analytics for insider threat detection. In: MILCOM 2015-2015 IEEE Military Communications Conference, pp. 915–922. IEEE, Tampa (2015). https://doi.org/10.1109/MILCOM.2015.7357562

63. Goeckel, K.: Reducing false positives in intrusion detection systems using data-mining techniques utilising support vector machines, decision trees, and naive bayes for off-line analysis. In: SoutheastCon 2016, pp. 1–6. IEEE, Norfolk (2016). https://doi.org/10.1109/SECON.2016.7506774

64. Ma, Y. et al.: A hybrid spectral clustering and deep neural network ensemble algorithm for intrusion detection in sensor networks. Sensors. 16(10), 1701 (2016)

65. Potluri, S., Ahmed, S., Diehrich, C.: Convolutional neural networks for multi-class intrusion detection system. In: International Conference on Mining Intelligence and Knowledge Exploration, pp. 225–238. Springer, Cluj-Napoca (2018)

66. Moustafa, N., Slay, J.: The significant features of the unswe-nb15 and the kdd99 data sets for network intrusion detection systems. In: 2015 4th International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS), pp. 25–31. IEEE, Kyoto (2015). https://doi.org/10.1109/BADGERS.2015.014

67. Set, I.D. https://archive.ics.uci.edu/ml/datasets_Pima+Indians+Diabetes. (2014) Accessed 10 Sept. 2014.

68. Kumar, V., Minz, S.: Feature selection: a literature review. SmartCR. 4(3), 211–229 (2014)

69. Forman, G.: An extensive empirical study of feature selection metrics for text classification. J. Mach. Learn. Res. 3, 1289–1305 (2002)

70. Forman, G.: Correlation-based feature selection for machine learning. Master’s Thesis. 2020. https://www.cs.waikato.uct.ac.nz/simmball/thesis.pdf. Accessed 16 April 2020. [Online].

71. Maaten, L.v.d., Hinton, G.: Visualising data using t-sne. J. Mach. Learn. Res. 9, 2579–2605 (2008)

72. Keim, D. et al.: Mastering the information age: solving problems with visual analytics. Eurographics Association, Goslar (2010)

73. Torgerson, W.S.: Multidimensional scaling: I. theory and method. Psychometrika. 17(4), 401–419 (1952)

74. Sammon, J.W.: A nonlinear mapping for data structure analysis. IEEE Trans. Comput. 100(5), 401–409 (1969)

75. Mardia, K., Kent, J., Bibby, J.: Multivariate analysis. In: Probability and Mathematical Statistics. Academic Press Inc. [Google Scholar]. (1979)

76. Williams, C.K.: On a connection between kernel pca and metric multidimensional scaling. Mach. Learn. 46(1-3), 11–19 (2002)

77. Tenenbaum, J.B., De Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science. 290(5500), 2319–2323 (2000)

78. Cao, Y., Wang, L.: Automatic selection of t-sne perplexity. (2017) arXiv preprint arXiv:1708.03229.

79. Kayacik, H.G., Zincker-Heywood, A.N., Heywood, M.I.: A hierarchical som-based intrusion detection system. Eng. Appl. Artif. Intell. 20(4), 439–451 (2007)

80. Suleiman, M., Issac, B.: Performance comparison of intrusion detection machine learning classifiers on benchmark and new datasets. Paper presented at 28th International Conference on Computer Theory and Applications, Alexandria, Egypt (2018)

81. Bagui, S. et al.: Using machine learning techniques to identify rare cyber-attacks on the unswe-nb15 dataset. Secur. Priv. 2(6), p. e91 (2019)

82. Akou, D. et al.: Intrusion detection with comparative analysis of supervised learning techniques and Fisher score feature selection algorithm. In: International Symposium on Computer and Information Sciences, pp. 141–149. Springer (2018)

83. Singh Parsvar, S., Raiwani, Y., Parwar, L.S.: Evaluation of network intrusion detection with features selection and machine learning algorithms on ciedy-2017 dataset. In: International Conference on Advances in
84. Gustavsson, V.: Machine learning for a network-based intrusion detection system: an application using zek and the CICIDS2017 dataset. Dissertation (2019)

85. Sharafaldin, I., Lashkari, A.H., Ghorbani, A.A.: A detailed analysis of the cicsids2017 data set. In: International Conference on Information Systems Security and Privacy, pp. 172–188. Springer (2018)

86. Yulianto, A., Sukarno, P., Suwastika, N. A.: Improving adaboost-based intrusion detection system (ids) performance on cicsids2017 dataset. In: Journal of Physics: Conference Series, vol. 1192, no. 1. IOP Publishing (2019) p. 012018.

87. Mahmud, M. et al.: Applications of deep learning and reinforcement learning to biological data. IEEE Trans. Neural Netw. Learn. Syst. 29(6), 2063–2079 (2018)

How to cite this article: Hammad M, Hewahi N, Elmedany W. T-SNERF: A novel high accuracy machine learning approach for Intrusion Detection Systems. IET Inf. Secur. 2021;15:178–190. https://doi.org/10.1049/ise2.12020