Information Fusion 90 (2023) 353–363

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full length article

Fusion of statistical importance for feature selection in Deep Neural Network-based Intrusion Detection System

Ankit Thakkar, Ritika Lohiya *

Institute of Technology, Nirma University, Ahmedabad, Gujarat 382 481, India

ARTICLE INFO

Keywords:
Intrusion Detection System
Deep Learning
Filter-based feature selection
Deep Neural Network
Standard deviation
Fusion of statistical importance

ABSTRACT

Intrusion Detection System (IDS) is an essential part of network as it contributes towards securing the network against various vulnerabilities and threats. Over the past decades, there has been comprehensive study in the field of IDS and various approaches have been developed to design intrusion detection and classification system. With the proliferation in the usage of Deep Learning (DL) techniques and their ability to learn data extensively, we aim to design Deep Neural Network (DNN)-based IDS. In this study, we aim to focus on enhancing the performance of DNN-based IDS by proposing a novel feature selection technique that selects features via fusion of statistical importance using Standard Deviation and Difference of Mean and Median. Here, in the proposed approach, features are pruned based on their rank derived using fusion of statistical importance. Moreover, fusion of statistical importance aims to derive relevant features that possess high discernibility and deviation, that assists in better learning of data. The performance of the proposed approach is evaluated using three intrusion detection datasets, namely, NSL-KDD, UNSW_NB-15, and CIC-IDS-2017. Performance analysis is presented in terms of different evaluation metrics such as accuracy, precision, recall, f-score, and False Positive Rate (FPR) and the results are compared with existing feature selection techniques. Apart from evaluation metrics, performance comparison is also presented in terms of execution time. Moreover, results achieved are also statistically tested using Wilcoxon Signed Rank test.

1. Introduction

The fundamental rationale of developing Intrusion Detection System (IDS) is to detect and classify network samples with precise classification accuracy and minimum false alarms [1,2]. Hence, various philosophical and analytical design principles should be taken into consideration while developing intrusion detection and classification system. Over past decades, there has been various efforts in designing an efficient IDS using Deep Learning (DL) techniques [3]. Furthermore, DL techniques such as Deep Neural Network (DNN) has emerged as one of the leading solutions for building an efficient IDS [4]. This is because, DNN has an intriguing characteristic attribute of performing end-to-end learning and in-depth analysis to derive patterns in data for prediction and classification [5]. Hence, DL techniques such as DNN can be considered as one of the intelligent techniques that performs implicit learning on high-dimensional data with ease. Apart from handling high-dimensional data, DNN offers high-level data abstraction and good generalization ability for underlying attack classification problem [6].

In this research, we aim to design a DNN-based intrusion detection and classification system by applying fusion of statistical importance by considering statistical measures for feature engineering on intrusion detection datasets. We have applied fusion of statistical importance using statistical measures to derive association and identify significance of features for feature selection [7]. For intrusion classification, DNN requires a large amount of data for learning and deriving patterns. In the field of IDS, various intrusion detection datasets have been developed for analysis and learning [8]. These datasets have been developed by capturing raw network traffic flowing through underlying network environment. Various networking tools such as Wireshark and Nmap are used for capturing raw network traffic [9]. Further, the captured data is stored in the form of pcap or tcpdump files, which are processed to extract network features from network packets consisting header and payload information [10]. Hence, intrusion detection datasets used for the performance evaluation comprises of high-dimensional network feature space for learning. However, considering network features, there is a possibility that intrusion detection datasets might consist of redundant and irrelevant features that conceivably would either affect or would not contribute towards prediction and classification process [11].

Consequently, considering the role and importance of feature engineering in intrusion detection and classification process, we aim

* Corresponding author.
E-mail addresses: ankit.thakkar@nirmauni.ac.in (A. Thakkar), 18ftphde30@nirmauni.ac.in (R. Lohiya).

https://doi.org/10.1016/j.inffus.2022.09.026
Received 23 February 2022; Received in revised form 26 September 2022; Accepted 28 September 2022
Available online 3 October 2022

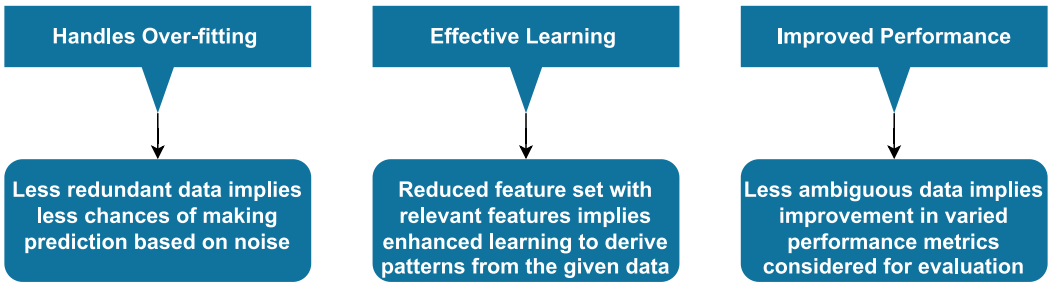

Fig. 1. Scientific contribution and importance of the proposed feature selection technique.

to design a novel feature engineering process that selects features based on statistical interpretation of features for the underlying intrusion detection datasets. Generating a reduced subset of interpretative features is a critical process and hence, we aim to design a novel filter-based feature selection technique that considers standard deviation, mean, and median statistical measures for deriving reduced feature subset for learning and performance enhancement of DNN-based IDS. Unlike other feature selection technique, filter-based feature selection approach aims at deriving feature subset without any influence of classification technique applied for learning and prediction [12].

1.1. Scientific contribution and importance of the proposed feature selection technique

With a goal of designing effective predictive model for the underlying classification problem, feature selection technique can be considered as a heuristic approach that may not guarantee optimal, perfect, or rational performance but is an adequate medium to achieve immediate and effective performance for the underlying classification problem [13]. Hence, the science and art of the proposed feature selection technique is based on the heuristics, namely, standard deviation and difference ($|Mean - Median|$) of the features in the given dataset. The scientific contribution and importance of the proposed feature selection technique is shown in Fig. 1.

The major contributions of our proposed work are summarized as follows.

- We have proposed a novel feature selection technique based on fusion of statistical importance of features for intrusion detection and classification.
- DNN is applied for learning and classification process using reduced feature subset.
- For fusion of statistical importance of features, statistical measures, namely, standard deviation, mean, and median are taken into consideration.
- The proposed approach is evaluated using three intrusion detection datasets, namely, NSL-KDD, UNSW_NB-15, and CIC-IDS-2017.
- Performance analysis of the proposed approach is presented in terms of varied evaluation metrics such as accuracy, precision, recall, f-score, and False Positive Rate (FPR).
- Performance of the proposed approach is compared with different feature selection techniques such as Chi-Square, Correlation-based Feature Selection (CFS), Recursive Feature Elimination, Genetic Algorithm (GA), Mutual Information (MI), Relief-f, and Random Forest (RF).
- Comparative analysis with existing feature selection techniques is presented using different evaluation metrics that have been considered as well as execution time.

The remainder of the paper is organized as follows. Section 2 presents overview of feature selection techniques for intrusion detection and classification. Section 3 describes the proposed feature selection technique for DNN-based IDS. Section 4 discusses experimental methodology adopted for intrusion detection and classification. Section 5 present and discuss result analysis of implemented techniques. Section 6 concludes the work presented in this paper.

2. Related work

Over the past years, various approaches have been proposed for intrusion detection and classification. In [14], a wrapper-based feature selection technique is proposed using Genetic Algorithm (GA) and Logistic Regression (LR). Here, in the proposed approach, GA along with LR is applied for feature selection and Decision Tree (DT) classifier is applied for classification. The performance of the proposed approach is evaluated using Weka tool with two intrusion detection datasets, namely, KDD CUP 99 and UNSW_NB-15 datasets. Result analysis revealed that the proposed feature selection technique achieved 99.90% Detection Rate (DR) and 0.1% False Alarm Rate (FAR) for KDD CUP 99 dataset with 18 features and 81.24% DR and 6.39% FAR for UNSW_NB-15 dataset with 20 features.

Flexible Mutual Information (FMI) feature selection technique is proposed in [15] for deriving reduced feature subset for intrusion detection and classification. Here, in the proposed approach, Least Square-Support Vector Machine (LS-SVM) is applied for classification and features are selected using FMI considering correlation among the features of the dataset. Moreover, FMI is a non-linear feature selection techniques that uses correlation as a metric for feature selection. For assessing the performance of the proposed approach three intrusion detection datasets are used, namely, Kyoto 2006, KDD CUP 99, and NSL-KDD datasets. The performance analysis of the proposed approach is presented in terms of DR and FAR.

Correlation-based Feature Selection (CFS) technique is applied in [16] for intrusion detection and classification. Here, in the proposed approach, DT classifier is applied for classification that uses reduced feature subset derived using CFS. The performance of the proposed approach is evaluated using NSL-KDD dataset consisting of 41 features. The proposed approach derives reduced feature subset with 14 features which is further used for intrusion detection and classification. Performance analysis of the proposed approach is presented in terms of accuracy and it can be inferred from the results that the proposed approach achieved accuracy of 90.30% for NSL-KDD dataset with 14 features.

Comparative analysis of various classifiers is performed in [17] using Weka tool. Here, for the comparative analysis various feature selection techniques are applied, namely, attribute evaluator, greedy stepwise, IG, and ranker technique. Further, two feature subsets are derived for intrusion detection and classification by performing defined number of simulations. It is inferred from the performance analysis that Random Forest (RF) classifier performs better in terms of overall performance for both derived feature subsets. Moreover, result analysis is presented in terms of Kappa statistic and accuracy for demonstrating the performance of each feature subset.

A filter-based feature selection technique using IG is proposed in [18] for intrusion detection and classification. Here, in the proposed

approach, classification is performed by integrating rule-based approach with multiple tree classifiers. The performance of the proposed approach is evaluated using UNSW_NB-15 dataset with 22 features derived using IG technique. Furthermore, results analysis is presented in terms of accuracy, f-score, and FAR.

Feature selection using RF is applied in [19], wherein feature ranking is performed using feature importance. Here, in the proposed study, feature importance of each feature is computed and features are ranked based on their feature importance value. This implies that feature with highest rank can be considered as most significant feature for intrusion detection and classification. For the prediction and classification, various classification techniques were implemented, namely, k-Nearest Neighbour (kNN), DT, Bagging Meta Estimator (BME), XGBoost, and RF. The performance of the proposed approach is assessed using UNSW_NB-15 dataset with reduced feature subset consisting of 11 features. Furthermore, result analysis is presented in terms of accuracy and f-score.

An ensemble two-tier IDS is designed in [20], wherein, hybrid feature selection techniques is implemented along with majority voting-based classification. Here, in the proposed approach, features are selected using hybrid technique designed using PSO, GA, and Ant Colony Optimization (ACO). Further, for the classification, rotation forest and bagging classifier are applied and predictions are generated using majority voting technique. The performance of the proposed approach is evaluated using KDD CUP 99 dataset with reduced feature subset consisting of 19 features. Moreover, the performance of the proposed approach is validated using 10-fold cross validation technique and result analysis is presented in terms of accuracy, precision, recall, and FAR.

A two-stage intrusion classification model is designed using RF classifier in [21]. Here, in the proposed approach, IG is applied as feature selection technique. In the first stage, detection of minority class is performed and in the second stage, majority class is detected. Prediction from each stage are combined to generate classification result. The performance of the proposed approach is assessed using UNSW_NB-15 dataset and results are presented in terms of accuracy and FAR.

RepTree-based two stage IDS is proposed in [22], wherein IG is used as feature selection technique. Here, in the initial stage underlying dataset is divided into three categories based on the protocol type and further, in the second stage classification is performed. The performance of the proposed approach is evaluated using UNSW_NB-15 dataset with a reduced feature set consisting of 20 features and results are presented in terms of accuracy. An incremental technique comprising of Extreme Learning Machine (IELM) and Advanced Principal Component (APCA) algorithm is proposed in [23] for intrusion detection and classification. Here, in the proposed approach, ELM is applied for classification and APCA is applied for adaptive feature selection. The performance of the proposed approach is assessed using UNSW_NB-15 dataset and results are presented in terms of accuracy, DR, and FAR.

Intrusion detection and classification system is designed in [25] using NB and MLP classifier. Here, in the proposed approach, combined feature selection technique is applied that consists of three feature selection techniques, namely, IG, GR, and ReliefF. Performance of the proposed approach is evaluated using KDD CUP 99 dataset and result analysis is presented in terms of accuracy and FAR. An ensemble framework along with feature selection is designed in [24] for intrusion detection and classification. Here, in the proposed approach, GR is applied for selecting significant features for learning and Bagging classifier is applied for classification. Performance evaluation of the proposed approach is conducted using NSL-KDD dataset and results are presented in terms of classification accuracy and FAR. The comparative summary of existing ML approaches for IDS is presented in Table Table 1.

2.1. Comparative analysis of existing approaches for IDS

The design and development of discussed research work for intrusion detection and classification using feature engineering is encouraging. However, varied IDS have been designed using different learning algorithms and feature selection techniques which adapt to unique learning strategy for feature selection as well as attack classification [26–28]. However, research gaps still exist with different learning mechanisms for intrusion detection and classification such as,

- Majority of the research work have designed IDS considering existing feature selection techniques using visualization tools such as WeKa [14,17].
- The existing techniques designed for intrusion detection and classification using feature selection techniques have been analysed and compared using outdated datasets which lack experimental scenario.

Hence, there is a scope for developing an enhanced technique for intrusion detection and classification. Therefore, we aim to design a novel feature selection approach for DNN-based IDS that uses fusion of statistical importance derived using standard deviation and absolute difference of mean and median to select relevant and contributing features for prediction and classification process. The application of statistical importance to select features is effective because it derives features based on statistical reasoning, which enhances the performance of designed DNN-based IDS with feature discernibility and deviation. Thus, novelties of our proposed work can be summarized as follows.

- A novel feature selection technique based on fusion of statistical importance of features is proposed for intrusion detection and classification.
- Experimental-based analysis of novel feature selection techniques is performed using recent datasets and datasets used in literature, namely, NSL-KDD, UNSW_NB-15, and CIC-IDS-2017.
- We have presented comparative analysis of the proposed feature selection technique with existing feature selection techniques.

3. Proposed feature selection technique for intrusion detection and classification

For intrusion detection and classification, various feature selection techniques are applied which are categorized in three categories, namely, filter-based feature selection technique, wrapper-based feature selection technique, and embedded feature selection technique [29]. In filter-based feature selection technique, reduced feature subset is derived based on certain relevancy criteria that defines significance of features pertaining to learning and classification [30]. Hence, in filter-based feature selection technique a relevancy score is derived and features are filtered based on the calculated score [30]. In wrapper-based feature selection technique, classification algorithm is considered and knowledge based is constructed to derive feature subset for learning and classification [31]. The knowledge base of features reveals the importance of features in refined form based on underlying classification algorithm. Feature selection using wrapper-based techniques is performed using predefined rules and conditions. However, performance of wrapper-based feature selection technique is dependent on the type of classification algorithm used [31]. Embedded feature selection technique is implemented by incorporating two phases that are learning phase and feature selection phase [32]. This implies that embedded technique employs selecting features separately in two phases, wherein outcomes of learning phase are used to add or delete features in feature selection phase [32].

Table 1
Comparative summary of existing Machine Learning (ML) approaches for IDS.

Ref	Technique	Feature selection	Dataset	Results
[16]	DT	CFS	NSL-KDD	· Accuracy for NSL-KDD: 90.30%
[15]	LS-SVM	FMI	Kyoto 2006, KDD CUP 99, and NSL-KDD	· DR for KDD CUP 99: 99.46% · DR for NSL-KDD: 98.76% · DR for Kyoto 2006: 99.64%
[17]	RF	Attribute evaluator, greedy stepwise, IG, and ranker	KDD CUP 99, UNSW_NB-15	Comparative analysis is presented in graphical format for feature selection technique considered.
[22]	RepTree	IG	NSL-KDD, UNSW_NB-15	· Accuracy for NSL-KDD: 89.85% · Accuracy for UNSW_NB-15: 88.95%
[14]	DT	GA	KDD CUP 99, UNSW_NB-15	· DR for KDD CUP 99: 99.90% · DR for UNSW_NB-15: 81.24%
[19]	kNN, DT, BME, XGBoost, and RF	Feature importance	UNSW_NB-15	· Accuracy for kNN: 71.01% Accuracy for DT: 74.22% Accuracy for BME: 74.64% Accuracy for XGBoost: 71.43% Accuracy for RF: 74.87%
[21]	RF	IG	UNSW_NB-15	· Accuracy for UNSW_NB-15: 85.78%
[24]	Bagging classifier	GR	NSL-KDD	Accuracy for NSL-KDD: 84.25%
[23]	IELM	APCA	NSL-KDD, UNSW_NB-15	· Accuracy for NSL-KDD: 81.22% · Accuracy for UNSW_NB-15: 70.51%
[20]	Rotation forest and Bagging classifier	PSO, ACO, and GA	KDD CUP 99	· Accuracy for KDD CUP 99: 72.52%
[25]	NB, MLP	Combined feature selection technique	KDD CUP 99	· Accuracy for NB: 93.00% · Accuracy for MLP: 97.00%
[18]	Rule-based multiple tree classifiers	IG	UNSW_NB-15	· Accuracy for UNSW_NB-15: 84.83%

3.1. Association between intrusion classification and feature selection

Intrusion detection datasets are developed by sniffing network packets flowing through network environment using various networking tools such as Wireshark and Nmap [26]. Captured network packets are accumulated in form of raw network files such as pcap files or tcpdump files. These files consist of various details regarding network communication which are extracted from network packet header and network packet payload. The details regarding network communication from captured network traffic serve as network features for designed IDS. Intrusion detection and classification system examines the network activities and analyses the data to inspect whether the analysed data flow is anomalous network traffic or normal network traffic [32]. IDS analyses the data to check whether system's confidentiality, integrity, or availability is compromised or not. While designing an IDS, various aspects are considered such as monitoring network, collecting data, statistically analysing the collected data, intrusion detection, intimidating the security administrator of an intrusive event, and responding to intrusions [8].

In literature various researchers have designed hybridized approaches by combining feature selection techniques with classification technique [26]. Considerably, feature selection technique is incorporated to enhance the performance of the designed IDS. However, the combination of feature selection and intrusion detection should focus on achieving increased classification accuracy with reduced number of false positives. Therefore, the designed IDS requires an efficient feature selection technique which is capable of extracting significant features from the underlying dataset. Hence, the proposed feature selection technique aims to select features by considering statistical characteristics of features.

3.2. Conceptualization of the proposed feature selection technique

Nowadays, enormous amount of network traffic is generated from various network resources. Features from flowing network traffic are studied for deriving patterns of normal and anomalous network traffic. However, network traffic data needs to examined with precision and effectiveness. Selecting relevant features plays a significant role in deriving pertinent information from a large number of data samples. Feature selection is one of the critical approaches that directs to select features from the underlying dataset which can contribute better in enhancing the predictive capability for the given classification problem. Hence, feature selection can be described as selection strategy adopted to remove irrelevant and redundant features for better representation of data.

In our study, a novel filter-based feature selection technique is designed to derive relevant features from the intrusion detection dataset that can contribute more towards learning and classification process. Hence, with an aim to enhance the performance of DNN-based IDS, a new and discerning feature selection technique named as Feature Selection via Standard Deviation and Difference of Mean and Median is proposed in our study. The proposed feature selection technique derives reduced feature subset with high discernibility and deviation. The application of standard deviation, mean, and median is effective in deriving features as these measures perform quantitative and statistical reasoning to derive relevant features for intrusion detection and classification [33]. The fusion of statistical importance using statistical measures aims at improving the performance of prediction and classification through quantitative and descriptive comparisons [33]. The conceptualization strategy of the proposed feature selection techniques is presented in Fig. 2.

3.3. Standard deviation

Standard deviation for features can be described as a statistical measure that measures the amount of variation or deviation in features from mean [34]. The standard deviation can be computed using Eq. (1) [34].

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}} \tag{1}$$

Here, in Eq. (1), σ represents standard deviation, N is the total number of samples, x_i represents each value from underlying feature, and μ represents mean of underlying feature.

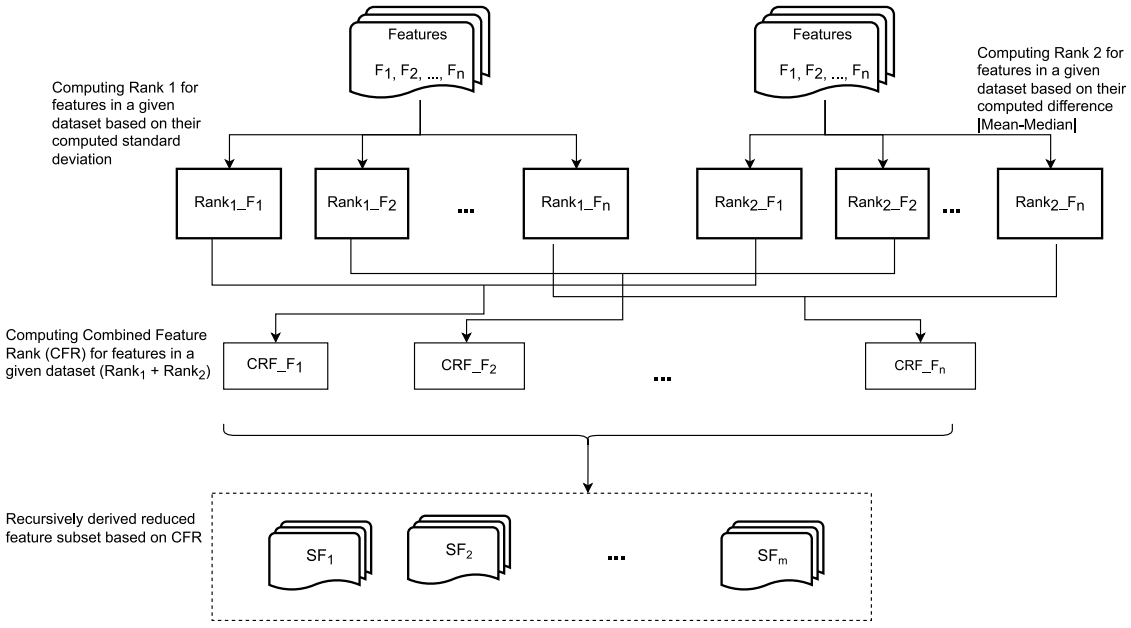

Fig. 2. Conceptualization strategy of the proposed feature selection technique.

Interpretation of standard deviation reveals that high standard deviation value indicates that feature is dispersed over large range of values and a low standard deviation value indicates that feature values are closely located with respect to mean [33]. Hence, feature selection using standard deviation chooses features with high standard deviation value because as feature values are extended over large range, effective prediction outcome can be achieved. Moreover, standard deviation represents features distinguishable capability and therefore, standard deviation of a feature manifests its differences on all samples. This implies that high standard deviation value reveals more differences the feature has on all samples [33].

3.4. Mean and median

Mean and Median can be defined as descriptive statistical measures that are used to characterize data distribution [35]. Moreover, these statistical measures represents relative magnitude of deviation in a data distribution [35]. For feature selection, we have utilized the absolute value of the difference between mean and median to derive relevant features from the dataset, which is represented using Eq. (2).

$$D = |Mean - Median| \tag{2}$$

Here, in Eq. (2), D represents absolute value of the difference between mean and median for a given feature. Interpretation of the difference of mean and median reveals that high difference value indicates deviation over a large range of values and hence, features with high difference value can be selected as relevant features from dataset for effective prediction and classification process [34].

3.5. Process of feature selection for intrusion detection and classification

The outcome of feature selection process is a set of relevant features that are strongly related with class output labels and contribute more towards the learning patterns from data. To quantify contribution of a given feature in classification process, we introduce combined feature rank that represents significance of a features. The combined feature rank is computed based on ranks derived using fusion of standard deviation and difference of mean and median. From the description of standard deviation and difference of mean and median is revealed that features with highest values possess strong discernibility and minimum redundancy. Hence, process of feature selection is described as follows.

- Compute the standard deviation (σ) of the features of dataset.
- Rank the features based on standard deviation value from high to low. Assign rank derived using standard deviation (σ) as $Rank_1$.
- Compute the absolute difference (D) of mean and median of the features of dataset.
- Rank the features based on difference value from high to low. Assign rank derived using difference (D) as $Rank_2$.
- Compute combined feature rank as Combined Feature Rank = $Rank_1 + Rank_2$.
- Recursively add features to feature subset based on combined feature rank until accuracy is not better than the previous derived feature subset.

The algorithm for recursive feature selection using proposed technique is presented in Algorithm 1. The derived feature subset is given input to DNN model for training and classification.

4. Experimental methodology

The proposed work implements DNN for intrusion detection and classification. DNN architecture is a multi-layered neural network structure that performs mathematical transformations on input data to derive and learn patterns for prediction and classification [36]. The experimental methodology of the proposed approach consists of various phases such as deciding on intrusion detection datasets required for the performance evaluation, data pre-processing for transforming data for ease of experimentation, feature selection to derived reduced feature subset for learning, training DNN with reduced feature subset, and performance evaluation. The schematic of the proposed approach is as shown in Fig. 3.

4.1. Dataset description

The performance of the proposed approach is evaluated using three intrusion detection datasets, namely, NSL-KDD, UNSW_NB-15, and CIC-IDS-2017. These datasets consist of wide variety of network features and have been developed in different network environments [8]. Moreover, these datasets consist of realistic as well as synthetic network traffic. Hence, performance of the proposed approach can be indisputably advocated using diversified network traffic from three different datasets. A brief description of each dataset is as follows.

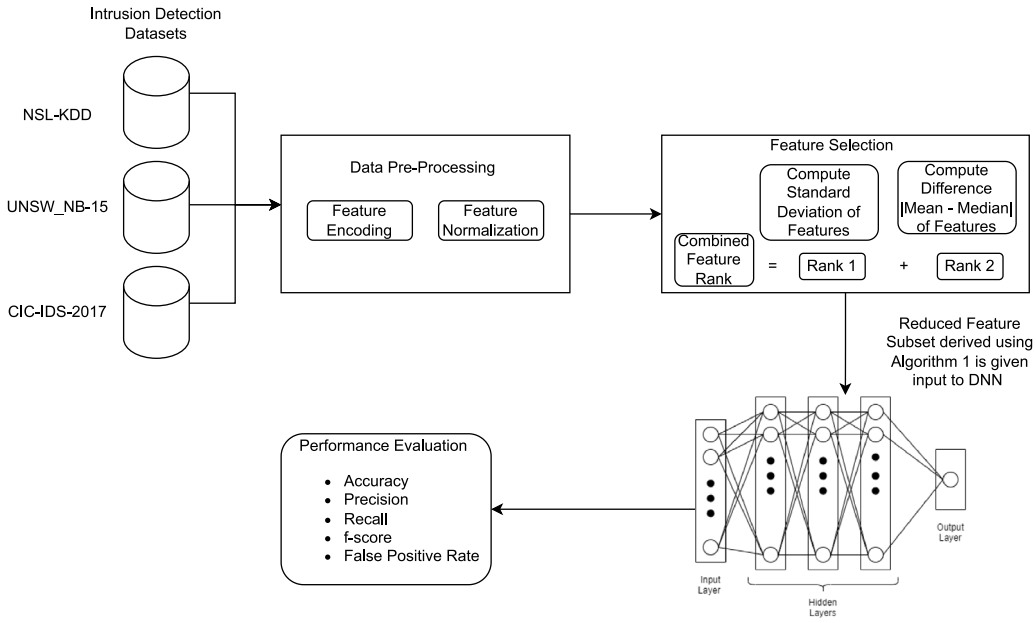

Fig. 3. Schematic of the proposed approach.

Algorithm 1 Recursive Feature Selection Using Fusion of Standard Deviation and Absolute Difference of Mean and Median

1: Consider Dataset D_l for intrusion detection and classification where D = {NSL-KDD, UNSW_NB-15, CIC-IDS-2017}.
2: For features of dataset D_l, calculate standard deviation for each feature using equation (1).
3: Sort features from high to low based on their standard deviation and rank them. Consider the assigned rank as $Rank_1$.
4: For features of dataset D_l, calculate absolute value of difference between mean and median of each feature using equation (2).
5: Sort features from high to low based on the absolute value of the difference and rank them. Consider the assigned rank as $Rank_2$.
6: Compute combined feature rank R by summing $Rank_1$ and $Rank_2$.
7: For each feature $F_i \in F$ of dataset D_l do,
8: Remove the highest rank feature F_i from F and update S_l as $S_l = S_l \cup F_i$.
9: Train DNN model on training set with S_l features and compute model accuracy.
10: Repeat Steps [8-9], for features F_i until increase in accuracy is recorded more than previous computed accuracy.
11: Store the derived relevant features in subset S_l for the Dataset D_l.
12: Use feature subset S_l for training DNN-based IDS for the dataset D_l.

- NSL-KDD Dataset: It is an intrusion detection dataset that has been developed by eliminating missing and duplicate samples from KDD CUP 99 dataset [37]. It consist of different categories of network features as well as samples for four attack categories and normal network traffic [38]. Moreover, a distinct training dataset and test dataset with 125,973 and 22,544 data samples, respectively, is present in NSL-KDD dataset [38].
- UNSW_NB-15 Dataset: It is an intrusion detection dataset that has been developed using IXIA Perfect Storm tool which captures the network packets flowing in the designed network testbed [39]. The dataset is developed with network traffic that describes security vulnerabilities and exploits along with normal network traffic. The network features of the designed dataset are extracted using software tools, namely, Argus and Bro-IDS [39]. Moreover, a distinct training and test datasets with 175,341 and 82,332 data samples, respectively, is present in UNSW_NB-15 dataset [39].
- CIC-IDS-2017 Dataset: It is one of the largest and recent intrusion detection datasets that is developed by sniffing real-time network packets flowing through the network [40]. The designed intrusion detection dataset covers wide range of network services, protocols, and modernistic attack categories. The dataset consist of data samples captured over the span of five days. Moreover, the designed dataset consist of distinctive wide range of network features extracted using CICFlowMeter tool [41].

The statistics of intrusion detection datasets used for experimentation is presented in Table 2.

4.2. Data pre-processing

Data pre-processing techniques are applied for ease of experimentation to convert the data for smooth processing and learning [36]. In the proposed work, two data pre-processing techniques are applied, namely, feature encoding and feature normalization. Feature encoding is performed to convert categorical features into numerical features [4]. Intrusion detection datasets used for experimentation consist of categorical features such as flag, service type, and protocol type. The categorical features are converted into numerical features by applying one-hot encoding technique. One-hot encoding is one of the common feature encoding technique that is applied for numeralization of categorical features [4]. Further, after feature encoding, feature normalization is performed, as datasets might consist of features that are in different dimensions and scale of values. Hence, for feature normalization standard scalar technique is applied that normalizes the features by subtracting the mean and scaling the feature values to unit variance.

4.3. Feature selection

Feature selection process is performed to derived reduced feature subset consisting of relevant and contributing features from considered intrusion detection dataset. Features are selected using proposed feature selection technique described in Section 3. The proposed feature selection technique is applied on intrusion detection datasets, namely, NSL-KDD, UNSW_NB-15, and CIC-IDS-2017. Application of proposed

Table 2
Statistics of the experimental datasets [8].

Criteria (↓)/Dataset (→)	NSL-KDD	UNSW_NB-15	CIC-IDS-2017
Type of network traffic	Real & Synthetic	Synthetic	Real
Number of features	41	42	79
Number of attack categories	4	9	7
Number of classes	5	10	15
Number of data samples	148 517	257 673	225 745
Number of samples in training set	125 973	175 341	165 730
Number of samples in test set	22 544	82 332	60 015

Table 3
Neural network architecture and configuration details [4].

Criteria	Values
Model	Sequential
Number of hidden layers [4]	3
Size of input	NSL-KDD: 21, UNSW_NB: 21, CIC-IDS-2017: 64
Number of neurons in hidden layers [4]	1024, 768, 512
Activation function for hidden layer [4]	ReLU
Activation function for output layer [4]	Sigmoid
Dropout techniques	Standard dropout (p = 0.1) (Derived using GridSearchCV)
Batch-size [4]	1024
Epochs [4]	300

feature selection technique results in reduced feature susbet of 21 features out of 41 features for NSL-KDD dataset, 21 features out of 42 features for UNSW_NB-15 dataset, and 64 features out of 79 features for CIC-IDS-2017 dataset. The reduced feature subset is given input to DNN model for learning and prediction.

4.4. Deep neural network for intrusion detection and classification

Multi-layered DNN architecture is designed for intrusion detection and classification. The designed DNN architecture consists of an input layer with input dimension equal to the number of features derived using feature selection, three fully connected dense hidden layers with varied number of neurons for data transformation and learning, and an output layer with one neuron for binary classification. The intricate layered structure of neurons learns patterns by exhibiting end-to-end learning and performs prediction for given input sample. With each fully connected layer ReLU activation function is used to strengthen effect of learning process [5]. Moreover, succeeding to every dense layer, a dropout layer is incorporated to achieve generalization and avoid co-adaptation in neural network [42]. For output layer, Sigmoid activation function is used to predict the class output label. Furthermore, performance of the DNN structure is evaluated by applying binary cross entropy loss function. The structure of DNN is presented in Fig. 4 and its configuration details are presented in Table 3.

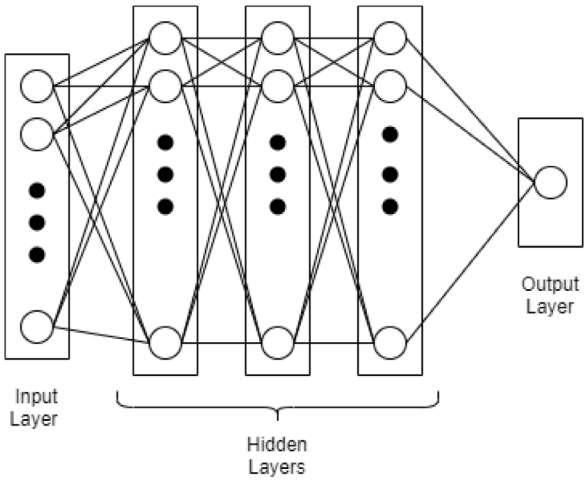

Fig. 4. Deep Neural Network.

4.5. Performance evaluation

Performance of the proposed approach is presented using varied evaluation metrics derived from confusion matrix, namely, accuracy, precision, recall, f-score, and FPR [26]. The evaluation metrics are expressed using Eqs. (3)–(7).

$$Accuracy = \frac{T_p + T_n}{T_p + F_p + F_n + T_n} \tag{3}$$

$$Precision = \frac{T_p}{T_p + F_p} \tag{4}$$

$$Recall = \frac{T_p}{T_p + F_n} \tag{5}$$

$$f - score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{6}$$

$$FPR = \frac{F_p}{F_p + T_n} \tag{7}$$

Here, in Eqs. (3)–(7), T_p, T_n, F_p, and F_n represent true positive, true negative, false positive, and false negative, respectively [26].

5. Result analysis

Experiments for evaluation of the proposed approach is performed on Intel(R) Core(TM) i5-8265U CPU processor with 64-bit Windows 10 operating system and 8.00 GB RAM using Python. The experiments are performed on pre-processed intrusion detection datasets namely, NSL-KDD, UNSW_NB-15, and CIC-IDS-2017 with reduced feature subset derived using proposed feature selection technique. Experiments are performed for ten runs and achieved results are averaged. For the performance analysis, we have compared our proposed feature selection technique with existing feature selection techniques that are described as follows.

- Recursive Feature Elimination (RFE) : In RFE, feature selection is performed by recursively eliminating features based on feature importance and deriving a feature subset that consist of relevant features with superior feature importance scores [12].
- Chi-Square: In Chi-Square feature selection technique reduced feature subset is derived by performing chi-square statistical test that measures the dependency among features [12].
- Correlation-based Feature Selection (CFS): CFS technique is based on the hypothesis that a good feature subset consist of features that are in strong correlation with the target class and in low correlation with each other [43]. Hence, in CFS features are selected based on their computed correlation coefficient score.
- Genetic Algorithm: In feature selection technique based on genetic algorithm, feature are selected based on their fitness values computed using defined fitness function. In [44], fitness function

Table 4
Results for NSL-KDD Dataset.

Technique	Feature selection	No. of selected features	Accuracy	Precision	Recall	f-score	FPR	Execution time (s)
DNN	Recursive Feature Elimination (RFE) [12]	13	98.94	99.39	98.75	99.33	0.012	32 519.045
DNN	Chi-Square [12]	13	98.92	99.92	98.73	99.32	0.012	31 613.115
DNN	Correlation-based Feature Selection (CFS) [43]	30	92.65	99.59	91.24	95.23	0.082	25 915.630
DNN	Genetic Algorithm [44]	23	94.90	95.10	94.30	94.70	0.094	35 569.016
DNN	Mutual information	13	98.89	99.90	98.70	99.30	0.0291	33 033.130
DNN	Relief-f	20	81.94	81.91	98.46	89.42	0.0530	36 045.110
DNN	Random forest	16	98.88	99.89	98.71	99.30	0.0210	34 761.620
DNN	Proposed feature selection technique	21	**99.84**	**99.94**	**98.81**	**99.37**	**0.011**	**22 318.015**

Note: The value in boldface indicates the best performance for the experiments conducted for NSL-KDD dataset.

Table 5
Results for UNSW_NB-15 Dataset.

Technique	Feature selection	No. of selected features	Accuracy	Precision	Recall	f-score	FPR	Execution time (s)
DNN	Recursive Feature Elimination (RFE) [12]	13	82.21	78.71	98.86	87.64	0.013	22 314.470
DNN	Chi-Square [12]	13	82.41	79.02	98.61	87.73	0.013	21 832.195
DNN	Correlation-based Feature Selection (CFS) [43]	30	75.34	67.43	98.29	79.99	0.017	21 766.215
DNN	Genetic Algorithm [44]	30	76.70	92.70	95.00	93.83	0.069	25 387.412
DNN	Mutual information	21	76.26	72.87	97.92	83.55	0.077	18 190.205
DNN	Relief-f	13	72.34	73.26	89.09	80.40	0.1090	18 643.650
DNN	Random forest	17	82.69	79.37	98.41	87.87	0.0518	18 152.130
DNN	Proposed feature selection technique	21	**89.03**	**95.00**	**98.95**	**96.93**	**0.011**	**13 913.500**

Note: The value in boldface indicates the best performance for the experiments conducted for UNSW_NB-15 dataset.

is defined using accuracy, f-score, and FPR, which is used to calculate fitness of features and further, features with high fitness values are chosen for intrusion detection and classification.

- Mutual Information: In mutual information, feature selection is performed by estimating dependency between the features. The selection of the feature relies upon non-parametric procedure, namely, entropy estimation [45].
- Relief-f: This feature selection technique is based on sensitive feature interactions, wherein feature score is derived for each feature which is further considered to rank features. The feature score is obtained by estimating feature value differences among the nearest neighbour instance pairs [45].
- Random Forest: Random forest is one of the popular classification technique that possess implicit feature selection capability. In random forest, features are selected based on their measure of impurity, namely, Gini index. Hence, while training random forest classifier, there is a possibility to determine how much each feature reduces the impurity. The more a feature reduces impurity, the more significant it is [46].

The experimental results for NSL-KDD, UNSW_NB-15, and CIC-IDS-2017 are presented in Tables 4, 5, and 6, respectively. It can be inferred from the results that the proposed feature selection technique achieved better results compared to existing features selection technique with DNN-based IDS for all the three intrusion detection datasets. For NSL-KDD dataset, the proposed approach achieved 99.84% accuracy with the derived reduced feature subset. Hence, an approximate increase of 1%–18% in accuracy is recorded with proposed feature selection technique for DNN-based IDS using NSL-KDD dataset. For UNSW_NB-15 dataset, the proposed approach achieved 89.03% accuracy with the derived reduced feature subset. Hence, an approximate increase of 7%–17% in accuracy is recorded with the proposed feature selection technique for DNN-based IDS using UNSW_NB-15 dataset. For CIC-IDS-2017 dataset, the proposed approach achieved 99.80% accuracy with the reduced feature subset. Hence, an approximate increase of 0.9%–2% in accuracy is recorded with the proposed feature selection technique for DNN-based IDS using CIC-IDS-2017 dataset.

Precision and recall evaluation metrics illustrate the relevancy and sensitivity of underlying classification technique for a given application problem. For the proposed approach, promising scores for precision and recall evaluation metrics are achieved for all the three datasets. For NSL-KDD dataset, approximate increase of 0.04%–18% in precision and 0.06%–7% in recall is recorded with precision of 99.94% and recall of 98.81% using proposed feature selection technique. For UNSW_NB-15 dataset, approximate increase of 3%–28% in precision and 0.09%–9% in recall is recorded with precision of 95.00% and recall of 98.95% using proposed feature selection technique. For CIC-IDS-2017 dataset, approximate increase of 0.05%–2% in precision and 0.67%–2% in recall is recorded with precision of 99.85% and recall of 99.94% using proposed feature selection technique.

It is interesting to study the performance of classification techniques with unbalanced dataset using f-score. This is because, f-score can be considered as one of the important performance measure which is a balanced metric that considers both precision and recall. For NSL-KDD dataset, 99.37% f-score value is achieved with proposed approach, which is approximately, 0.07%–10% more compared to other feature selection technique. For UNSW_NB-15 dataset, 96.93% f-score value is achieved with proposed approach, which is approximately, 3%–17% more compared to other feature selection technique. For CIC-IDS-2017 dataset, 99.89% f-score value is achieved with proposed approach, which is approximately, 0.59%–2% more compared to other feature selection technique. Moreover, the proposed feature selection approach has outperformed with respect to minimum FPR for DNN-based IDS compared to other feature selection techniques.

Apart from the evaluation metrics, the proposed approach can also be compared based on the execution time recorded that consist of pre-processing, feature selection, training, and classification. The execution time for all the three intrusion detection datasets is presented in Tables 4–6. It can be inferred from the execution time that with derived feature subset the proposed DNN-based IDS has recorded less execution time even though the number of features are higher than few of the existing techniques that have been implemented.

Table 6
Results for CIC-IDS-2017 Dataset.

Technique	Feature selection	No. of selected features	Accuracy	Precision	Recall	f-score	FPR	Execution time (s)
DNN	Recursive Feature Elimination (RFE) [12]	13	98.81	98.00	98.00	98.00	0.041	35 214.235
DNN	Chi-Square [12]	13	98.15	98.20	98.93	98.56	0.062	35 517.115
DNN	Correlation-based Feature Selection (CFS) [43]	54	97.78	97.77	97.00	97.38	0.094	32 261.510
DNN	Genetic Algorithm [44]	38	98.00	98.89	97.77	98.32	0.069	40 552.215
DNN	Mutual information	35	98.00	98.17	99.82	98.98	0.0178	32 031.522
DNN	Relief-f	35	98.99	99.07	99.07	99.07	0.0801	32 045.130
DNN	Random forest	37	98.90	99.90	98.70	99.30	0.0601	32 029.250
DNN	Proposed feature selection technique	64	**99.80**	**99.85**	**99.94**	**99.89**	**0.012**	**27 719.360**

Note: The value in boldface indicates the best performance for the experiments conducted for CIC-IDS-2017 dataset.

5.1. Comparison with existing studies and future scope

The hypothesis of feature selection technique reveals that feature selection is an essential part of a learning model that facilitates the model to extract and learn features and thereby reduce the complexity of the model [26]. Feature engineering can be performed by selecting or extracting relevant features from the dataset. In our proposed approach we aim to focus on enhancing the performance of DNN-based IDS by proposing a novel feature selection technique that selects features via fusion of statistical importance using Standard Deviation and Difference of Mean and Median. Result analysis of the proposed approach communicates that the proposed approach performs better to existing feature selection techniques considered for performance comparison. Apart from comparative analysis with existing feature selection technique, we have also presented comparative analysis with existing research work in the field of intrusion detection and classification as shown in Table 7.

Considering and comparing with the research work and their achieved results following key insights can be derived.

- It can be deduced from result analysis that DL techniques performs better compared to ML techniques for intrusion detection and classification. There are various factors that contribute in better performance of DL-based IDS such as, efficacy to handle high-dimensional data, better feature learning capability, and effective learning strategy. The proposed approach is able to achieve to improved performance for NSL-KDD dataset with an approximate increase in accuracy of 26.27% compared to [47].
- Comparing the results of the proposed approach with other DL techniques presented in [48,49], it can be deduced that the proposed approach achieved better results for NSL-KDD dataset in terms of accuracy, wherein increase in accuracy is reported approximately 9% and 1% compared to [48,49], respectively. Moreover, the proposed approach has achieved improved performance in terms of FPR for NSL-KDD dataset with approximate decrease of 7% and 6% compared to [48,49], respectively.
- Moreover, comparable performance is achieved for other performance metrics such as precision, recall, and f-score.

However, for the performance recorded for UNSW_NB-15 dataset in [48,49] is better in terms of varied performance metrics such as accuracy, precision, recall, and f-score. This is because based on the exploratory analysis of UNSW_NB-15 dataset consist of high number of outliers and skewed data, which can be effectively handled by CNN and LSTM architectures [50]. However, the proposed approach records better performance in terms of FPR for UNSW_NB-15 dataset compared to [48,49]. Hence, in future, it would be promising to consider analysing the resiliency of IDS by optimizing the neural network architecture using nature-inspired algorithms or by using nature-inspired algorithms as feature selection techniques.

5.2. Complexity analysis

In this section, we discuss the time complexity of Algorithm 1. For analysing the time complexity, assume that N is the number of data samples in the underlying dataset, d is the number of features, m is the number of features in feature subset S, n is the number of nested subset of features. The computational significance of Algorithm 1 is to derive feature subset S to enhance the process of DNN-based IDS by minimizing the generalization error and increasing the predictive capability.

The proposed feature selection requires to compute standard deviation and difference of mean and median for each feature. The time complexity of calculating standard deviation, mean, median, and combined rank of all features is $O(dNlogN)$. Time complexity of recursively eliminating one feature from the feature subset is $O(\frac{Nd^2}{2})$ [52]. Hence, the time complexity of Algorithm 1 is $max[O(dNlogN), O(\frac{Nd^2}{2})]$.

5.3. Energy consumption analysis

For profiling energy consumption analysis for different datasets, it is intriguing to note that the energy consumption for a given task refers to the core power usage during the task execution time [53]. This implies that the energy consumption is directly proportional to the execution time during which the power is consumed. From the Tables 4–6, it can be inferred that the proposed approach recorded less execution time compared to the existing feature selection techniques for all the three intrusion detection datasets considered for performance evaluation. Hence, from the result analysis, it can be deduced that the proposed approach consumed less energy compared to other existing feature selection techniques considered for comparative analysis.

5.4. Statistical significance and discussions

The attained results are also statistically validated using Wilcoxon signed-rank test for all the performance measures considered for the experimentation. Significance of the results achieved can be expressed using p-value, wherein p-value should be less than 0.05 [54]. It can be inferred from the Table 8, that the p-value obtained for all the three datasets considered for experimentation is less than 0.05. Hence, the results achieved are statistically significant.

Considering the role and importance of feature engineering in intrusion detection and classification process, the proposed feature selection technique recursively derives features based on their statistical properties. Focus of the proposed feature selection approach is to recursively derive significant features from the underlying dataset based on their computed combined rank. Single-feature ranking procedure assumes that the features in the underlying dataset are independent of each other. However, there are often correlations among features that should be considered to impose feature redundancy while feature selection. Hence, multi-feature ranking can be considered that takes correlation as well as combined feature to derive reduced feature subset. Thus, this can serve as an important future research direction which can be consider prospective scope in the field of feature engineering for intrusion detection and classification.

Table 7
Comparison with existing studies.

Ref.	Technique	Feature selection	Dataset	Result analysis
[51]	Decision Tree (DT)	Linear correlation coefficient	KDD CUP 99	Accuracy: 95.03%, Detection Rate: 95.23%, FPR: 1.65%
[47]	Ensemble tree classifier	CFS-Bat algorithm	NSL-KDD	Results for NSL-KDD: Accuracy of 73.57%, Detection Rate of 73.6% and FPR of 12.92%
[48]	Optimized CNN	Hierarchical multi-scale LSTM	NSL-KDD, ISCX, UNSW_NB-15	· NSL-KDD dataset Accuracy: 90.67%, Precision: 86.71%, Recall: 95.19%, f-score: 91.46%, FPR: 8.86%, and Training time: 5118 s. · ISCX dataset Accuracy: 95.33%, Precision: 100%, Recall: 94.77%, f-score: 97.61%, FPR: 7.84%, and Training time: 54 480 s. · UNSW_NB-15 dataset Accuracy: 96.33%, Precision: 100%, Recall: 95.87%, f-score: 98.13%, FPR: 5.87%, and Training time: 30 665 s.
[49]	Black widow optimized Conv LSTM	Artificial bee colony	NSL-KDD, UNSW_NB-15, ISCX, CIC-IDS-2018	· NSL-KDD dataset accuracy: 98.67%, Precision: 97.48%, Recall: 100%, f-score: 98.73%, FPR: 7.50%, and Training time: 4675.45 s. · UNSW_NB-15 dataset accuracy: 98.66%, Precision: 100%, Recall: 98.77%, f-score: 98.77%, FPR: 4.48%, and Training time: 26 721.2 s. · ISCX dataset accuracy: 97.00%, Precision: 100%, Recall: 95.78%, f-score: 99.67%, FPR: 5.76%, and Training time: 48 761.05 s. · CSE-CIC-IDS-2018 dataset accuracy: 98.25%, Precision: 97.48%, Recall: 98.67%, f-score: 98.18%, FPR: 2.52%, and Training time: 22 713.02 s.
Our Study	DNN	Proposed feature selection	NSL-KDD, UNSW_NB-15, CIC-IDS-2017	· NSL-KDD dataset accuracy: 99.84%, Precision: 99.94%, Recall: 98.81%, f-score: 99.37%, FPR: 1.1%, and Execution time: 22 318.015 s. · UNSW_NB-15 dataset accuracy: 89.03%, Precision: 95.00%, Recall: 98.95%, f-score: 96.93%, FPR: 1.1%, and Execution time: 13 913.50 s. · CIC-IDS-2017 dataset accuracy: 99.80%, Precision: 99.85%, Recall: 99.94%, f-score: 99.89%, FPR: 1.2%, and Execution time: 27 719.36 s.

Table 8
Wilcoxon signed-rank test results.

Dataset	p-value
NSL-KDD	0.0027
UNSW_NB-15	0.0053
CIC-IDS-2017	0.0054

6. Concluding remarks

The study proposes a novel feature selection technique based on fusion of statistical importance using standard deviation and difference of mean and median for enhancing the performance of intrusion detection and classification. The proposed feature selection technique aims at deriving reduced feature subset that consist of features that have attributes such as high discernibility and deviation. For prediction and classification Deep Neural Network (DNN) technique is applied that considers reduced feature subset for learning and deriving patterns in data. Performance evaluation of the proposed approach is performed using three intrusion detection datasets, namely, NSL-KDD, UNSW_NB-15, and CIC-IDS-2017. The performance of the proposed approach is demonstrated in terms of accuracy, precision, recall, f-score, False Positive Rate (FPR), and execution time. From the experiments performed, it can be deduced that the proposed approach achieved better performance compared to existing feature selection techniques for all the three intrusion detection datasets with reduced execution time. Hence, the derived features using proposed feature selection technique were able to enhance the performance of DNN-based IDS.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors are unable or have chosen not to specify which data has been used.

References

[1] A. Thakkar, R. Lohiya, Role of swarm and evolutionary algorithms for intrusion detection system: A survey, Swarm Evol. Comput. 53 (2020) 100631.
[2] R. Lohiya, A. Thakkar, Application domains, evaluation datasets, and research challenges of IoT: A systematic review, IEEE Internet Things J. (2020).
[3] A. Thakkar, R. Lohiya, A review on machine learning and deep learning perspectives of IDS for IoT: Recent updates, security issues, and challenges, Arch. Comput. Methods Eng. (2020) 1–33, http://dx.doi.org/10.1007/s11831-020-09496-0.
[4] A. Thakkar, R. Lohiya, Analyzing fusion of regularization techniques in the deep learning-based intrusion detection system, Int. J. Intell. Syst. (2021).
[5] M.A. Chang, D. Bottini, L. Jian, P. Kumar, A. Panda, S. Shenker, How to train your DNN: The network operator edition, 2020, arXiv preprint arXiv:2004.10275.
[6] R. Lohiya, A. Thakkar, Intrusion detection using deep neural network with antirectifier layer, in: Applied Soft Computing and Communication Networks, Springer, 2021, pp. 89–105.
[7] F.E. White, Data Fusion Lexicon, Technical Report, Joint Directors of Labs Washington DC, 1991.
[8] A. Thakkar, R. Lohiya, A review of the advancement in intrusion detection datasets, Procedia Comput. Sci. 167 (2020) 636–645.
[9] G. Bagyalakshmi, G. Rajkumar, N. Arunkumar, M. Easwaran, K. Narasimhan, V. Elamaran, M. Solarte, I. Hernández, G. Ramirez-Gonzalez, Network vulnerability analysis on brain signal/image databases using nmap and wireshark tools, IEEE Access 6 (2018) 57144–57151.
[10] A. Gharib, I. Sharafaldin, A.H. Lashkari, A.A. Ghorbani, An evaluation framework for intrusion detection dataset, in: 2016 International Conference on Information Science and Security (ICISS), IEEE, 2016, pp. 1–6.
[11] G. Creech, J. Hu, Generation of a new IDS test dataset: Time to retire the KDD collection, in: 2013 IEEE Wireless Communications and Networking Conference (WCNC), IEEE, 2013, pp. 4487–4492.
[12] A. Thakkar, R. Lohiya, Attack classification using feature selection techniques: a comparative study, J. Ambient Intell. Humaniz. Comput. 12 (1) (2021) 1249–1266.
[13] O. Almomani, A feature selection model for network intrusion detection system based on PSO, GWO, FFA and GA algorithms, Symmetry 12 (6) (2020) 1046.
[14] C. Khammassi, S. Krichen, A GA-LR wrapper approach for feature selection in network intrusion detection, Comput. Secur. 70 (2017) 255–277.
[15] M.A. Ambusaidi, X. He, P. Nanda, Z. Tan, Building an intrusion detection system using a filter-based feature selection algorithm, IEEE Trans. Comput. 65 (10) (2016) 2986–2998.
[16] B. Ingre, A. Yadav, Performance analysis of NSL-KDD dataset using ANN, in: 2015 International Conference on Signal Processing and Communication Engineering Systems, IEEE, 2015, pp. 92–96.
[17] T. Janarthanan, S. Zargari, Feature selection in UNSW-NB15 and KDDCUP'99 datasets, in: 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE), IEEE, 2017, pp. 1881–1886.

[18] V. Kumar, D. Sinha, A.K. Das, S.C. Pandey, R.T. Goswami, An integrated rule based intrusion detection system: analysis on UNSW-NB15 data set and the real time online dataset, Cluster Comput. 23 (2) (2020) 1397–1418.
[19] N.M. Khan, N. Madhav C, A. Negi, I.S. Thaseen, Analysis on improving the performance of machine learning models using feature selection technique, in: International Conference on Intelligent Systems Design and Applications, Springer, 2018, pp. 69–77.
[20] B.A. Tama, M. Comuzzi, K.-H. Rhee, TSE-IDS: A two-stage classifier ensemble for intelligent anomaly-based intrusion detection system, IEEE Access 7 (2019) 94497–94507.
[21] W. Zong, Y.-W. Chow, W. Susilo, A two-stage classifier approach for network intrusion detection, in: International Conference on Information Security Practice and Experience, Springer, 2018, pp. 329–340.
[22] M. Belouch, S. El Hadaj, M. Idhammad, A two-stage classifier approach using reptree algorithm for network intrusion detection, Int. J. Adv. Comput. Sci. Appl. 8 (6) (2017) 389–394.
[23] J. Gao, S. Chai, B. Zhang, Y. Xia, Research on network intrusion detection based on incremental extreme learning machine and adaptive principal component analysis, Energies 12 (7) (2019) 1223.
[24] N.T. Pham, E. Foo, S. Suriadi, H. Jeffrey, H.F.M. Lahza, Improving performance of intrusion detection system using ensemble methods and feature selection, in: Proceedings of the Australasian Computer Science Week Multiconference, 2018, pp. 1–6.
[25] A.A. Salih, M.B. Abdulrazaq, Combining best features selection using three classifiers in intrusion detection system, in: 2019 International Conference on Advanced Science and Engineering (ICOASE), IEEE, 2019, pp. 94–99.
[26] A. Thakkar, R. Lohiya, A survey on intrusion detection system: feature selection, model, performance measures, application perspective, challenges, and future research directions, Artif. Intell. Rev. (2021) 1–111.
[27] Y. Xin, L. Kong, Z. Liu, Y. Chen, Y. Li, H. Zhu, M. Gao, H. Hou, C. Wang, Machine learning and deep learning methods for cybersecurity, IEEE Access (2018).
[28] A.L. Buczak, E. Guven, A survey of data mining and machine learning methods for cyber security intrusion detection, IEEE Commun. Surv. Tutor. 18 (2) (2016) 1153–1176.
[29] L.-H. Li, R. Ahmad, W.-C. Tsai, A.K. Sharma, A feature selection based DNN for intrusion detection system, in: 2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM), IEEE, 2021, pp. 1–8.
[30] T.-S. Chou, K.K. Yen, J. Luo, Network intrusion detection design using feature selection of soft computing paradigms, Int. J. Comput. Intell. 4 (3) (2008) 196–208.
[31] S. Zaman, F. Karray, Features selection for intrusion detection systems based on support vector machines, in: Consumer Communications and Networking Conference, 2009. CCNC 2009. 6th IEEE, IEEE, 2009, pp. 1–8.
[32] S. Aljawarneh, M. Aldwairi, M.B. Yassein, Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model, J. Comput. Sci. 25 (2018) 152–160.
[33] J. Xie, M. Wang, S. Xu, Z. Huang, P.W. Grant, The unsupervised feature selection algorithms based on standard deviation and cosine similarity for genomic data analysis, Front. Genet. 12 (2021).
[34] R. de Nijs, T.L. Klausen, On the expected difference between mean and median, Electron. J. Appl. Statist. Anal. 6 (1) (2013) 110–117.
[35] T. Pham-Gia, T.L. Hung, The mean and median absolute deviations, Math. Comput. Modelling 34 (7–8) (2001) 921–936.
[36] P. Chen, Y. Guo, J. Zhang, Y. Wang, H. Hu, A novel preprocessing methodology for DNN-based intrusion detection, in: 2020 IEEE 6th International Conference on Computer and Communications (ICCC), IEEE, 2020, pp. 2059–2064.
[37] U. Repository, NSL-KDD dataset, 2009, URL https://www.unb.ca/cic/datasets/nsl.html (accessed April 22, 2019).
[38] L. Dhanabal, S. Shantharajah, A study on NSL-KDD dataset for intrusion detection system based on classification algorithms, Int. J. Adv. Res. Comput. Commun. Eng. 4 (6) (2015) 446–452.
[39] N. Moustafa, J. Slay, UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set), in: 2015 Military Communications and Information Systems Conference (MilCIS), IEEE, 2015, pp. 1–6.
[40] I. Sharafaldin, A.H. Lashkari, A.A. Ghorbani, Toward generating a new intrusion detection dataset and intrusion traffic characterization, in: ICISSP, 2018, pp. 108–116.
[41] R. Panigrahi, S. Borah, A detailed analysis of CICIDS2017 dataset for designing intrusion detection systems, Int. J. Eng. Technol. 7 (3.24) (2018) 479–482.
[42] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (1) (2014) 1929–1958.
[43] N. Gopika, M.E.A. Meena Kowshalaya, Correlation based feature selection algorithm for machine learning, in: 2018 3rd International Conference on Communication and Electronics Systems (ICCES), IEEE, 2018, pp. 692–695.
[44] Z. Liu, Y. Shi, A hybrid IDS using GA-based feature selection method and random forest, Int. J. Mach. Learn. Comput. 12 (2) (2022).
[45] Y. Zhang, X. Ren, J. Zhang, Intrusion detection method based on information gain and relieff feature selection, in: 2019 International Joint Conference on Neural Networks (IJCNN), IEEE, 2019, pp. 1–5.
[46] X. Li, W. Chen, Q. Zhang, L. Wu, Building auto-encoder intrusion detection system based on random forest feature selection, Comput. Secur. 95 (2020) 101851.
[47] Y. Zhou, G. Cheng, S. Jiang, M. Dai, Building an efficient intrusion detection system based on feature selection and ensemble classifier, Comput. Netw. 174 (2020) 107247.
[48] P.R. Kanna, P. Santhi, Unified deep learning approach for efficient intrusion detection system using integrated spatial–temporal features, Knowl.-Based Syst. 226 (2021) 107132.
[49] P.R. Kanna, P. Santhi, Hybrid intrusion detection using MapReduce based black widow optimized convolutional long short-term memory neural networks, Expert Syst. Appl. 194 (2022) 116545.
[50] N. Sharma, N.S. Yadav, S. Sharma, Classification of UNSW-NB15 dataset using exploratory data analysis using ensemble learning, EAI Endorsed Trans. Ind. Netw. Intell. Syst. 8 (29) (2021) e4.
[51] S. Mohammadi, H. Mirvaziri, M. Ghazizadeh-Ahsaee, H. Karimipour, Cyber intrusion detection by combined feature selection algorithm, J. Inf. Secur. Appl. 44 (2019) 80–88.
[52] X. Ding, F. Yang, F. Ma, An efficient model selection for linear discriminant function-based recursive feature elimination, J. Biomed. Inform. 129 (2022) 104070.
[53] S. Hajiamini, B.A. Shirazi, A study of DVFS methodologies for multicore systems with islanding feature, in: Advances in Computers, Vol. 119, Elsevier, 2020, pp. 35–71.
[54] S. Taheri, G. Hesamian, A generalization of the wilcoxon signed-rank test and its applications, Statist. Papers 54 (2) (2013) 457–470.