A Hard Voting Ensemble Model of the Logistic Regression, Support Vector Machine and Random Forest for Network Intrusion Detection

Jasim Mohammed Dahr[1] and Yaghoub Farjami[2]
[1]Directorate of Education in Basrah, Basrah, Iraq.
[2]Department of Computer and Information Technology, University of Qom, Qom, Iran.
E-mail: Jmd20586@gmail.com and farjami@qom.ac.ir

Keywords: SVM, ensemble, logistic regression, random forest, intrusion, hard voting, intrusion, web packets, hybrid model, misprediction

Received: March 11, 2025

The rapid evolving landscape of cybersecurity, the need for robust and efficient intrusion detection systems (IDS) has never been more critical. The real-time network traffic environments are plagued with the challenges posed by traffic routing and complex network behaviours. To this end, this paper proposes four hybrid/ensemble models for the detection of intrusion across complex networks infrastructures by combining logistic regression (LR) and support vector machines (SVM), and random forest (RF). The paper leverages the hard voting ensemble strategy to mix the interpretability of LR; fusing decision capacity of RF; and classification efficacy of SVM, to enhance detection accuracy and reduce false positive rates. During the experimentation of the proposed ensemble models, the two standard datasets were acquired, that is, the KDD Cup 1999 and CSE-CIC-IDS2018, for the training and testing phases after the ENN-SMOTE method data resampling strategies. The results showed that, the resampled binary-class dataset (KDD Cup 1999), the SVM-RF obtains best accuracy of 99.42%. The biggest precision score of 99.94% was computed for the LR-RF model. The recall measure of 99.01% was attained by the SVM-LR-RF model. F1-score of 99.37% was observed for the SVM-RF model. Similarly, upon ENN-SMOTE method resampling of the multi-class dataset (CSE-CIC-IDS2018), accuracy of 92.06% achieved by the SVM-RF model. The precision of 99.68% was witnessed for the SVM-RF. With the recall score, the SVM-LR-RF model offers the widest margin of 90.76%. The F1-score of 94.73% was recorded for the SVM-LR-RF model. The Asymptotic Significance (2-sided test) of 0.043 is less than p-value at the significance level of 5%. The paper established that, the ensembles models performances with the ENN-SMOTE technique were more significant than the RFE for data preprocessing.

Povzetrek: Prispevek predstavi ansambelski model s trdim glasovanjem, ki združuje logistično regresijo, podporne vektorske stroje in naključni gozd za izboljšano zaznavo vdorov v omrežjih.

1 Introduction

Organizations attempt to address cyber-attacks by removing suspicious activity, threats and risks across network infrastructures [1]. It became commonplace to adopt specialized software defense techniques to keep-up with network data scrutiny [2]. The fresh security threats coupled with enormous traffic transmitted across wireless networks continue to pose new necessities for the machine learning-associated solutions from the standpoints of the physical and functional differences, especially from the network topology and device meta data of numerous computing devices [3]. The use of intrusion detection systems (IDS) is common, though less-effective in case of anomaly or signature-affiliated IDS, which utilizes deep learning and machine learning algorithms to improve performance [4].

There are main weaknesses include: high false alarm, use of predefined threats and attacks, and relatively controlled detection operations. Intrusion prevention system (IPS) performs intrusion-repelling tasks for network and computer system by leveraging on IDS' log files [5]. IPS takes necessary actions upon identifying packet dropping activities and unauthorized addresses. Both approaches are emboldened by certain levels of intelligence for recognizing and classifying abnormal behaviours of network packets [6]. Machine learning and deep learning algorithms play serious functions in providing improved intelligences for the IDS and IPSs during detection and rerouting operations on networks.

In recent time, IDS combines hybrid models, algorithms or approaches with the goal of maximizing their respective advantages and reducing their disadvantages. Hybrid models improve detection accuracy and resilience against different kinds of cyber threats by integrating, for instance, machine learning algorithms with statistical techniques or heuristic approaches. This dual or multifaceted strategy can result in more complete security solutions and improve adaptation to changing attack patterns [7]. The value of hybrid models resides in their capacity to overcome the drawbacks of conventional

single-algorithm systems, which may find it difficult to handle intricate or unusual threats.
In [8], the authors had demonstrated that hybrid approaches not only yield greater detection rates but dramatically reduce false positive rates, which are crucial in preserving the reliability of intrusion detection systems.

The author in [9] revealed that, hybrid models perform better in terms of accuracy and detection rates as well as greater generalization across various datasets like pcap.

The main objectives of this paper include:

- To preproess and optimize binaey-class and multi-class network intrusion datasets from the Kaggle standard repositories using ENN-SMOTE technique.
- To develop four hybrid models for the intrusion traffic detection across wireless networks based on SVM, RF, and LR algorithms using hard voting technique.
- To evaluate the effectiveness of the ensemble models with ENN-SMOTE and Recursive Feature Selection using measures like accuracy, precision, recall, F1-score, confusion matrix, type I error and type II error.

2 Related works

In [10], the authors developed a hybrid Adaptive Ensemble for Intrusion Detection (HAEnID) comprising Stacking Ensemble (SEM), a Bayesian Model Averaging (BMA), and a Conditional Ensemble method (CEM). Using the CIC-IDS 2017, the HAEnID achieved accuracy of 97 to 98%. Further feature selection raises the accuracy to 98.79% for the BMA-M (20). There is the need to lower false alarm rates and increase the reliability especially for multi-class dataset than binary-class.
Sayem et al. [11] proposed an ensemble model comprising the base learner and the meta-learner. The base learner had convolutional neural networks, long short-term memory, and gated recurrent units, and the meta-learner had a deep neural network model. UNSW-15 and CICIDS-2017 datasets were used for validation, which gave accuracy of 90.6% and 99.6% and an F1-score of 90.5% and 99.6% accordingly. More complex features and multi-class dataset can be utilized to test the model.
In 2024, Saheed and Misra [12] developed an ensemble of wolf optimizer (GWO) with a decision tree, random forest, K-nearest neighbor, and multilayer perceptron for intrusion classification tasks. UNSW-NB15, BoT-IoT were used for training and testing, which offered accuracy of 100% for GWO, and 99.9% for DR, Precision of 99.59%, ROC of 99.40%, and False Alarm Rate of 1.5. Future to sample more complex and multi-class datasets.
In 2025, Almania et al. evolved AIDS model composed of Fuzzy c-means clustering, K-Nearest Neighbors (KNN), and weight mapping. The imbalanced datasets issues were overcome with ensemble classifiers Random Forest (RF) and Decision Tree (DT). Network traffic data (DoS, R22, U2R) were curated for the analysis. The AIDS offered accuracy of 97.7% and a false alarm rate of 2.0%. Though, detection rate can be pursued in the future works. Again, the computational complexity and data imbalanced must be handled while more adaptive models can be explored for real-world applications [13].
In [14], the authors introduced stack classifier model for improving the classification of intrusions in IoT datasets. The K-Best feature selection algorithm and ensemble modelling were combined in optimizing essential classification metrics through combination of strengths of the single machine learning models. The Ton IoT dataset was used for validation which provides accuracy of 99.99%, recall of 99.99%, F1-score of 99.98%, and low false positive rates. But, there is the need to consider the different threat types, and network infrastructures.
In 2024, Jemili et al. hybridized Random Forest (RF), XGBoost, and decision trees (DT) for intrusion detection in big data. The model was validated with N-BaIoT, NSL-KDD, and CICIDS2017. The model results attained accuracy of 97% by capitalizing on the consensus of diverse classifiers. There is need to raise the level detection rate for emerging threats [15].
In 2023, Hnamte and Hussain were motivated by deep learning's exceptional performance in various detection and identification tasks by presenting an intelligent and efficient network intrusion detection system (NIDS) based on DL for attack detection. The hybrid of CNN and LSTM model was trained with real-time traffic datasets namely; CICIDS2018 and Edge_IIoT. The performance of the model using multiclass classification achieved a 100% and 99.64% accuracy rates respectively when trained and tested with the datasets. The ensemble model had better performance than single model despite its extended runtime [16].
A hybrid model of convolutional neural network and bidirectional long short-term memory was advanced by Bowen et al. [17] to improve the malicious traffic recognition. Authors validated the proposed model using for CIC-IDS2017, IoT-23 in which accuracy of 98.00% and 99.00% were attained with multi-class and binary-class datasets. However, there is the need to explore more classifiers.
In 2023, Almarshdi et al. combined the convolutional neural network and bidirectional long short-term memory models for inspecting the intrusion traffic inside the UNSW-NB15 dataset. The imbalance in the data was resolved with a synthetic data generation technique. The outcomes revealed that, the hybrid model obtained accuracy of 92.10%. More threats types can be investigated in the future work [18].
The hybrid model of the convolutional neural network and long short-term memory algorithm was implemented by Yassen et al. [19] for multi-class detection tasks such as network intrusion based on secondary datasets (CICID2018 and Edge_IIoT). After training and testing phases, the hybrid model realized accuracy of 100% and 99.64%. However, the ensemble models and local dataset could be implemented subsequently.

The convolutional neural network and GRU algorithms were hybridized to reinforce the performance for intrusion detection in CICIDS-2017 dataset [20]. The results revealed that, the hybrid model produced accuracy of 98.73% and false positive rate of 0.075. Nevertheless, the performance of the hybrid can be tested with complex datasets to measure the accuracy, false alarm, and time elapsed in local settings.

In 2023, Alomari et al. proposed a hybrid algorithm of dense and long short-term memory algorithms for investigating the malware intrusion traffic using dissimilar feature generation strategies. The first case reduced features further by 42.42% from the initial 18.18%, which gave rise to accuracy of 5.84% with tradeoff of 0.07%. The other case minimized the features by 93.50% from the original 81.77%, which realized accuracy of 9.44% with tradeoff of 3.79%. Though, more forms of network intrusion traffic datasets could be included in later studies [21].

The performances of the long short-term memory and GRU model were mixed to from a hybrid model for intrusion detection in Internet of Things networks by [3]. The Harris Hawk optimization and fractional derivative mutation were applied to IoT-23 and ME-WMVEDL datasets for the feature selections. The outcomes indicated that, the hybrid model attained accuracy of 98.125 and 97.34%. More so, the proposed model realized AUC-ROC of 0.9982 and 0.9994 respectively. Though, the class imbalance resampling of dataset was not performed.

Gohari et al. hybridized convolutional neural network and long short-term memory algorithms to detect network traffic using CICAndMal2017. The hybrid classifier eliminates preprocessing data stage which speedups the binary classification tasks. The outcomes showed that, the detection accuracy of 97.79% for binary data, 98.90% for category, and 98.90% for others. It demonstrates the superiority of hybrid models for intrusion detection tasks. More complex data features could be investigated in future [22].

The summary of the related studies reviewed including authors, objectives, methodology, datasets, results and weaknesses as presented in Table 1.

Table 1: Summary of related studies.

Authors	**Objectives**	**Methodology**	**Datasets**	**Results**	**Weaknesses**
Bowen et al. (2023)	Network traffic classification.	Hybrid model of convolutional neural network and bidirectional long short-term memory.	CIC-IDS2017, IoT-23	Accuracy: 98.00%, 99.00%	More classifiers can be investigated.
Almarshdi et al. (2023)	Intrusion traffic inspection.	Hybrid of Convolutional neural network and bidirectional long short-term memory models.	UNSW-NB15	Accuracy: 92.10%	Binary dataset only.
Yassen et al. (2023)	Network intrusion detection.	Hybrid of the convolutional neural network and long short-term memory algorithms.	CICID2018 and Edge_IIoT	Accuracy: 100%, 99.64%.	Multi-class dataset only.
Henry et al. (2023)	Network intrusion detection.	Hybrid of convolutional neural network and GRU algorithms.	CICIDS-2017	Accuracy: 98.73% False positive rate: 0.075	Less complex dataset.
Alomari et al. (2023)	Malware intrusion traffic detection.	Hybrid of dense and long short-term memory algorithms with features reduction.	Malware	Accuracy: 9.44% with tradeoff of 3.79%.	Low datasets features.
Sanju (2023)	IoT intrusion detection.	Hybrid of long short-term memory and GRU model with features selection based on Harris Hawk optimization and fractional derivative mutation.	IoT-23 and ME-WMVEDL	Accuracy: 98.125, 97.34%. AUC-ROC: 0.9982, 0.9994.	Class imbalance resampling not performed.

Gohari et al. (2021)	Network traffic detection.	Hybrid of convolutional neural network and long short-term memory algorithms.	CICAndMal2017	Accuracy: 97.79% - binary data. 98.90% - category data 98.90% - others.	Less complex dataset utilized.
Ahmed et al. (2024)	Intrusion detection	novel Hybrid Adaptive Ensemble: Bayesian Model Averaging, Stacking Ensemble, Conditional Ensemble method.	CIC-IDS 2017	Accuracy: 97-98%. Feature selection: 98.79%.	High false alarm and low detection rate.
Sayem et al. (2024)	Network intrusion detection.	Base learner: Convolutional neural networks, long short-term memory, and gated recurrent units, The meta-learner: deep neural network model	UNSW-15 and CICIDS-2017	Accuracy: 90.6% and 99.6%. F1-score: 90.5% and 99.6%.	Few performance metrics and less complex datasets utilised.
Almotairi et al. (2024)	IoT Intrusion detection.	Stack classifier.	Ton IoT.	Accuracy: 99.99%. Recall: 99.99%, F1-score: 99.98%	Limited threats data.
Saheed & Misra (2024)	IoT intrusion detection.	Ensemble of wolf optimizer (GWO) with a decision tree, random forest, K-nearest neighbor, and multilayer perceptron.	UNSW-NB15, BoT-IoT.	Accuracy - GWO: 100%, DR: 99.9%, Precision: 99.59%, ROC: 99.40%, and FAR: 1.5	Limited multi-class dataset.
Almania et al. (2025)	Network intrusion detection.	AIDS model composed of Fuzzy c-means clustering, K-Nearest Neighbors (KNN), and weight mapping. Data class imbalance with Random Forest (RF) and Decision Tree (DT).	Network traffic: DoS, R22, U2R	Accuracy: 97.7%, False alarm rate: 2.0%.	High computational complexity, data imbalance and low detection rate.
Jemili et al. (2024)	Intrusion detection in big data.	Hybrid of RF, XGBoost, and DT.	N-BaIoT, NSL-KDD, CICIDS2017	Accuracy: 97%	Low detection rate.
Hnamte and Hussain (2023)	Network intrusion detection.	hybrid of CNN and LSTM.	CICIDS2018, Edge_IIoT.	Accuracy: 100% and 99.64%.	Extended runtime.
This paper	Network intrusion detection.	Ensemble models of SVM- RF, SVM- LR, LR-RF, and SVM-LR+RF.	KDD CUP 1999, CSE-CIC-IDS2018.	Accuracy, Precision, F1-Score, Recall, confusion matrix,	Ensemble strategy of hard voting only.

		ENN-SMOTE data preprocessing strategy for binary and multi-class datasets.		AUC-ROC and runtime complexity measured.	

3 Research methodology

The paper attempts to enhance the wants to investigate the following research questions to achieve the objectives set:

i. What are the implications of applying ENN-SMOTE technique on binary-class and multi-class datasets 'accuracy and false positive rates?
ii. Do hybrid models for the intrusion traffic detection across wireless networks for binary-class datasets outperform multi-class datasets after addressing imbalance problems?
iii. Do ensembles model fit into real-world intrusion traffic detection across networks?

Consequently, this study proposes the hybrid strategy to leverage on the advantages of the LR, RF, and SVM based on hard voting strategies in building the ensemble models of RF+SVM, LR+SVM, and LR+RF+SVM. This model aims to improve intrusion detection accuracy by preprocessing the data with ENN-SMOTE technique and fusing the strong classification skills of SVM with the probabilistic insights of LR, and multiple-decision trees of RF as shown in Figure 1.

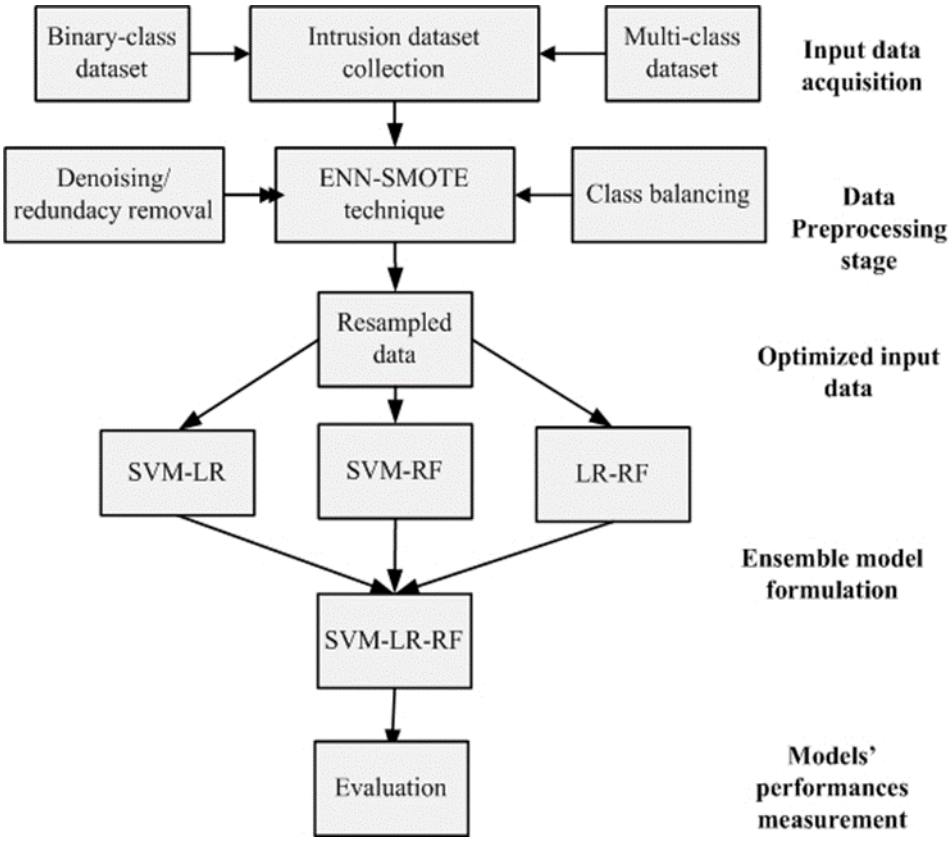

Figure 1. The proposed hybrid models for the network intrusion detection.

From Figure 1, the input data acquisition receives the intrusion datasets from the binary-class and multi-class, which contain begnin and different attacks monitored and logged on the standard repository. The second stage is the data preprocessing to remove noise and redistribute the class-make up in order to address the undersampling or oversampling problems for classifiiers. The ENN-SMOTE resampling technique was adopted for this study. The resampled data is the optimized input data to the proposed ensemble/hybrid models. The study investigates the hybrid models performances with the optimized datasets to determine the influences on the following: SVM-LR, SVM-RF, LR-RF, and SVM-LR-RF. The final stage is the models' performance measurement using standard metrics like accuracy, precision, recall, F1-score, confusion matrix and runtime. The outcomes are used to determine the best classifiers to be implemented in intrusion detection system of enterprises and organization in terms of speed and effectiveness.

Data Collection: The first of the secondary dataset was collected from Kaggle monitored on Wireshark tool. KDD CUP 1999 comprises 41 quantitative and qualitative features of normal and attack data (3 qualitative and 38 quantitative features). The class variable is Normal and Anomalous categories of traffic data for a wide variety of intrusions created in a military network setup. The network intrusion detection is composed of 22544 rows and 41 columns for the Test_data.csv file, and 25192 rows and 42 columns for the Train_data.csv.

The second dataset is made up of 36417 rows and 13columns of URLs from begnin and malicious websites. The URL dataset known as CSE-CIC-IDS2018 contains 7311 of phishing attacks, 7776 of benign, 7930 of Defacement, 6707 of malware attacks, and 6693 of spam attacks from the InfogramALL.csv file.

Data preprocessing: Data cleaning is the foundational step in preparing the dataset for analysis, which involves conversion of the categorical features and removal of the irrelavant features. The StandardScalar () was applied for scaling the features. The scalar.fit () and scalar. transform () were applied on the numeric features to prepare for modeling purposes. This ensures that the dataset remains as complete as possible while minimizing the risk of skewed analysis.

The class imbalance solution was addressed using ENN-SMOTE technique to reduce the dimensionality and noise in both datasets. The dataset dimension before and after the class balancing as follows:

KDD CUP 1999: The original dataset sample class count: int64(1): 13449, int64(0): 11743. Also, the resampled dataset sample class count: int64(1): 13332, int64(0): 13322. The class distributions for the original dataset and the ENN-SMOTE method resampled class distribution are given in Table 2.

Table 2: The outcomes of the ENN-SMOTE technique on KDD CUP 1999.

Class	**Count (Before ENN-SMOTE)**	**Class**	**Count (After ENN-SMOTE)**
Normal	13449	Normal	13332
anomaly	11743	anomaly	13322

Figure 2 shows the graphical representations of the class distributions before and after ENN-SMOTE method's operation on the KDD CUP 1999 dataset, which improvement in the eventual class distribution and noise reduction.

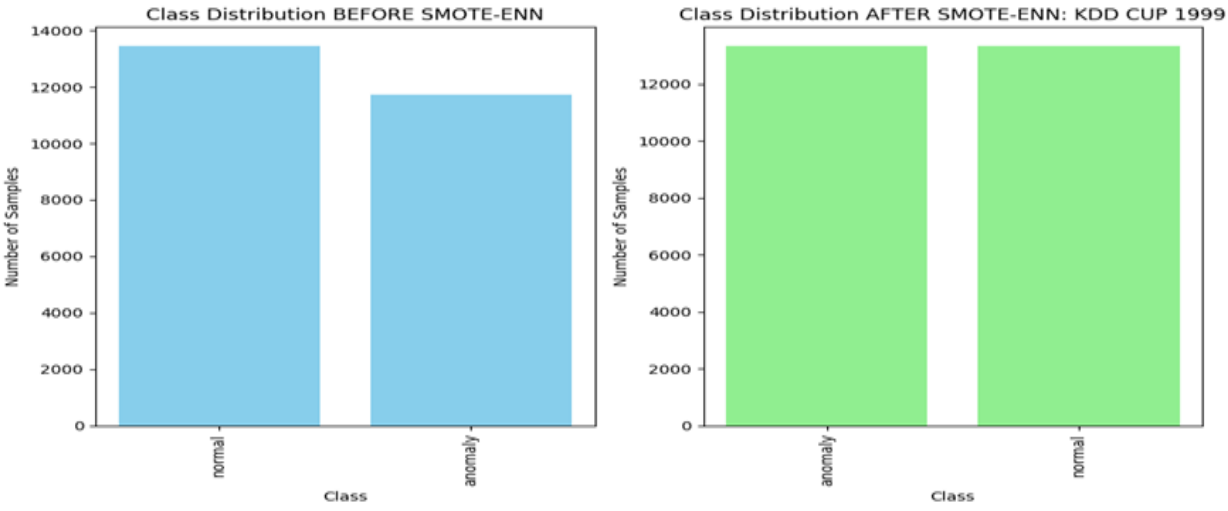

Figure 2: The class distribution for the ENN-SMOTE technique applied on KDD CUP 1999 dataset.

CSE-CIC-IDS2018: The orignal dataset sample class count: int64(0): 7930, int64(1): 7776, int64(3): 7311, int64(2): 6707, int64(4): 6693. Moreso, the resmapled dataset sample class count: int64(4): 7536, int64(2): 7524, int64(0): 7300, int64(1): 7180, int64(3): 5637. The class distributions for the original dataset and the ENN-SMOTE method resampled class distribution are given in Table 3.

Table 3: The outcomes of the ENN-SMOTE technique on CSE-CIC-IDS2018.

Class	**Count (Before ENN-SMOTE)**	**Class**	**Count (After ENN-SMOTE)**
Defacement	7930	Defacement	7300
Benign	7776	Benign	7180
Phishing	7311	Phishing	7524
malware	6707	malware	5637
Spam	6693	Spam	7536

Figure 3 shows the graphical representations of the class distributions before and after ENN-SMOTE method's operation on the CSE-CIC-IDS2018 dataset, which advancement across the multi-class eventual class with minimized noise.

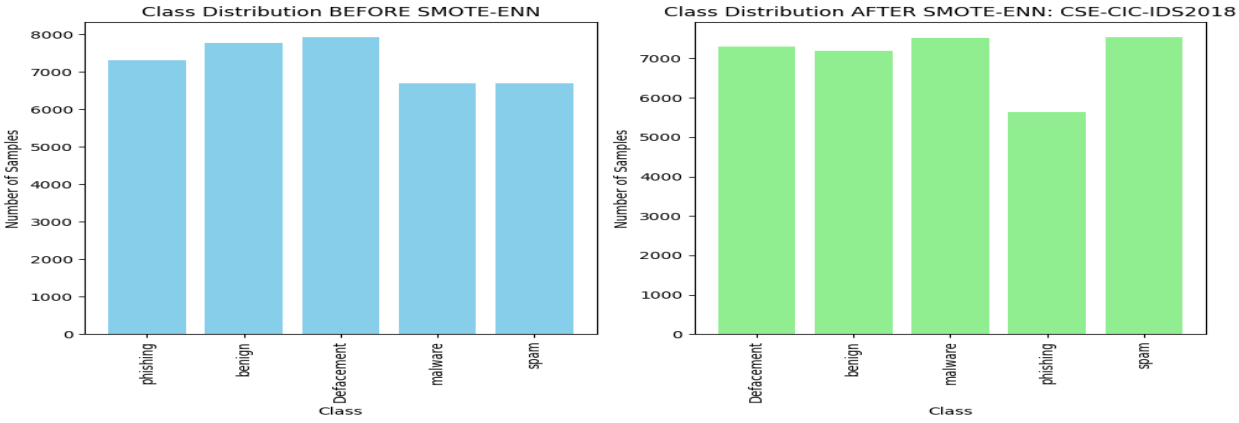

Figure 3: The class distribution for the ENN-SMOTE technique applied on CSE-CIC-IDS2018 dataset.

Programming Language: Python 3.8 or above on Google Colaboratory environment.
Machine Learning Libraries: Scikit-learn, imblearn, time, TensorFlow/PyTorch.
Data Processing Tools: Pandas and NumPy for data manipulation, SciPy for statistical computations.
Visualization Tools: Matplotlib and Seaborn for plotting data and model performance metrics.
Stream Processing Tools: Apache Kafka or Apache Flink for real-time network traffic analysis.
Data format: Microsoft Excel CSV for storing captured traffic logs and network packets.
Hyperparameters setting of the base classifiers:
Support Vector Machine (SVM): kernel = rbf, probability = False, random state=42.
Logistic Regression (LR) = maximum number of iteration =1000, random state=42.
Random Forest Classifier (RF): number of estimators =100, random state=42.
Performance Evaluation Metrics: The hybrid models' performances were assessed using various evaluation metrics, including and runtime, to ensure that it effectively distinguishes between normal and intrusive traffic. The key metrics used in the evaluation process:
Accuracy measures the percentage of correctly classified instances (both normal and malicious traffic) out of the total instances.
Precision evaluates how many of the detected intrusions were actual attacks.
Higher precision means fewer false positives.
Recall (or sensitivity) measures how well the model detects actual intrusions among all instances of intrusion. A high recall indicates the model is good at detecting intrusions but might produce more false positives.
F1 Score is the harmonic mean of precision and recall, providing a balanced metric when the dataset has imbalanced classes.
Area under curve (AOC ROC) is the measures the effectiveness of classification models on binary-class or multi-class datasets.
Confusion matrix provides a comprehensive view of the model's performance, showing the number of true positives, true negatives, false positives, and false negatives.
Runtime measure the time taken to perform model training and validation in seconds of the CPU clock.
The comparisons of the proposed ensemble models outcomes after preprocessing of dataset with ENN-SMOTE class imbalance technique against the optimal features within the dataset based Recursive Features Selection (RFE) techniques. The Related Samples Wilcoxon Signed Rank Test used to measure the significance of the outcomes attained by this study.

4 Results and discussion

Cyber-threats continue to evolve in complexity and frequency; organizations face significant challenges in protecting their networks from malicious activities. Traditional security measures often fall short in detecting sophisticated attacks, leading to severe data breaches and financial losses. This paper attempts to provide an effective network intrusion detection system (NIDS) for enterprises, which is capable of identifying both normal and abnormal network traffic, allowing for timely responses to potential threats. The primary challenge lies in developing a machine learning-based models such as SVM-LR, SVM-RF, LR-RF, and SVM-LR-RF ensemble models. Two resampled binary-class and multi-class network intrusion datasets whose classification performances scores are presented in the proceeding subsections.

4.1 Support vector machine and logistic regression ensemble model

Table 4 presents the outcomes of the SVM+LR ensemble model for the detection of intrusion traffic composed of binary-class and multi-class acquired the Kaggle data repository. The performance of SVM+LR ensemble model measured with accuracy, precision, recall, F1-score, AUC ROC and confusion matrices for the both datasets.

Table 4: The outcomes of the SVM+LR Ensemble model.

Dataset	Accuracy	Precision	Recall	F1-score	AUC ROC	Confusion matrix
KDD Cup 1999	0.9803	0.9890	0.9685	0.9786	0.9795	[[3997 38] [111 3412]]
CSE-CIC-IDS2018	0.8556	0.9906	0.8242	0.8998	0.89.77	[[2266 67] [1511 7082]]

From Table 4, the SVM+LR ensemble model with ENN-SMOTE resampled KDD Cup 1999 was best for accuracy of 98.03%, recall of 96.85%, F1-score of 97.86% against the resampled CSE-CIC-IDS2018 data, which performed well for precision of 99.06%. These outcomes can be attributed to denoising and class-balancing undertaken by the study. Moreso, the SVM-LR model's performance with type I error of 2.70% and type II error of 1.10% are acceptable for the missing on the positive and negative for the KDD Cup 1999 dataset. Whereas the SVM-LR model achieved type I error of 40.00% and type II error of 0.94% as acceptable for misprediction of the positive and negative classes (acceptable) in the CSE-CIC-IDS2018 dataset. Therefore, the research question can be answered in the affective as ENN-EMOTE improved and optimized the datasets.

4.2 Support vector machine and random forest ensemble model

Table 5 presents the outcomes of the SVM-RF ensemble model for the detection of intrusion traffic comprising of acquired binary-class and multi-class datasets from the Kaggle. The performance of SVM+LR ensemble model measured with accuracy, precision, recall, F1-score, AUC ROC and confusion matrices for the both datasets.

Table 5: The outcomes of the SVM+RF Ensemble model.

Dataset	Accuracy	Precision	Recall	F1-score	AUC ROC	Confusion matrix
KDD Cup 1999	0.9942	0.9991	0.9884	0.9937	0.9937	[[4032 3] [41 3482]]
CSE-CIC-IDS2018	0.9141	0.99681	0.8936	0.9424	0.9415	[[2308 25] [914 7679]]

From Table 5, the SVM+RF ensemble model with ENN-SMOTE resampled KDD Cup 1999 offered the highest accuracy of 99.42%, precision of 99.91, recall of 98.84%, F1-score of 99.37%, and overtook the outcomes with the resampled CSE-CIC-IDS2018 data. These outcomes can be attributed to denoising and class-balancing as most suitable for the binary-class data. The SVM-RF model generated the type I error of 1.00% and type II error of 0.09% as mispredictions for the positive class and negative class (acceptable) in the KDD Cup 1999 dataset. Similarly, the SVM-RF model scored the type I error of 28.37% and type II error of 0.32% as the mispredictions of the positive class (unacceptable) and negative class (acceptable) in the CSE-CIC-IDS2018 dataset. Therefore, this answers the research question that the ENN-EMOTE technique improved and optimized the dataset for high performance of the model.

4.3 Logistic regression and random forest ensemble model

Table 6 presents the outcomes of the LR-RF ensemble model for the detection of intrusion traffic comprising of acquired binary-class and multi-class datasets available at the Kaggle. The performance o LR-RF ensemble model measured with accuracy, precision, recall, F1-score, AUC ROC and confusion matrices for the both datasets.

Table 6. The outcomes of the LR+RF Ensemble model

Dataset	Accuracy	Precision	Recall	F1-score	AUC ROC	Confusion matrix
KDD Cup 1999	0.9843	0.9994	0.9668	0.9828	0.9831	[[4033 2] [117 3406]]
CSE-CIC-IDS2018	0.8688	0.9967	0.8360	0.9093	0.9129	[[2309 24] [1409 7184]]

From Table 6, the LR+RF ensemble model with ENN-SMOTE resampled KDD Cup 1999 offered the highest accuracy of 98.43%, precision of 99.94, recall of 96.68%, F1-score of 98.28%, and then trailed by the outcomes from the resampled CSE-CIC-IDS2018 data. These outcomes can be attributed to denoising and class-balancing was highly effective for the binary-class data during the model validation. The LR-RF model generated the type I error of 0.87% and type II error of 1.12% as mispredictions for the positive class and negative class (acceptable) in the KDD Cup 1999 dataset. Similarly, the LR-RF model scored the type I error of 26.01% and type II error of 0.94% as mispredictions of the positive class (unacceptable) and negative class (acceptable) in the CSE-CIC-IDS2018 dataset. Therefore, the answer to the research question is the ENN-EMOTE technique improved and optimized the dataset for both binary-class and multi-class classification tasks of the LR-RF ensemble model.

4.4 The proposed ensemble model

Table 7 shows the outcomes of the SVM-LR-RF ensemble model validated with the detection of intrusion traffic datasets acquired the Kaggle. The performance of SVM-LR-RF ensemble model measured with accuracy, precision, recall, F1-score, AUC ROC and confusion matrices for both binary-class and multi-class datasets.

Table 7: The outcomes of the SVM-LR-RF ensemble model.

Dataset	Accuracy	Precision	Recall	F1-score	AUC ROC	Confusion matrix
KDD Cup 1999	0.9902	0.9889	0.9901	0.9895	0.9902	[[3996 39] [35 3488]]
CSE-CIC-IDS2018	0.9206	0.9906	0.9076	0.9473	0.9379	[[2259 74] [794 7799]]

From Table 7, the proposed ensemble model validated using the ENN-SMOTE resampled KDD Cup 1999 offered the largest accuracy of 99.02%, recall of 99.01%, F1-score of 98.95%, and then trailed by the outcomes from the resampled CSE-CIC-IDS2018 data except for precision of 99.06%. The outcomes are possible due the denoising and class-balancing processes offered ENN-SMOTE method during preprocessing phase. Also, the binary-class data are best fitted for proposed ensemble model. The SVM-LR-RF model generated the type I error of 0.87% and type II error of 1.12% as mispredictions for the positive class and negative class (acceptable) in the KDD Cup 1999 dataset. Similarly, the SVM-LR-RF model scored the type I error of 26.01% and type II error of 0.94% as mispredictions of the positive class (unacceptable) and negative class (acceptable) in the CSE-CIC-IDS2018 dataset. The answer to the research question is that the ENN-EMOTE technique improved and optimized the dataset when applied at preprocessing phase in either case of dataset.

4.5 Models' validation performances compared

The models were trained and tested with 70% and 30% portions of the original intrusion datasets acquired from KDD Cup 1999 and CSE-CIC-IDS2018 available at the Kaggle. The prediction outcomes of the various hybrid models and mispredictions errors for intrusion detection are given in Table 8.

Table 8: Summary of the ensemble models' performances with KDD Cup 1999 dataset.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC (%)	Type I Error (%)	Type Error II (%)	Runtime (s)
Without ENN-SMOTE technique								
SVM+LR+ RF	99.17	99.15	99.06	99.11	99.16	0.82	0.85	9.89
SVM + RF	99.43	99.89	98.89	99.39	99.40	0.96	0.11	7.75
SVM + LR	97.75	99.15	96.00	97.55	97.64	0.35	0.85	10.15
LR + RF	98.12	99.91	96.05	97.95	97.99	3.33	0.09	4.90
With ENN-SMOTE technique								
SVM + LR + RF	99.02	98.89	99.01	98.95	99.02	0.89	1.11	8.93
SVM + RF	99.42	99.91	98.84	99.37	99.37	1.01	0.09	5.99
SVM + LR	98.03	98.90	96.85	97.86	97.95	2.70	1.10	8.07
LR + RF	98.43	99.94	96.68	98.28	98.31	2.82	0.06	2.83

The results in Table 8 indicated that, focusing on the binary-class dataset (KDD Cup 1999), the SVM-RF obtains highest accuracy of 99.42%, followed by SVM-LR-RF at 99.02%, SVM-LR at 98.03%, and LR-RF at 98.43%. Precision score of 99.94% was achieved by the LR-RF, and trailed by the SVM-RF at 99.91%, the SVM-LR at 98/90%, and LR-RF at 96/88%. With the recall measure, the SVM-LR-RF ensemble model attained the highest score of 99.01%, following is the SVM-RF model at 98.84%, SVM-LR at 96.85%, and LR-RF at 96.68%. The F1-score of 99.37% was obtained by SVM-RF model, the biggest margin, outperforming SVM-LR-RF at 98.95%, LR-RF at 98.28%, and 97.86%. The misclassification performance of the SVM-LR-RF model was best with type I error of 0.89% and type II error of 1.11%.

Table 9 shows the performances of the ensemble models using CSC-CIC-IDS2018 dataset for the various measures.

Table 9. Summary of the ensemble models' performances with CSE-CIC-IDS2018 dataset.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC ROC (%)	Type I Error (%)	Type Error II (%)	Runtime (s)
Without ENN-SMOTE technique								
SVM + LR+ RF	94.11	96.04	96.49	96.26	90.91	13.17	3.96	31.98
SVM + RF	95.91	99.35	95.43	97.35	96.56	14.71	0.65	31.92
SVM + LR	90.57	95/96	91.88	93.88	88.82	8.48	4.04	23.90
LR + RF	93.53	99.32	92.40	95.74	95.04	22.27	0.68	5.64
With ENN-SMOTE technique								
SVM + LR+ RF	92.06	99.06	90.76	94.73	93.79	26.01	0.94	37.23
SVM + RF	91.41	99.68	89.36	94.24	94.15	28.37	0.32	34.26
SVM + LR	85.56	99.06	82.42	89.98	89.77	40.01	0.94	28.54
LR + RF	86.88	99.67	83.60	90.93	91.29	37.90	0.33	9.64

On the other hand, Table 9 presents the multi-class dataset (CSE-CIC-IDS2018) resampled with the ENN-SMOTE method gave the high accuracy score of 92.06%, immediate undertaken by SVM-RF at 91.41%, LR-RF at 86.88%, and lastly by SVM-LR model at 85.56%. In terms of precision measure, SVM-RF got 99.68% to place best, LR-RF (99.67%); while SVM-RF and LR-RF-SVM at 99.06% were the joint lowest. Considering the recall score, the SVM-LR-RF model had the biggest value of 90.76%, closely underperformed by SVM-RF at 89.36%, LR-RF at 83.60%, and SVM-LR at 82.42%. The highest F1-score of 94.73% was computed for the SVM-LR-RF model, and the rest following: SVM-RF (94.24%), LR-RF (90.93%), and SVM-LR (89.98%). In terms of misprediction scores showed that, the SVM-LR-RF model produces the least type I error of 26.01% and type II error of 0.94%, which implies the high classification rate.

The various dataset sizes (10-folds) and matching time complexity measured for the two datasets. The distinct ensemble models were trained and validated using 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90% of the KDD Cup 1999 data as presented in Figures 2.

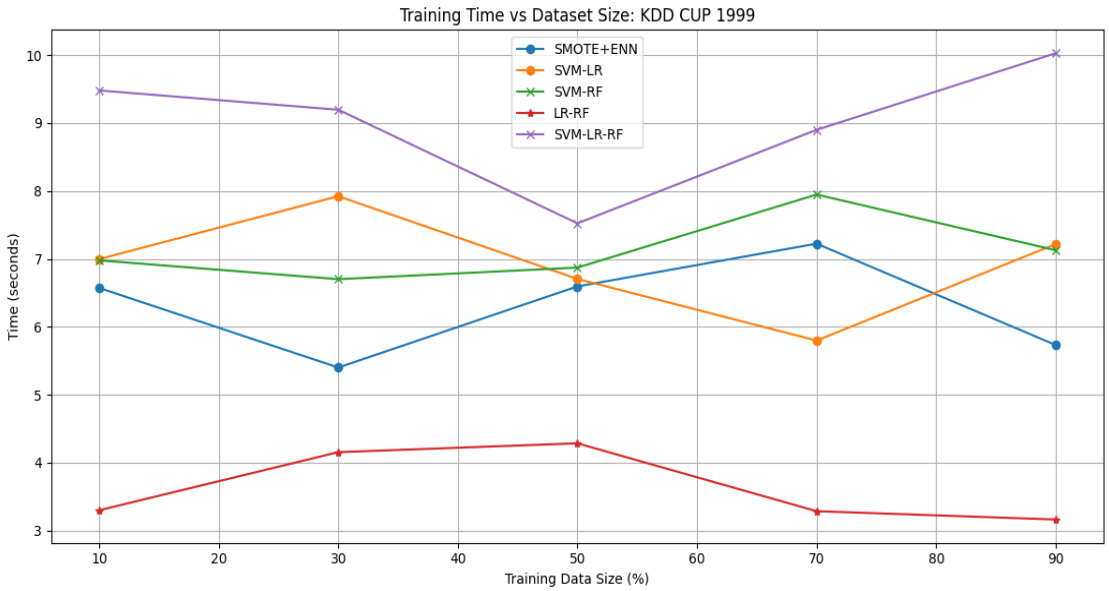

Figure 2: The ensemble models time complexity measured during training and validation phases with KDD CUP 1999 dataset.

In Figure 2, the SVM-LR-RF model commenced the training and validation phases at runtime of 9.5s at 10% of data, then decrease to lowest runtime of 7.5s at 50% of the data. Thereafter, the SVM-LR-RF model's runtime increases with additional data to peak at 10s upon full data usage. On the opposite, the LR-RF model took lowest runtime of 2s at 10% of data and increases until 50% data usage at runtime of 4.2s before sliding to 3.1s upon full data training and validation phases.

The several ensemble models were trained and validated with 10-fold sizes of the CSE-CIC-IDS2018 dataset are presented in Figures 3.

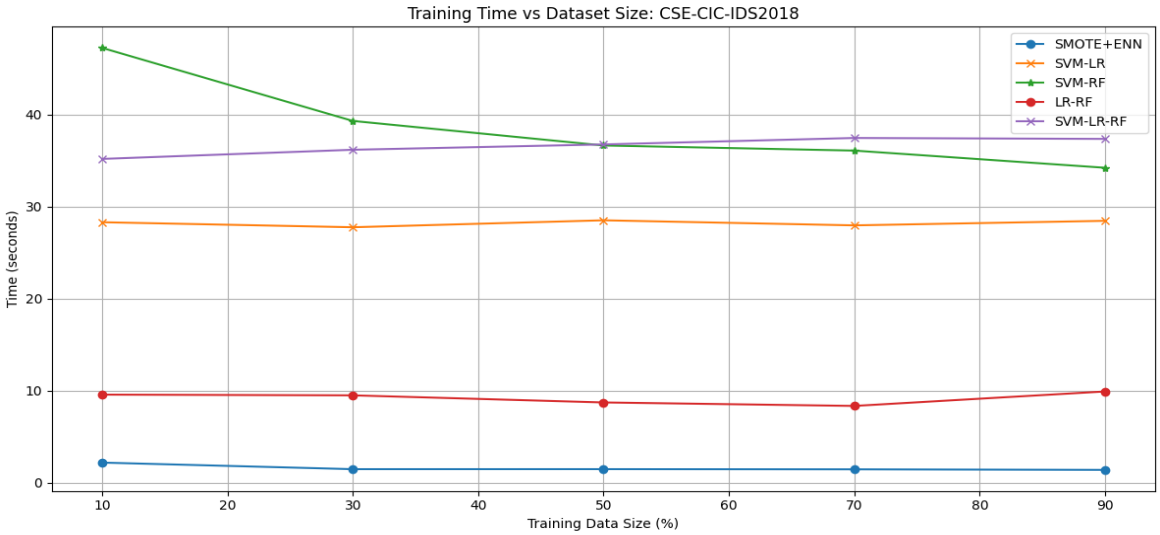

Figure 3: The ensemble models time complexity measured during training and validation phases with CSE-CIC-IDS2018 dataset.

From Figure 3, the SVM-LR-RF model started the training and validation phases at runtime of 35s at 10% of data, then increases steadily until 90% of data with runtime of 38s. Similar to the tends presented in Figure 2 for the LR-RF model, the runtime of 28s was observed at the 1st fold of data, and relatively steady after 9th fold of the data at 28.8s to terminate model's training and validation phases.

4.6 Discussion of findings

The paper introduced the ENN-SMOTE technique to address overfitting, noise reduction and class distribution balancing. This approach raises the hybrid ensemble models' performances in the cases of binary-class and multi-class data used for the validation phases. The SVM-LR-FR model offers more superior interpretability, less computational overhead, and better generalizability for multi-class datasets than binary-class data in case of intrusion detection tasks. It was found that, multi-class datasets require more corporation and synergy of the weak learners within the hybrid/ensemble models; thereby minimizing the potential redundancy as it is the case of the binary-class due to lower noise and fewer class distribution using ENN-SMOTE method.

The misclassification performance of the SVM-LR-RF model achieves the type I error of 0.89% and type II error of 1.11% for KDD Cup 1999 dataset, and the type I error of 26.01% and type II error of 0.94% for CSE-CIC-IDS2018 dataset. These imply that, the effectiveness of the ensemble models with larger size of weak classifiers for different resampled class of datasets.

The runtime complexity of training and validation of the SVM-LR-RF model was largest at 38s followed by the LR-RF model at the runtime of 28.8s using CSE-CIC-IDS2018 dataset. The opposite was the case of KDD CUP 1999 dataset in that, the LR-RF model had the fastest runtime of 3.1s, and the SVM-LR-RF model's runtime was biggest at 10s. These explain the influence of data size on the runtime performances of the ensemble models.

The Wilcoxon Signed Ranks Test measures the related samples from the RFE-based data preprocessing approach (bet) against ENN-SMOTE technique. The samples were drawn from the classification results, and type I error, and type II error computed. The Asymptotic Significance (2-sided test) p-value = 0.043, which less than the significance level of 0.05. This implies that, the median differences between the data samples of RFE and ENN-SMOTE techniques are NOT equal to 0. The null hypothesis is rejected. The paper established that, that the ensembles models with ENN-SMOTE performance increases caused significant differences.

5 Conclusion

The paper established four hybrid models by mixing SVM-RF, SVM-LR, LR-RF, and SVM-LR-RF models to raise the accuracy and minimize rate of false-alarms during the intrusion detection on enterprise networks. The two standard intrusion datasets belong to binary-class and multiclass, then preprocessed with ENN-SMOTE imbalance class and denoising solution. The resampled datasets were used for the training and testing of ensemble

models developed. The results indicated that, focusing on the binary-class dataset (KDD Cup 1999), the SVM-RF obtains highest accuracy of 99.42%, followed by SVM-LR-RF at 99.02%. Precision of 99.94% for LR-RF was highest, and the SVM-RF at 99.91% in second place. For the recall measure, the SVM-LR-RF model attained 99.01% and the SVM-RF model at 98.84%. The F1-score of 99.37% for the SVM-RF model, and SVM-LR-RF at 98.95% were highest.

On the contrary, the multi-class dataset (CSE-CIC-IDS2018) resampled with the ENN-SMOTE method offered accuracy of 92.06%, and SVM-RF at 91.41%. The precision score of 99.68% was highest for the SVM-RF, before the LR-RF at 99.67%. Focusing on the recall measure, the SVM-LR-RF model had the widest margin of 90.76%; thereafter the SVM-RF at 89.36% was closest. The F1-score of 94.73% for the SVM-LR-RF model was biggest before SVM-RF (94.24%).

The SVM-LR-RF model achieves the lowest misprediction for type I error of 0.89% and type II error of 1.11% for KDD Cup 1999 dataset. The same trend of misprediction for the type I error of 26.01% and type II error of 0.94% in CSE-CIC-IDS2018 dataset. The ensembles models' performances were largely satisfactory before of their relative high margin of values computed for the evaluation metrics. The reasons are due to the hard voting in which majority vote from the sets of decision trees from the single models when the final decision are made during classification tasks. Also, the ENN-SMOTE method is excellent for removing noise and balancing classes of datasets. Therefore, the SVM-RF and SVM-LR-RF ensemble models are the most effective classifier when dealing both the binary-class and multi-class data for the network intrusion detection.

These findings suggest that the proposed models not only enhanced the intrusion traffic detection rates but also adapts well with the complexities of real-world network traffic behaviours. By integrating the proposed ensemble models for advanced intrusion detection systems, organizations can significantly bolster their cybersecurity posture against increasingly sophisticated threats rather data feature redundancy removal. Future works could explore other high-performance classifiers, SMOTE techniques and more complex dataset to enhance detection capabilities and computational complexity of the proposed ensemble models. Again, it may interest to investigate the effect of combining the data preprocessing approaches of SMOTE and RFE techniques on performances of the hybrid models.

Code Availability
Authors shall provide the source codes utilized in the paper upon reasonable request-only.

References

[1] U. Dixit, S. Bhatia, and P. Bhatia, "Comparison of different machine learning algorithms based on intrusion detection system," in *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, IEEE, 2022, pp. 667–672, doi: 10.1109/com-it-con54601.2022.9850515.

[2] S. K. Shandilya, S. Upadhyay, A. Kumar, and A. K. Nagar, "AI-assisted Computer Network Operations testbed for Nature-Inspired Cyber Security based adaptive defense simulation and analysis," *Future Generation Computer Systems*, vol. 127, pp. 297–308, 2022, doi: 10.1016/j.future.2021.09.018.

[3] P. Sanju, "Enhancing intrusion detection in IoT systems: A hybrid metaheuristics-deep learning approach with ensemble of recurrent neural networks," *Journal of Engineering Research*, vol. 11, no. 4, pp. 356–361, 2023, doi: 10.1016/j.jer.2023.100122.

[4] K. A. Dhanya, S. Vajipayajula, K. Srinivasan, A. Tibrewal, T. S. Kumar, and T. G. Kumar, "Detection of network attacks using machine learning and deep learning models," *Procedia Comput Sci*, vol. 218, pp. 57–66, 2023, doi: 10.1016/j.procs.2022.12.401.

[5] M. Landauer, F. Skopik, M. Frank, W. Hotwagner, M. Wurzenberger, and A. Rauber, "Maintainable log datasets for evaluation of intrusion detection systems," *IEEE Trans Dependable Secure Comput*, vol. 20, no. 4, pp. 3466–3482, 2022, doi: 10.1109/tdsc.2022.3201582.

[6] F. Zola, L. Segurola-Gil, J. L. Bruse, M. Galar, and R. Orduna-Urrutia, "Network traffic analysis through node behaviour classification: a graph-based approach with temporal dissection and data-level preprocessing," *Comput Secur*, vol. 115, p. 102632, 2022, doi: 10.1016/j.cose.2022.102632.

[7] K. G. Reddy and P. S. Thilagam, "trust-based hybrid ids for rushing attacks in wireless mesh Networks," in *Recent Advances in Computer Based Systems, Processes and Applications*, CRC Press, 2020, pp. 49–57, doi: 10.1201/9781003043980-7.

[8] T. Wisanwanichthan and M. Thammawichai, "A double-layered hybrid approach for network intrusion detection system using combined naive bayes and SVM," *IEEE access*, vol. 9, pp. 138432–138450, 2021, doi: 10.1109/access.2021.3118573.

[9] P. Yadav and S. C. Sharma, "Unveiling the cutting edge: a comprehensive survey of localization techniques in WSN, leveraging optimization and machine learning approaches," *Wirel Pers Commun*, vol. 132, no. 4, pp. 2293–2362, 2023, doi: 10.1007/s11277-023-10630-x.

[10] U. Ahmed *et al.*, "Explainable AI-based innovative hybrid ensemble model for intrusion detection," *Journal of Cloud Computing*, vol. 13, no. 1, p. 150, 2024, doi: 10.1186/s13677-024-00712-x.

[11] I. M. Sayem, M. I. Sayed, S. Saha, and A. Haque, "ENIDS: a deep learning-based ensemble framework for network intrusion detection systems," *IEEE transactions on network and service management*, 2024, doi: 10.1109/tnsm.2024.3414305.

[12] Y. K. Saheed and S. Misra, "A voting gray wolf optimizer-based ensemble learning models for intrusion detection in the Internet of Things," *Int J Inf Secur*, vol. 23, no. 3, pp. 1557–1581, 2024, doi: 10.1007/s10207-023-00803-x.

[13] M. Almania, A. Zainal, F. A. Ghaleb, A. Alnawasrah, and M. Al Qerom, "Adaptive Intrusion Detection System with Ensemble Classifiers for Handling Imbalanced Datasets and Dynamic Network Traffic," *Journal of Robotics and Control (JRC)*, vol. 6, no. 1, pp. 114–123, 2025.

[14] A. Almotairi, S. Atawneh, O. A. Khashan, and N. M. Khafajah, "Enhancing intrusion detection in IoT networks using machine learning-based feature selection and ensemble models," *Systems Science & Control Engineering*, vol. 12, no. 1, p. 2321381, 2024, doi: 10.1080/21642583.2024.2321381.

[15] F. Jemili, R. Meddeb, and O. Korbaa, "Intrusion detection based on ensemble learning for big data classification," *Cluster Comput*, vol. 27, no. 3, pp. 3771–3798, 2024, doi: 10.1007/s10586-023-04168-7.

[16] V. Hnamte and J. Hussain, "DCNNBiLSTM: An efficient hybrid deep learning-based intrusion detection system," *Telematics and Informatics Reports*, vol. 10, p. 100053, 2023, doi: 10.1016/j.teler.2023.100053.

[17] B. Bowen, A. Chennamaneni, A. Goulart, and D. Lin, "BLoCNet: a hybrid, dataset-independent intrusion detection system using deep learning," *Int J Inf Secur*, vol. 22, no. 4, pp. 893–917, 2023, doi: 10.1007/s10207-023-00663-5.

[18] R. Almarshdi, L. Nassef, E. Fadel, and N. Alowidi, "Hybrid Deep Learning Based Attack Detection for Imbalanced Data Classification.," *Intelligent Automation & Soft Computing*, vol. 35, no. 1, 2023, doi: 10.32604/iasc.2023.026799.

[19] M. S. Yassen, A. A. Raghdah, and A. B. Mohammed, "Employing hybrid ANOVA-RFE with machine and deep learning models for enhanced IoT and IIoT attack detection and classification," *Ingenierie des Systemes d'Information*, vol. 28, no. 4, p. 1003, 2023, doi: 10.18280/isi.280420.

[20] A. Henry *et al.*, "Composition of hybrid deep learning model and feature optimization for intrusion detection system," *Sensors*, vol. 23, no. 2, p. 890, 2023, doi: 10.3390/s23020890.

[21] E. S. Alomari *et al.*, "Malware detection using deep learning and correlation-based feature selection," *Symmetry (Basel)*, vol. 15, no. 1, p. 123, 2023, doi: 10.3390/sym15010123.

[22] M. Gohari, S. Hashemi, and L. Abdi, "Android malware detection and classification based on network traffic using deep learning," in *2021 7th International Conference on Web Research (ICWR)*, IEEE, 2021, pp. 71–77, doi: 10.1109/icwr51868.2021.9443025.