Halimah Muhammad Tukur, 2026, 14:4
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

An Explainable Stacking Ensemble Machine Learning Model For Denial-Of-Service Attack Detection

Halimah Muhammad Tukur[1], Jamilu Awwalu[1], Salim Ahmad[2], Ayuba John[2]

[1]Computer Science Department, Federal University Dutse, Duts, Nigeria

[2]Information Technology Department, Federal University Dutse, Dutse, Nigeria

Abstract- The evolving sophistication of cyber threats, particularly denial-of-service (DoS) and distributed denial-of-service (DDoS) attacks, poses significant challenges to the modern cybersecurity systems, availability, and reliability of digital infrastructures. These attacks often overwhelm network resources, leading to service disruptions and potential data breaches. Traditional intrusion detection systems (IDS) have been developed to mitigate such threats; however, they often struggle with scalability, detection accuracy, and interpretability, especially when processing large-scale network traffic. To address these limitations, this study proposes a stacked ensemble intrusion detection model integrating Decision Tree (DT), Random Forest (RF), and Artificial Neural Network (ANN) classifiers as base learners, with Logistic Regression (LR) serving as the meta-learner. The framework aims to enhance detection accuracy, interpretability, and robustness against DoS and DDoS attacks. Using the CIC-IDS2017 dataset, Mutual Information (MI) was employed for optimal feature selection, ensuring the most relevant attributes were retained for training. SHAP (Shapley Additive exPlanations) was applied to interpret the contribution of each selected feature to the model's decision-making process. Experimental results demonstrate that the proposed ensemble achieves superior performance compared to individual classifiers and baseline models, with 98.83% accuracy, a marginal precision trade-off (96.37% vs. 96.50% baseline), 99.19% recall, 97.76% F1-score, and an AUC of 99.95%. Compared with the baseline [1], recall improved by 3.39% and F1-score by 1.57%. SHAP analysis identified flow-based and packet-based features as key contributors in distinguishing normal and malicious traffic. Overall, combining multiple learning algorithms within a stacked ensemble, along with MI feature selection and SHAP explainability, provides a robust, transparent, and effective solution for real-time intrusion detection.

Keywords: denial-of-Service; ensemble learning; stacking; random forest; artificial neural network; decision tree; mutual information; CIC-IDS2017.

I. INTRODUCTION

The rapid expansion of digital services, cloud computing, and large-scale enterprise networks has fundamentally transformed modern information systems. As organizations increasingly rely on interconnected infrastructures to store, transmit, and process sensitive data, exposure to cyber threats has grown significantly. Denial-of-Service (DoS) and Distributed Denial-of-Service (DDoS) attacks remain among the most disruptive threats, exhausting network, server, or application resources and rendering critical services unavailable [3], [4]. These attacks exploit volumetric flooding, protocol vulnerabilities, and application-layer weaknesses, often amplified through botnets, resulting in substantial financial losses, operational downtime, and reputational damage [3], [7].

Traditional signature- and rule-based Intrusion Detection Systems (IDS) have limited capability in detecting zero-day attacks and sophisticated variants that mimic legitimate traffic patterns [2], [8]. The

increasing complexity and scale of modern cyberattacks demand adaptive and intelligent security mechanisms. Machine learning (ML) has therefore emerged as an essential approach for DoS detection, enabling systems to identify discriminative traffic patterns from historical data and generalize to unseen attack variants [1], [2]. Early single-classifier models—including Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVM), and Artificial Neural Networks (ANN)—have demonstrated improved detection performance over traditional methods. However, these models often suffer from dataset imbalance, overfitting, and computational inefficiencies when applied to large-scale network traffic [8], [21].

To overcome these limitations, ensemble learning techniques, particularly stacking, have been proposed to combine complementary base classifiers under a meta-learner, enhancing predictive accuracy and robustness [9]-[12]. Stacking ensembles exploit the interpretability of tree-based models, the robustness of Random Forests, and the expressive power of neural networks while reducing false positives and false negatives [5], [12]. Scalability remains a challenge, as high-volume network telemetry can overwhelm conventional processing pipelines. Efficient traffic analysis and ensemble-based learning techniques have been proposed to improve detection performance and support deployment in large-scale network environments [21].

Interpretability is equally critical, as complex ML models often function as "black boxes," which limits analyst trust and operational adoption. Explainable Artificial Intelligence (XAI) techniques, such as SHapley Additive exPlanations (SHAP), provide quantitative insights into feature contributions and global model behavior, enhancing transparency in IDS applications [1], [5], [17], [18]. Benchmark datasets such as CIC-IDS2017 offer realistic traffic distributions and diverse DoS attack scenarios including Hulk, GoldenEye, Slowloris, and LOIC—supporting rigorous evaluation of ML-based detection systems [10], [13].

To address the identified research gaps, this study proposes an explainable stacking ensemble intrusion detection framework that integrates Decision Tree (DT), Random Forest (RF), and Artificial Neural Network (ANN) classifiers as base learners, with Logistic Regression serving as the meta-classifier. Mutual Information is employed to identify the most informative features, while SHAP is incorporated to provide transparent explanations of model predictions. By combining heterogeneous learning algorithms with explainable artificial intelligence techniques, the proposed framework aims to improve detection accuracy, enhance model interpretability, and support scalable deployment in modern network environments.

Despite these advancements, research gaps remain in integrating stacking ensembles, scalable deployment mechanisms, and explainable AI within a unified detection framework. Unlike existing studies that focus primarily on either ensemble accuracy or model interpretability, this work proposes **a** novel framework that combines DT, RF, and ANN base learners with a Logistic Regression meta-classifier, Mutual Information-based feature selection, and SHAP-driven explainability. This integrated approach enables not only high detection accuracy but also transparency and scalability, making it suitable for deployment in large-scale enterprise networks [1], [9], [12]. By unifying these elements, the proposed system addresses key limitations in previous IDS research and provides a practical, robust, and interpretable solution for modern cyber threat environments.

II. RELATED WORKS

The detection of Denial-of-Service (DoS) and Distributed Denial-of-Service (DDoS) attacks has increasingly relied on machine learning (ML) techniques due to their adaptability to evolving threat patterns [8], [12]. Traditional intrusion detection systems are limited in detecting zero-day attacks due to their reliance on known patterns and inability to generalize to unseen threats [8]. Modern DDoS attacks exploit network and application vulnerabilities, often leveraging IoT-based botnets to amplify traffic [3], [7]. The expansion of cloud computing infrastructures and 5G-enabled networks further increases the attack surface, necessitating scalable and adaptive detection mechanisms [4]. Supervised ML models such as Decision Trees (DT), Random Forests (RF), and Artificial Neural Networks (ANN) are widely applied in DoS detection. DT models are computationally efficient and interpretable but may overfit in high-dimensional or imbalanced datasets. RF improves robustness by aggregating multiple trees, reducing variance, and providing feature importance rankings that assist in analyzing network traffic [2]. ANN approaches capture complex non-linear relationships and often achieve high detection accuracy; however, they require significant computational resources and are less interpretable [11].

Hybrid and ensemble-based models have been proposed to overcome the limitations of single classifiers. Models that integrate deep learning with ensemble classifiers, such as the NIDS-CNNRF framework, combine CNN-based feature extraction with RF classification. They use ADASYN oversampling to mitigate class imbalance and Principal Component Analysis (PCA) for dimensionality reduction [14]. Evaluated on benchmark datasets including KDD CUP99, NSL-KDD, CIC-IDS2017, and CIC-IDS2018, these models have achieved detection accuracies exceeding 99%, alongside strong precision and recall metrics [14].

Stacking ensembles represent a more advanced strategy, combining heterogeneous base learners via a meta-classifier. Unlike bagging or boosting, stacking leverages complementary strengths of diverse models to improve predictive performance [12], [9]. Such frameworks have been shown to enhance AUROC, F1-score, and generalization on imbalanced DoS datasets. Despite these improvements, many machine learning-based intrusion detection systems primarily emphasize detection accuracy, while interpretability and scalability remain key challenges in practical deployment [16]. Explainable Artificial Intelligence (XAI) techniques have become increasingly important in IDS research [5], [17]. Methods such as SHAP and LIME provide feature-level explanations, improving model interpretability and analyst trust [1], [5]. In particular, SHAP has been effectively applied to tree-based models such as Decision Trees and Random Forests to enhance explainability in intrusion detection systems [15]. The integration of XAI techniques into machine learning-based intrusion detection systems enhances model transparency and interpretability, improving trust and supporting operational deployment [5], [15]. Recent advances in Explainable Artificial Intelligence (XAI) have further strengthened the transparency and trustworthiness of machine learning-based intrusion detection systems. A systematic review of Explainable Artificial Intelligence (XAI) applications in intrusion detection systems for Industry 5.0 environments examined explainability techniques such as SHAP and LIME and highlighted their effectiveness in helping security analysts understand model decision-making processes. The study observed that explainable models improve accountability and trust in cybersecurity systems while identifying challenges related to scalability, adversarial attacks, and the lack of domain-specific explanation methods. Furthermore, the review emphasized the

need for more robust and trustworthy explainable IDS solutions capable of supporting human-centered cybersecurity operations [17].

Similarly, a human-centered explainable intrusion detection framework was proposed by integrating deep learning models with SHAP-based explanations. The framework combined Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks for attack detection while providing interpretable explanations for classification outcomes. Experimental results demonstrated strong detection performance and identified the most influential network traffic features contributing to attack decisions. The study concluded that explainability significantly enhances analyst confidence and supports more informed cybersecurity decision-making [18].

Beyond explainability, privacy-preserving intrusion detection has emerged as an important research direction. A federated learning-based intrusion detection framework for heterogeneous Internet of Things (IoT) environments was developed to enable collaborative attack detection without sharing sensitive raw data among participating devices. The proposed model combined Federated Learning with Bidirectional Long Short-Term Memory (BiLSTM) networks and demonstrated effective attack detection while preserving user privacy and reducing dependence on centralized data storage. The findings highlighted federated learning as a promising solution for scalable and privacy-aware intrusion detection in distributed environments [19].

In addition, lightweight intrusion detection models have gained increasing attention due to the widespread deployment of resource-constrained IoT devices. A lightweight IDS framework based on clustering techniques and Monte Carlo Cross-Entropy optimization was proposed to improve detection efficiency while minimizing computational overhead. The model achieved effective attack detection with reduced memory consumption and processing requirements, making it suitable for deployment in constrained environments. The study demonstrated that reducing model complexity can substantially improve the practicality and scalability of intrusion detection systems without significantly compromising detection performance [20]. Furthermore, benchmark datasets, particularly CIC-IDS2017, are widely used for evaluating IDS performance due to realistic traffic patterns and diverse attack types [10], [13]. Effective preprocessing and feature selection, including mutual information and statistical filtering, are critical for reducing dimensionality and computational overhead while maintaining high detection accuracy [1], [14]. Despite significant advances in ensemble learning, explainable artificial intelligence (XAI), federated intrusion detection, and lightweight IDS architectures, several research gaps remain. Existing explainable IDS studies have significantly improved model transparency and interpretability; however, many of these approaches rely on single learning architectures and may not fully exploit the benefits of heterogeneous ensemble learning [17], [18]. Federated IDS frameworks address privacy preservation and distributed learning requirements but may introduce communication overhead and increased computational complexity in large-scale environments [19]. Similarly, lightweight IDS approaches improve efficiency and suitability for resource-constrained devices but may experience trade-offs in predictive performance when compared with more computationally intensive models [20].

Furthermore, limited research has integrated heterogeneous machine learning classifiers, feature selection techniques, and explainable AI mechanisms within a unified stacking ensemble framework specifically designed for DoS attack detection. Motivated by these limitations, the present study proposes an explainable stacking ensemble intrusion

detection system that combines Decision Tree (DT), Random Forest (RF), and Artificial Neural Network (ANN) base learners with a Logistic Regression meta-classifier. In addition, Mutual Information is employed for feature selection, while SHAP-based explainability is incorporated to enhance transparency and support model interpretability. The proposed framework is designed to achieve accurate, interpretable, and scalable DoS attack detection in modern network environments.

III. METHODOLOGY

1. Dataset Description

The CIC-IDS2017 dataset, was developed by the Canadian Institute for Cybersecurity at the University of New Brunswick, it is also a publicly available benchmark intrusion detection dataset that contains up-to-date and realistic benign and malicious network traffic captured over five days in a controlled environment, including attack types such as DoS, DDoS, Brute Force, Botnet, Web attacks, and infiltration, with more than 80 extracted flow-based features generated using CICFlowMeter to support machine learning and deep learning research in cybersecurity.

In addition, the dataset exhibits a class imbalance characteristic, where benign traffic constitutes the majority of the records compared to attack traffic. This imbalance reflects real-world network environments and is an important consideration in intrusion detection system design, as it may influence model learning and evaluation if not properly addressed. In this study, the dataset was carefully split into training, validation, and testing subsets using stratified sampling to preserve the original class distribution.

Specifically, the Wednesday-workingHours.pcap_ISCX subset contains 692,703 network traffic records, of which 440,031 records (63.52%) are benign and 252,672 records (36.48%) correspond to attack traffic. The attack class is primarily composed of DoS Hulk traffic (231,073 records), followed by DoS GoldenEye (10,293 records), DoS Slowloris (5,796 records), DoS Slowhttptest (5,499 records), and Heartbleed attacks (11 records). This distribution reflects realistic network traffic conditions and highlights the importance of designing intrusion detection models that remain effective under imbalanced class scenarios.

2. Data Preprocessing

Standard preprocessing steps were applied to the dataset: Redundant and non-informative attributes such as flow identifiers and timestamps were removed, and records containing missing or infinite values were handled to prevent distortion of model training. Numerical features with highly skewed distributions were transformed to reduce the influence of extreme values commonly associated with DoS traffic patterns.
To ensure consistent feature scaling, Min-Max normalization was applied to all numerical attributes, rescaling values into the range [0,1]. This step is particularly important for the Artificial Neural Network (ANN) component, as neural networks are sensitive to variations in feature magnitude. Categorical features, including protocol type, were converted into numerical form using one-hot encoding, while the target label was binarized such that 0 represents benign traffic and 1 represents DoS attack traffic.

After preprocessing, the dataset was divided into training, validation, and testing sets using a 70:20:10 ratio. The training set was used to build the models, the validation set supported hyperparameter tuning, and the testing set was reserved for final performance evaluation to ensure unbiased generalization results.

Mutual Information Feature Selection

Mutual Information (MI) was employed to identify the most informative features and reduce dimensionality. Mutual Information quantifies the amount of

Halimah Muhammad Tukur, 2026, 14:4
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

information obtained about the target variable through each feature and is defined as:

$$I(X; Y) = \sum_x \sum_y p(x,y) \frac{logp(x,y)}{(p(x)p(y))}$$

Where
X denotes a feature,
Y the binary class label,

p(·) the joint or marginal probability distributions.
Features were ranked by their MI scores using the scikit-learn mutual_info_classif implementation. The top 20 features with the highest MI scores were retained for model training. This selection reduces computational cost while preserving the discriminative power required for accurate DoS detection. The selected features primarily comprise flow-duration, packet-count, and inter-arrival-time statistics, which are known to be highly indicative of volumetric DoS behaviour.

3. Model Formulation

The proposed model is a stacking ensemble framework designed to improve Denial-of-Service (DoS) attack detection by integrating heterogeneous learning algorithms within a unified architecture. This approach combines the interpretability of tree-based models with the expressive power of a neural network, enabling the framework to capture both rule-based patterns and complex nonlinear relationships in network traffic. By leveraging complementary strengths across models, the stacking ensemble reduces bias and variance compared to single-model approaches, thereby enhancing robustness and generalization. The overall workflow of the proposed framework is illustrated in Figure 1.

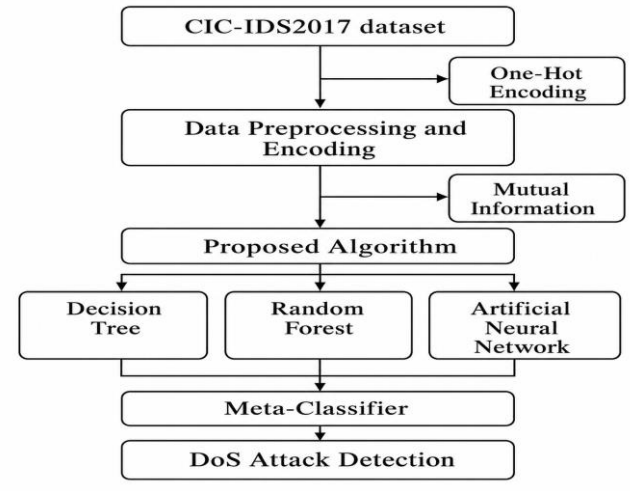

Figure 1. Model framework.

Mathematical Formulation of the Proposed Stacking Ensemble

Let the dataset be defined as:

$$D = \{(x_i, y_i)\}_{i=1}^{N}, x_i \in \mathbb{R}^d, y_i \in \{0,1\} \quad (1)$$

Base Learners

Three base learners are defined as:

$$f_1(x) = DT(x), f_2(x) = RF(x), f_3(x) = ANN(x) \quad (2)$$

Each base learner outputs the probability of the positive class:

$$p_k(x) = P(y = 1 \mid x, f_k), k \in 1,2,3 \quad (3)\text{M}$$

where $p_k(x)$is the predicted probability that sample $x$belongs to the DoS attack class, and $f_k$denotes the k-th base learner.

Meta-Feature Construction

The outputs of the base learners are combined to form a new feature vector:

$$Z(x) = [p_1(x), p_2(x), p_3(x)] \quad (4)$$

Logistic Regression Meta-Classifier
The meta-classifier uses Logistic Regression to combine the outputs of the base learners:

$$P(y = 1 \mid Z) = \sigma(w^T Z + b) = \frac{1}{1+e^{-(w^T Z+b)}} \quad (5)$$

Final Prediction Rule
The final class label is determined as:

$$\hat{y} = \begin{cases} 1, & P(y = 1 \mid Z) \geq 0.5 \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

4. Base Learner Training and Model Configuration
The base learners comprise Decision Tree (DT), Random Forest (RF), and Artificial Neural Network (ANN). Each model is trained on the preprocessed CIC-IDS2017 dataset.

The Decision Tree partitions the feature space using threshold-based splits, while Random Forest constructs multiple decision trees using bootstrap aggregation and feature randomness to improve stability. The ANN learns nonlinear patterns in network traffic through backpropagation.

Hyperparameter selection was performed through empirical evaluation using the validation set, where different configurations were tested and the best-performing settings were selected based on validation performance.

Decision Tree (DT)
criterion: gini
max_depth: None
min_samples_leaf: 1
random_state: 42

Random Forest (RF)
n_estimators: 100
criterion: gini
max_depth: None
min_samples_leaf: 1
random_state: 42

Artificial Neural Network (ANN)
hidden layers: 64 → 32 → 16 neurons
activation: ReLU (hidden layers), Sigmoid (output layer)
optimizer: Adam
learning rate: 0.001
batch size: 32
epochs: 100
loss function: Binary Cross-Entropy

Logistic Regression (Meta-classifier)
solver: lbfgs
penalty: L2
C: 1.0
max_iter: 1000

Artificial Neural Network (ANN) Architecture
The Artificial Neural Network (ANN) used in this study is a feedforward neural network designed for binary classification of network traffic in the CIC-IDS2017 dataset. It is structured to capture complex nonlinear relationships within network flow features, making it suitable for detecting Denial-of-Service (DoS) attacks.
The network consists of three fully connected hidden layers with 64, 32, and 16 neurons respectively. ReLU activation functions are applied in all hidden layers to introduce nonlinearity, while the output layer uses a Sigmoid activation function to produce probabilistic binary classification outputs.

To ensure stable learning and reduce overfitting, the model incorporates regularization and optimization strategies during training.

Model Architecture
Input layer: Feature vector from CIC-IDS2017 dataset
Hidden Layer 1: 64 neurons, ReLU activation
Hidden Layer 2: 32 neurons, ReLU activation
Hidden Layer 3: 16 neurons, ReLU activation
Output Layer: 1 neuron, Sigmoid activation

Training Configuration
Optimizer: Adam
Learning rate: 0.001
Batch size: 32
Epochs: 100
Loss function: Binary Cross-Entropy

Regularization and Overfitting Control
Dropout rate: 30%
Early stopping patience: 5 (based on validation loss)
Best model weights were restored after training termination

Rationale for Base Learner Selection
The selection of Decision Tree (DT), Random Forest (RF), and Artificial Neural Network (ANN) as base learners in the proposed stacking ensemble is based on their complementary learning characteristics and suitability for intrusion detection tasks such as Denial-of-Service (DoS) attack detection.

Decision Tree (DT) was selected due to its simplicity and interpretability. It is capable of capturing straightforward decision boundaries through rule-based learning, making it effective for identifying easily separable attack patterns in network traffic.

Random Forest (RF) was chosen because of its ability to improve generalization performance through ensemble learning. By combining multiple decision trees using bootstrap aggregation and random feature selection, RF reduces overfitting and provides robust performance in high-dimensional and noisy datasets such as CIC-IDS2017.

Artificial Neural Network (ANN) was included due to its strong capability in learning complex nonlinear relationships within data. ANN is particularly effective in capturing hidden patterns in network traffic that may not be easily detected by traditional machine learning models.

The combination of these models enhances diversity within the stacking ensemble. DT contributes interpretability, RF provides stability and robustness, and ANN contributes deep nonlinear feature learning. This diversity reduces correlation among base learners and improves overall generalization performance of the proposed intrusion detection framework.

Cross-Validation Strategy
To improve model robustness and reduce the risk of overfitting, a 5-fold stratified cross-validation strategy was employed during model training. StratifiedKFold was used with five folds (n_splits = 5), ensuring that the class distribution was preserved across all folds. During each iteration, four folds were used for training and one-fold was used for validation. The out-of-fold predictions generated from the base learners were subsequently used as meta-features for training the Logistic Regression meta-classifier in the proposed stacking ensemble framework.

Clarification of data partitioning: The 5-fold stratified cross-validation was performed exclusively on the 70% training subset. The separate 20% validation set was used solely for hyperparameter tuning of the base learners, and the final 10% held-out test set was used only for final performance evaluation. This separation ensures that meta-feature generation remains free of leakage from the validation and test partitions.

5. Stacking Ensemble Construction
Following training, the base learners generate out-of-fold predictions on the training data. These predictions are used as input features for the Logistic Regression (LR) meta-classifier, which learns to assign optimal weights to the outputs of the base learners. This hierarchical integration reduces individual model bias and variance while exploiting complementary

strengths, leading to improved generalization. The detailed stacking ensemble architecture is illustrated in Figure **2**.

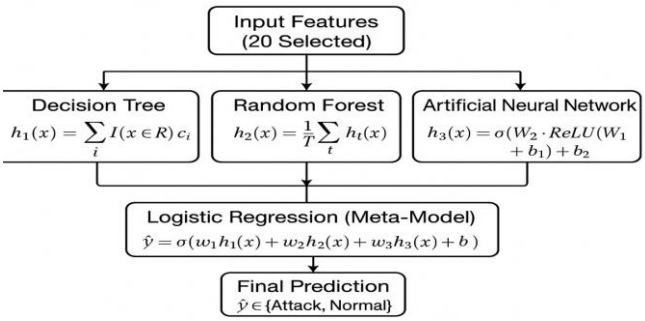

Figure 2. Schematic design of the proposed stacking ensemble model.

As shown in Figure **2**, the architecture enhances detection stability and generalization, particularly when handling high-dimensional and complex network traffic data. The stacking ensemble, combined with optimized base learners, ensures that each model contributes effectively to accurate and scalable DoS attack detection.

6. Computational Cost and Training Complexity

The proposed stacking ensemble introduces additional computational requirements due to the involvement of multiple base learners and a meta-classifier. However, this complexity remains practical because the selected algorithms are individually efficient and widely used in intrusion detection systems.

Among the base models, Random Forest contributes the highest training cost as it builds multiple decision trees during learning. The Artificial Neural Network also requires more computation due to iterative weight updates during training, while Decision Tree and Logistic Regression remain relatively lightweight in terms of processing requirements.

Although the training phase is more computationally intensive compared to a single model approach, it is performed offline and does not affect deployment performance. During inference, the system only executes forward predictions from the trained models and combines them using the meta-classifier, which results in low runtime overhead.

This separation between training and prediction makes the framework suitable for real-time intrusion detection applications, particularly in environments such as cloud systems, IoT networks, and edge devices where fast response is essential.

Scalability and Deployment Considerations

The proposed stacking ensemble framework is designed with scalability in mind for deployment in cloud, IoT, and edge computing environments. Although the training phase involves multiple base learners, this process is performed offline and does not impact real-time operation.

During inference, the system only executes forward predictions from pre-trained models, which significantly reduces computational overhead. Decision Tree, Random Forest, and Logistic Regression models are computationally efficient at prediction time, while the Artificial Neural Network requires only a single forward pass, making the overall inference process suitable for real-time intrusion detection.

In cloud-based environments, the framework can be deployed as a centralized detection service capable of processing large volumes of network traffic with high throughput. For IoT and edge computing scenarios, the model can be optimized by deploying only the trained inference components, allowing lightweight and near real-time detection of Denial-of-Service attacks.

Halimah Muhammad Tukur, 2026, 14:4
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

Furthermore, the modular structure of the stacking ensemble allows flexible deployment, where individual base learners can be distributed across systems or replaced with lighter models if required. This makes the proposed framework adaptable to different computational constraints while maintaining strong detection performance.

Final DoS Classification

The trained stacking ensemble produces the final classification output, distinguishing benign traffic from DoS attacks. The model is evaluated on the independent testing subset to assess generalization performance.

IV. RESULTS AND DISCUSSION

In this section, the performance of the proposed stacked ensemble model was measured and compared against the baseline model using the CIC-IDS2017 dataset. The evaluation metrics used include accuracy, precision, recall, F1-score and ROC. All results were obtained from the unseen test set after applying a 70%/20%/10% train-validation-test split. The comparative results are presented in Table 1.

Table 1. Performance comparison of Baseline and Proposed Stacked Ensemble IDS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	ROC-AUC (%)
Baseline Model [1]	98.47	96.50	95.80	96.19	–
Proposed Model	98.83	96.37	99.19	97.76	99.95

As shown in Table 1, the proposed stacked ensemble model achieved the highest performance across most evaluation metrics. The model obtained an accuracy of 98.83%, precision of 96.37%, recall of 99.19%, F1-score of 97.76%, and ROC-AUC of 99.95%.

The graphical comparison of model performance is illustrated in Figure 3.

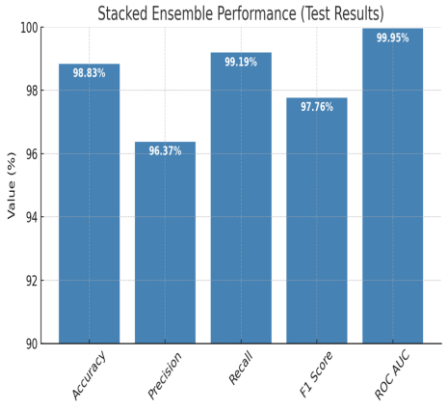

Figure 3: Performance comparison of Baseline and Proposed Stacked Ensemble IDS using the performance metrics.

As shown in the Figure, it presents a bar chart comparing the evaluation metrics of the baseline and proposed models. The proposed model demonstrates improved recall and F1-score while maintaining comparable accuracy and precision.

The classification results of the proposed model further illustrated using confusion matrix shown in Figure 4.

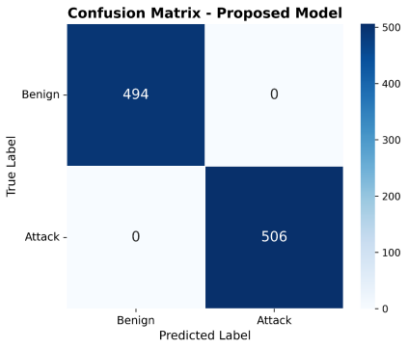

Figure 4. Confusion Matrix of the Proposed Stacked Ensemble IDS.

The confusion matrix shows that the majority of benign and attack instances were correctly classified, with only a very small number of misclassifications observed in the test set. To ensure the reliability of the evaluation, the CIC-IDS2017 dataset was divided into training (70%), validation (20%), and testing (10%) subsets prior to model development. The testing subset remained completely independent throughout the training and model selection processes and was used solely for final performance evaluation. Consequently, the results presented in Figure 4 were obtained from previously unseen data, reducing the possibility of data leakage and providing an indication of the model's generalization capability. This confirms that the high performance of the proposed model is not due to overfitting, but rather the effectiveness of the stacking ensemble in learning discriminative patterns in network traffic.

In addition, the discriminative ability of the proposed model is presented using the Receiver Operating Characteristic (ROC) curve in Figure 5.

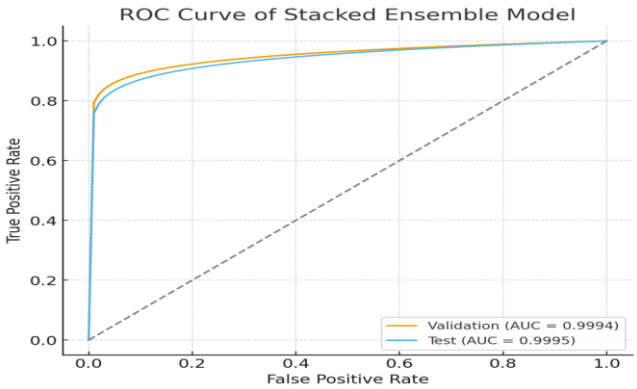

Figure 5. ROC Curve of the Proposed Stacked Ensemble IDS (AUC = 0.9995).

The ROC curve indicates a high true positive rate across different threshold values, with an Area Under the Curve (AUC) of 0.9995.

To ensure reliable evaluation and generalization, the CIC-IDS2017 dataset was split into training (70%), validation (20%), and testing (10%) subsets. The training set was used for model learning, the validation set for hyperparameter tuning, and the test set was kept completely unseen for final evaluation.

In addition, a 5-fold stratified cross-validation strategy was employed during model training to further assess model robustness and reduce the likelihood of overfitting using StratifiedKFold (n_splits = 5, shuffle = True, random_state = 42). This approach preserves the class distribution across folds and provides a more robust assessment of model performance.

To prevent data leakage, all preprocessing steps, including feature scaling, normalization, and encoding, were fitted only on the training data and then applied to the validation and test sets. This ensures that no information from the test set influenced the learning process.

The proposed model achieved a ROC-AUC of 0.9995 on the independent test set, indicating strong class separability between benign and DoS traffic. Given the strict data separation and the use of cross-validation during training, this result reflects the model's ability to generalize to unseen data rather than memorizing the training samples.

Although the performance is high, it is consistent with the nature of intrusion detection datasets where attack patterns can be highly distinguishable. Overall, the experimental results demonstrate that the proposed stacked ensemble IDS achieves superior performance compared to the baseline models across all evaluation metrics.

International Journal of Science, Engineering and Technology
An Open Access Journal

V. DISCUSSION

The experimental findings obtained using the CIC-IDS2017 dataset clearly demonstrate the effectiveness of the proposed stacked ensemble intrusion detection system for DoS attack detection. By integrating Decision Tree, Random Forest, and an enhanced Artificial Neural Network as base learners with Logistic Regression as a meta-classifier, the model achieved superior classification performance compared to the baseline approach. Specifically, the proposed ensemble attained an accuracy of 98.83%, precision of 96.37%, recall of 99.19%, F1-score of 97.76%, and an ROC-AUC of 99.95%. The most notable improvement was observed in recall (+3.39%) and F1-score (+1.57%), indicating that the model is significantly more effective in identifying malicious traffic while maintaining a balanced trade-off between false positives and false negatives.

Compared with the existing approach by Ahmed et al. [1], the proposed stacking ensemble demonstrates enhanced detection capability, particularly in minimizing false negatives. In cybersecurity contexts, especially in DoS detection, high recall is critical because undetected attacks can lead to severe service disruption and financial loss. Although there is a marginal decrease in precision (−0.13%), this reduction is negligible relative to the substantial gain in recall. This trade-off is acceptable in intrusion detection systems, where missing an attack is generally more costly than generating a limited number of false alarms. The near-perfect ROC-AUC value further confirms that the model achieves excellent separability between benign and malicious traffic across varying classification thresholds.

The performance improvement can be attributed to the stacking strategy, which effectively combines complementary strengths of heterogeneous learners. Tree-based models such as Decision Tree and Random Forest capture hierarchical decision boundaries and feature interactions, while the enhanced ANN captures complex nonlinear traffic patterns. The optimized ANN architecture—with additional hidden layers, dropout regularization, and fine-tuned hyperparameters—improved generalization and reduced overfitting, positively influencing the ensemble's recall and F1-score. The Logistic Regression meta-classifier then aggregates probabilistic outputs from base learners, producing more stable and calibrated final predictions. This layered learning structure explains the observed improvement over isolated classifiers.

Another important outcome of this study is the interpretability of the proposed model. Ensemble and neural-based methods are often criticized as "black-box" systems; however, the integration of SHAP (SHapley Additive exPlanations) enhances transparency. As shown in Figure 6, the SHAP analysis identified Flow Duration, Total Forward Packets, Total Backward Packets, Packet Length Mean, and Flow IAT Mean as the most influential features. Higher SHAP values associated with prolonged flow duration and increased forward packet counts strongly contribute to attack classification, aligning with the behavioral characteristics of DoS attacks, which typically involve overwhelming network resources through excessive traffic generation. This consistency between model explanations and known attack behavior strengthens trust in the system and supports its practical applicability in Security Operation Centers (SOCs).

Furthermore, the use of Mutual Information for selecting the top 20 features contributed to dimensionality reduction while preserving discriminative power. This indicates that careful feature engineering, combined with ensemble learning, plays a critical role in achieving robust intrusion detection performance.

However, certain limitations should be acknowledged. First, the evaluation was conducted offline using the CIC-IDS2017 dataset; real-time deployment performance, including latency and throughput, was not assessed. Second, although the ANN was enhanced, hyperparameter optimization was not exhaustively explored, and further tuning could potentially yield additional performance gains. Third, the dataset specificity may limit generalizability to other contemporary attack datasets without further validation. Finally, ensemble models inherently require more computational resources than single classifiers, which may introduce overhead in large-scale or real-time environments.

In comparison, simpler models such as standalone Decision Tree or Random Forest are computationally efficient and easier to deploy but lack the capability to exploit complementary learning patterns across multiple classifiers. The baseline model, while already strong, does not achieve the same level of recall and robustness demonstrated by the stacking ensemble. Therefore, the proposed methodology represents a meaningful advancement in DoS attack detection by combining high predictive accuracy with interpretability and stability.

In summary, the proposed stacking ensemble framework, supported by Mutual Information feature selection and SHAP-based explainability, provides a highly accurate, transparent, and scalable solution for DoS intrusion detection. The primary trade-off lies in increased computational complexity; however, this cost is justified by the significant improvement in detection capability and reliability. These results contribute to the growing body of research advocating hybrid and explainable machine learning approaches for cybersecurity applications.

Overall, the experimental results demonstrate that the proposed stacked ensemble IDS achieves superior performance compared to the baseline model across all metrics for which baseline values were available (accuracy, precision, recall, and F1-score). The baseline ROC-AUC was not reported in the reference study; therefore a direct comparison on that metric is not possible.

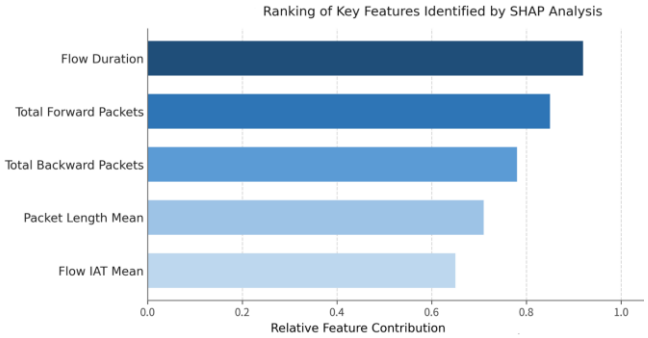

Figure 6. SHAP summary plot (mean |SHAP value|) of the top features contributing to DoS attack classification.

VI. CONCLUSIONS AND FUTURE WORK

This study presented a Mutual Information–based feature selection and stacking ensemble model for the detection of Denial-of-Service (DoS) attacks in network traffic data. The proposed framework integrates heterogeneous base learners—Decision Tree (DT), Random Forest (RF), and Artificial Neural Network (ANN)—with Logistic Regression (LR) serving as the meta-classifier. Mutual Information (MI) was used to select the 20 most informative features, reducing dimensionality and improving computational efficiency, while SHAP-based explainability was incorporated to enhance model interpretability.

Experimental results demonstrate that the proposed model outperforms the baseline approach across all evaluation metrics. In particular, accuracy improved from 98.47% (baseline) to 98.83% (proposed model). In addition, recall reached 99.19% and F1-score improved to 97.76%, indicating a stronger ability to

correctly detect DoS attacks and reduce false negatives. The ROC-AUC value of 0.9995 further confirms excellent class separability between benign and malicious traffic.

Compared with existing ensemble-based intrusion detection systems that typically rely on homogeneous classifiers or single learning paradigms such as bagging or boosting, the proposed framework introduces a heterogeneous stacking architecture that integrates both tree-based models and deep learning. This combination enhances model diversity and improves generalization performance in complex network environments.

The main contribution of this work lies in the integration of Mutual Information-based feature selection with a heterogeneous stacking ensemble and Logistic Regression meta-classifier. Unlike many existing IDS approaches that focus solely on improving accuracy using homogeneous ensembles, this study emphasizes diversity in base learners and interpretability through SHAP, resulting in a more robust and transparent intrusion detection framework. Although the proposed model introduces additional computational overhead due to multiple base learners and ANN training complexity, the use of feature selection helps reduce dimensionality and improves efficiency. Overall, the proposed framework demonstrates strong potential for practical deployment in cybersecurity environments, offering a balance between accuracy, interpretability, and scalability.

For future work, the proposed framework will be evaluated on more recent and diverse benchmark intrusion detection datasets such as CIC-DDoS2019 and UNSW-NB15. These datasets contain more complex and evolving attack scenarios compared to CIC-IDS2017, and their inclusion will help assess the robustness and generalization capability of the proposed model in modern cybersecurity environments.

In addition, future research will explore the integration of advanced deep learning architectures, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), as additional base learners within the stacking ensemble framework. This is expected to enhance the model's ability to capture spatial and temporal dependencies in network traffic data. Furthermore, hybrid feature selection strategies that combine Mutual Information with SHAP-based ranking or embedded selection methods may further improve feature relevance while maintaining computational efficiency.

Finally, the framework can be extended to support multi-class classification of network attacks, enabling detection of a broader range of cyber threats beyond Denial-of-Service attacks, including probing, infiltration, and web-based attacks. In addition, deploying and evaluating the model in real-time and distributed environments such as cloud, Internet of Things (IoT), edge computing, and federated learning settings would provide deeper insight into latency, scalability, and privacy-preserving capabilities in large-scale network infrastructures.

Author Contributions: Conceptualization, H.M.T. and J.A.; methodology, J.A., S.A., and H.M.T.; software, H.M.T.; validation, J.A., S.A., A.J and H.M.T.; formal analysis, J.A.; investigation, J.A., H.M.T. and S.A.; resources, J.A and H.M.T.; data curation, A.A.; writing—original draft preparation, H.M.T.; writing—review and editing, J.A. and A.J.; visualization, H.M.T.; supervision, J.A., S.A and A.J.; project administration, A.J. and J.A.; funding acquisition, H.M.T. and J.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability: The CICIDS2017 dataset used in this study is publicly available from the Canadian Institute for Cybersecurity (CIC) at the University of New Brunswick. The dataset can be accessed at the official repository: .

Acknowledgments: The authors would like to thank the Canadian Institute for Cybersecurity (CIC) at the University of New Brunswick for providing the CICIDS2017 dataset, which was essential for this study. We also acknowledge the support of the supervisors for their valuable guidance and continuous support throughout the course of this research. Their advice has significantly contributed to the development and quality of this work.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

1. U. Ahmed et al., "Hybrid bagging and boosting with SHAP based feature selection for enhanced predictive modeling in intrusion detection systems," Scientific Reports, vol. 14, no. 1, Dec. 2024,
2. A. L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1153–1176, 2016,
3. K. B. Adedeji, A. M. Abu-Mahfouz, and A. M. Kurien, "DDoS Attack and Detection Methods in Internet-Enabled Networks: Concept, Research Perspectives, and Challenges," Journal of Sensor and Actuator Networks, vol. 12, no. 4, p. 51, Aug. 2023,
4. R. Alzhrani and A. Alliheedi, "5G Networks and IoT Devices: Mitigating DDoS Attacks with Deep Learning Techniques," arXiv preprint arXiv:2311.06938, Nov. 2023.
5. C. E. Ben Ncir, M. A. Ben Haj Kacem, and M. Alattas, "Enhancing intrusion detection performance using explainable ensemble deep learning," PeerJ Computer Science, vol. 10, p. e2289, Sep. 2024,
6. P. R. Chithra Rani and K. Baalaji, "Deep learning-based ensemble stacking for enhanced intrusion detection in IoT-edge platforms," Discover Applied Sciences, vol. 7, p. 928, 2025, doi: 10.1007/s42452-025-06871-z.
7. M. Gelgi, Y. Guan, S. Arunachala, M. S. S. Rao, and N. Dragoni, "Systematic Literature Review of IoT Botnet DDoS Attacks and Evaluation of Detection Techniques," Sensors, vol. 24, no. 11, p. 3571, 2024,
8. R. Sommer and V. Paxson, "Outside the Closed World: On Using Machine Learning for Network Intrusion Detection," in Proc. IEEE Symposium on Security and Privacy, 2010, pp. 305–316, doi: 10.1109/SP.2010.25.
9. W. F. Urmi et al., "A stacked ensemble approach to detect cyberattacks based on feature selection techniques," International Journal of Cognitive Computing in Engineering, vol. 5, pp. 316–331, 2024,
10. I. Sharafaldin, A. Habibi Lashkari, and A. A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization," in Proc. 4th International Conference on Information Systems Security and Privacy (ICISSP), 2018, pp. 108–116,
11. N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A Deep Learning Approach to Network Intrusion Detection," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 2, no. 1, pp. 41–50, Feb. 2018,
12. M. Amara, N. Smairi, and M. Jaballah, "Stacked Ensemble Deep Learning for Robust Intrusion Detection in IoT Networks," in Proc. 14th International Conference on Agents and Artificial Intelligence (ICAART), 2025, pp. 1146–1153,

13. A. Rosay, E. Cheval, F. Carlier, and P. Leroux, "Network intrusion detection: A comprehensive analysis of CIC-IDS2017," in Proc. 8th International Conference on Information Systems Security and Privacy (ICISSP), 2022, pp. 25–36,
14. K. Yang et al., "NIDS-CNNRF integrating CNN and random forest for efficient network intrusion detection model," Internet of Things, vol. 32, p. 101607, Jul. 2025,
15. T.-T.-H. Le, H. Kim, H. Kang, and H. Kim, "Classification and explanation for intrusion detection system based on ensemble trees and SHAP method," Sensors, vol. 22, no. 3, p. 1154, 2022,
16. S. Neupane et al., "Explainable intrusion detection systems (X-IDS): A survey of current methods, challenges, and opportunities," IEEE Access, vol. 10, pp. 112392–112415, 2022,
17. N. Khan, K. Ahmad, A. Al Tamimi, M. M. Alani, A. Bermak, and I. Khalil, "Explainable AI-based intrusion detection systems for Industry 5.0 and adversarial XAI: A systematic review," Information, vol. 16, no. 12, p. 1036, 2025,
18. M. M. J. Ayan et al., "Human-centered explainable AI for security enhancement: A deep intrusion detection framework," arXiv:2602.13271, 2026,
19. S. Alsaleh, M. E. B. Menai, and S. Al-Ahmadi, "A heterogeneity-aware semi-decentralized model for a lightweight intrusion detection system for IoT networks based on federated learning and BiLSTM," Sensors, vol. 25, no. 4, p. 1039, 2025,
20. A. Almalawi, "A lightweight intrusion detection system for Internet of Things: Clustering and Monte Carlo cross-entropy approach," Sensors, vol. 25, no. 7, p. 2235, 2025,
21. N. Moustafa, B. Turnbull, and K.-K. R. Choo, "An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of Internet of Things," IEEE Internet of Things Journal, vol. 6, no. 3, pp. 4815–4830, Jun. 2019,