Journal of Artificial Intelligence and Metaheuristic

♦ ISSN: 2833-5597 Vol: 09, No. 02, 2025 ♦ PP. 01–13

Article DOI: https://doi.org/10.54216/JAIM.090201

FULL LENGTH ARTICLE

Hybrid Ensemble Learning for Flow-Level IoT Traffic Classification Using ACI Dataset: Towards Scalable and Real-Time Threat Detection

El-Sayed M. El-Kenawy[1,2] **Sini Raj Pulari**[3,*] **Shriram K Vasudevan**[4]

[1] School of ICT, Faculty of Engineering, Design and Information and Communications Technology (EDICT), Bahrain Polytechnic, PO Box 33349, Isa Town, Bahrain

[2] Applied Science Research Center, Applied Science Private University, Amman, Jordan

[3] Department of CSE, Vignan's Foundation for Science, Technology and Research, Guntur, Andhra Pradesh, India

[4] Intel India Pvt. Ltd., Bengaluru, India

Emails: skenawy@ieee.org ▪ sinikishan@gmail.com ▪ shriram.kris.vasudevan@intel.com

Received: December 18, 2024 Revised: February 10, 2025 Accepted: May 02, 2025 ⋆ Corresponding author

ABSTRACT

Internet of Things devices, which spread across consumer industrial and critical infrastructure domains, have boosted the quantity of diverse network traffic and its high frequency. The increasing scale of IoT networks causes problems securing the diverse data flow within these networks, threatening system performance and management capabilities. Analyzing network traffic with traditional methods based on signature identification and rule detection becomes ineffective for new traffic activity patterns and system behavior. Due to extensive growth in IoT networks, developing intelligent data-based classification systems that can process IoT traffic quickly and at large operational scales becomes essential. A detailed model of flow-level data-based machine learning operations for IoT traffic classification utilizes features extracted from the Army Cyber Institute (ACI) IoT dataset. The dataset encompasses statistical, temporal, and protocol-specific attributes for benign and malicious network flows. Our methodology first conducts a strict data preprocessing stage, which involves numerous operations such as cleaning the data, normalizing it and encoding the labels, and performing a feature correlation analysis before preparing the learning algorithms with a suitable quality and balanced dataset. Various classification models underwent training, including Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Naive Bayes and SGD Classifiers, and statistical learners. Our proposed hybrid ensemble method combines weighted voting between a deep learning neural network, a Random Forest model, and an XGBoost classifier to overcome the limitations of single classifiers. This ensemble model aimed to make the system more resilient while lowering bias and enhancing its ability to understand various IoT traffic patterns. A complete set of evaluation metrics assessed the models, using accuracy, precision, recall, F1-score, Hamming loss, Matthews correlation coefficient (MCC) and Cohen's Kappa plus balanced accuracy and log loss for assessment. The chosen metrics allowed researchers to monitor model performance from global and detailed perspectives when dealing with imbalanced classes and similar patterns between legitimate and malicious network traffic. The ensemble methodology produces superior results than individual classifiers demonstrated through experimental results under all performance metrics evaluation. The complex nature of network environments demonstrates that model fusion achieves excellent results when tracking non-easy-to-classify traffic patterns. The ensemble approach proves excellent generalization properties and optimized performance for real-time IoT implementations because of its ability to adapt continuously while maintaining high accuracy levels. This proposed framework adds to intelligent IoT traffic analysis research while demonstrating how deep learning and traditional machine learning methods enhance ensemble systems. The system develops an expandable and clear quantitative solution that can be implemented for advanced network security systems and traffic monitoring applications across smart cities industrial settings, and critical infrastructure frameworks.

Keywords: IoT Traffic Classification ▪ Ensemble Learning ▪ Deep Learning ▪ Flow-Based Analysis

© 2025 The Author(s). Published by ASPG. This article is licensed under CC BY 4.0.

1. INTRODUCTION

The Internet of Things (IoT) has permanently reshaped modern digital ecosystems by surge in its adoption and creation. Global devices in the billions now communicate through IoT technologies, which manufacture massive, interminable flows of distinct high-frequency information from household machines to industrial equipment, environmental sensors, and urban infrastructure. The vast amount of data transfers has become essential for running real-time analytics, which serves operations in critical fields such as climate-tracking and precision agriculture, intelligent transportation, and energy-efficient urban development [1, 2]. At the same time, the expanding IoT ecosystem produces great chances alongside various demanding obstacles. The growth of IoT networks increases complex traffic patterns beyond measure because traffic management, prediction systems, and security solutions face escalating needs. High dimensionality and stochastic tendencies make forecasting and classifying IoT network traffic challenging. The communication patterns of IoT devices differ from standard computing systems by showing unpredictable behavior together with specific protocols and sensitive contexts. Industrial devices express their behaviors across ranges that depend on hardware setups, environmental factors, different application circumstances, and firmware version releases. Traffic from IoT networks makes detection harder because it carries small data payloads with bursty patterns combined with asymmetric network flows [3, 4].

The innovative behavior of IoT systems surpasses traditional detection methods because these methods cannot adapt to new characteristics. Network security measures incorporated machine learning (ML) and deep learning (DL) into their operations since traditional detection systems demonstrated their limitations [5, 6]. The core advantage of employing ML techniques in analyzing IoT networks is their ability to detect relationships between various data sets and intricate connection patterns. Standard network planning systems prove inadequate when managing contemporary IoT traffic since these methodologies cannot cope with high device quantities and multiple communication protocols that change their environment. ML-based models accept large quantities of data to process traffic patterns alongside traffic adaptation at the edge and cloud levels while making automated decisions [7, 8]. IoT systems implemented with ML technology generate various capabilities, including traffic detection, device recognition, and network intrusion prevention, and policy execution to enhance IoT networks' performance and resilience.

IoT traffic patterns' complete range and complexity exceed what any single model architecture can effectively replicate. Random Forests and Gradient Boosting Machines use their inherent interpretability to achieve generalization power on structured flow-level data according to classical ML algorithms. Through deep neural networks, we can extract abstract nonlinear patterns from data. At the same time, they need large training volumes and get affected by noisy or imbalanced data [9]. The fundamental trade-offs between attributes have driven researchers to develop ensemble learning, which unifies several models into unified frameworks, delivering superior accuracy and stability compared to their basic elements. The complex nature of IoT traffic classification makes ensemble learning a highly beneficial method for effective classification. Ensemble methods utilize multiple classifiers with a standard dataset to produce prediction outputs that improve robustness due to diversified model approaches. The approach decreases bias and variance by adding a redundancy layer that aids anomaly detection across various traffic patterns. The up-and-coming traffic types and network condition fluctuations in IoT spaces make ensemble models the best choice for maintaining reliable outcomes through their adaptable design. Weighted ensembling allows models to combine diversity with component strength evaluation by assigning importance ratings to individual components. Weighted ensemble approaches give models importance weights through confidence measurement and statistical and contextual information instead of providing all models with identical status. The ensemble optimizes its decision strategy through this adjustment approach to best fit different traffic patterns dynamically. Such models work best for multiclass classifications since they must detect slight differences between classes with high precision and sensitivity levels.

This research analyzes data from the Army Cyber Institute (ACI) IoT network traffic dataset, providing detailed real-time device interaction records available to the public. Flow-level attributes in the dataset include source and destination IPs, ports and protocol types, packet statistics and timing details, TCP flags, and entropy-based indicators [10]. The flow-level features give a complete understanding of network actions for implementing sophisticated ML pipelines to detect device roles while monitoring traffic anomalies in wired and wireless environments. The proposed study develops an intelligent system for IoT traffic classification by implementing hybrid ensemble framework design techniques. The framework uses multiple learning principles that work together inside a weighted ensemble system that analyzes data at different abstraction layers. The ensemble uses multiple analytical perspectives from its components to interpret network data better because different analytic lenses enhance context-aware analysis [11]. Achieving this work's primary goal involves developing a single consolidated framework that handles traffic classification, device fingerprinting, and anomaly detection needs in advanced distributed IoT infrastructure systems. The research aims to develop next-generation network monitoring systems that combine accuracy and scalability with adaptivity and interpretability for deployment in actual network environments.

2. LITERATURE REVIEW

Vehicle traffic has surged dramatically with escalating network device numbers and the popularization of Internet of Things devices, so researchers have become active in developing sophisticated techniques for network traffic analysis, anomaly detection, and cyber-threat defense strategies. Detecting contemporary cyber threats becomes more challenging because advanced security methods are insufficient to deal with modern threats. Researchers now employ Machine Learning techniques to improve network monitoring and prediction systems because these methods enhance adaptability, resilience, and intelligence performance.

Research initiated the establishment of classical and deep ML algorithms for LTE network edge traffic load forecast-

ing. Research outcomes revealed that Gradient Boosting produced optimal predictive results. Still, SVMs delivered faster training time than the tested models, including Linear Regression, Gradient Boosting, Random Forest, Bagging, Huber Regression, Bayesian Regression, and Support Vector Machines (SVM). Bayesian Regression provided an optimal performance-speed ratio, and Huber Regression demonstrated high resistance to data errors. The study focused on performance outcomes and deployment feasibility while making their code available to the public for research replication and evaluation [12]. A signature-based detection system for identifying typical home environment-based malicious behavior was introduced within the field of IoT security research. The research assessed how the Mirai botnet utilizes DNS and Telnet protocols by implementing passive sniffing analysis with cloud-based analytics. The detection system demonstrated 98.35% accuracy for DNS traffic together with 99.33% accuracy for Telnet traffic, resulting in a total detection effectiveness of 98.84%. The research stands out because it provides security monitoring solutions to everyday users through systems that simplify the application of security features even for non-technical users [13].

Detecting anomalies remains challenging when IoT network data are weakly labeled or unlabeled. Nguyen and Kashef proposed TS-IDS, a traffic-aware self-supervised approach for IoT intrusion detection that learns flow relationships among network entities, making use of traffic structure when labeled examples are limited [14]. Feature selection is also important for high-dimensional traffic data. Kushwaha et al. proposed a multi-method stacked feature-selection approach for IoT intrusion detection that combines several selection techniques to reduce the feature set, and evaluated it on CICIDS 2017 data [15].

Lazzarini et al. proposed DIS-IoT, a stacking ensemble that combines four deep learning models for binary and multiclass IoT intrusion detection. They evaluated the method on the ToN_IoT, CICIDS2017, and SWaT datasets, reporting low false-positive rates and strong multiclass performance [16]. A complementary 2023 contribution introduced CICIoT2023, a real-time dataset and benchmark with traffic from 105 IoT devices and 33 attacks across seven categories, supporting machine-learning evaluations of benign-versus-malicious, grouped, and multiclass traffic [17].

Research teams established a smart-home testbed containing different devices to analyze IoT device traffic patterns because they understood the need to detect and monitor individual IoT devices. The research used IoTTGen for novel traffic simulation followed by entropy-based behavioral analysis of network participants. The machine learning models trained using entropy features demonstrated 94% classification precision and proved their effectiveness under irregular IoT environment conditions [18]. Smart city traffic congestion creates two types of difficulties relating to city traffic management and environmental sustainability. The OWENN and Intel 80,286 microprocessor collaboration built a new traffic signal management system for intelligent control. The procedure comprises multiple phases, which include gathering information and extracting characteristics before utilizing them for classification, followed by real-time system control implementation. The OWENN algorithm achieved 98.23% accuracy together with a 96.69% F-score, which surpassed current traffic management solutions [19].

The classification of IoT data needs to be effective to secure and manage network traffic flows. Developers created IoTHunter, a Deep Packet Inspection-based system for automated traffic flow keyword extraction to detect IoT devices. Network flow labeling becomes precise through IoTHunter by linking keywords with MAC addresses. The system achieved high precision when identifying different types of IoT devices as part of testing on public IoT data, which enhances network behavior tracking and anomaly alerting capabilities [20]. Research analysts developed an edge-intelligent framework that uses machine learning algorithms to detect malicious IoT traffic while identifying IoT devices during rising IoT complexity. Furthermore, the method derives flow-based attributes at network boundaries to establish device identification while discerning traffic patterns and identifying abnormal behavior immediately. Random Forest yielded the highest performance metrics throughout the benchmark process, which led to 94.5% device-type identification accuracy, 93.5% traffic-type classification accuracy and 97% malicious traffic detection level [21].

Research work presented a machine learning framework that handled traffic classification within SDN-IoT networks to improve network performance while meeting strict QoS criteria [22]. SDN architecture gained application programming abilities, which enabled the authors to incorporate ML techniques into the control layer. Various classifiers underwent a comparative evaluation in this study by assessing Random Forests, Decision Trees and K-Nearest Neighbors (KNN). The work applied Sequential Feature Selection (SFS) along with Shapley Additive Explanations (SHAP) as feature selection strategies to accomplish model input optimization and lower input features. Combining the Random Forest algorithm with the SFS feature selection method yielded optimum results by reaching 0.833 accuracy with six chosen features from available inputs. Such findings demonstrate that compact, efficient models work well for IoT implementations prioritizing fast computations.

Researchers explored IoT device classification methodologies using traffic behaviors to achieve better QoS management alongside security threat identification and network visibility improvements [23]. This research highlighted the difficulty in selecting features and designing algorithms since ML classifiers behave according to the data nature, deployment environment, and extracted feature attributes. The authors performed an extensive comparative study, which utilized publicly available traffic traces derived from 20 IoT devices throughout 20 days. The research rated algorithms using metrics including classification accuracy while reporting time measurements and processing speed to offer suitable recommendations for each application. A research paper introduced a traffic classification system with ML capabilities to detect IoT traffic types between normal and harmful conditions [24]. Network security enhancement was the principal reason, but the system also focused on congestion control through traffic management. The research established traffic classification as fundamental because it enabled proper identification of malicious traffic from regular flows to implement preventive security measures. Flow-level features processed through ML

models enabled the system to route verified traffic correctly to service nodes while protecting against exposure to unsafe data streams.

These research investigations demonstrate that machine learning is essential in enhancing IoT traffic management systems. The growing requirements in IoT-driven networks find solutions through ML-based classification systems that provide adaptable, scalable, and efficient operations. These models' interpretability and computational feasibility become better due to the addition of SHAP and SFS feature selection methods. The continuous development of IoT infrastructure requires integrating real-time processing and edge computing with policy enforcement for ML-driven traffic classification to build secure and resilient infrastructure networks.

3. DATASET

3.1 Dataset Preparation

The research utilizes a comprehensive dataset of IoT network communication features that the Army Cyber Institute recorded in their IoT traffic dataset. The network traffic features include statistical packet measures from forward and backward paths, timing statistics about interpacket delays, packet dimensions, and control signal patterns. Machine learning model training requires an essential preliminary step where the interrelationship between numerical features must be analyzed, and the dataset needs to be adequately prepared. The correlation analysis examined the fundamental relationships between features, which allowed an evaluation of how features support or duplicate each other. Identifying correlated features holds crucial importance during model input preparation because multicollinearity affects both stability and interpretability of linear models. Understanding feature relationships that show weak or no correlations enables the investigator to defend including those features even though they might operate independently from other features within the learning environment.

A selected subset of numerical features displays their Pearson correlation matrix through Figure 1. All matrix cells measure variable correlations by numbers between -1 and 1. A correlation value of 1 indicates a perfect positive relationship between variables, while -1 represents a perfect negative relationship, and 0 signifies no linear relationship between variables. Each element along the primary matrix diagonal shows full matching correlations between the features and themselves. The heatmap uses diverging colors to indicate a stronger positive correlation through darker red tones, while darker blues represent a negative correlation. The heatmap reveals that most feature combinations demonstrate a weak to average interdependence scale. The network behavior measurement set includes Fwd IAT Mean and Bwd IAT Mean and Flow Duration which demonstrate minimal connectedness with the packet-based measurement set consisting of Packet Length Mean and ACK Flag Count. The model benefits from diverse features since this diversity helps prevent information overlap between features and increases expressive modeling capabilities. Approximately two weak correlations emerge between some of the features in the dataset. Total Fwd Packet shows a slight positive relationship with Packet Length Mean yet Flow Duration demonstrates almost no correlation to Fwd PSH Flags or ACK Flag Count. The temporal characteristics in the dataset follow their distinct pattern from protocol-level control indicators, strengthening the observation of multidimensional IoT traffic patterns. The analysis results justify using algorithms that process linear and nonlinear relationship types because the correlation method detects only linear dependencies. Fault detection demands ensemble models together with neural networks because these models can detect intricate patterns. This analytic analysis of data internal patterns enables a better understanding of the data structure while streamlining the feature selection process through guided engineering principles. The model would identify strongly related features for reduction techniques and weakly related ones for potential emphasis due to uniqueness. The predictive model achieves better quality and stability by acquiring new knowledge https://www.kaggle.com/code/monagaffer12345/iot-2023-dataset.

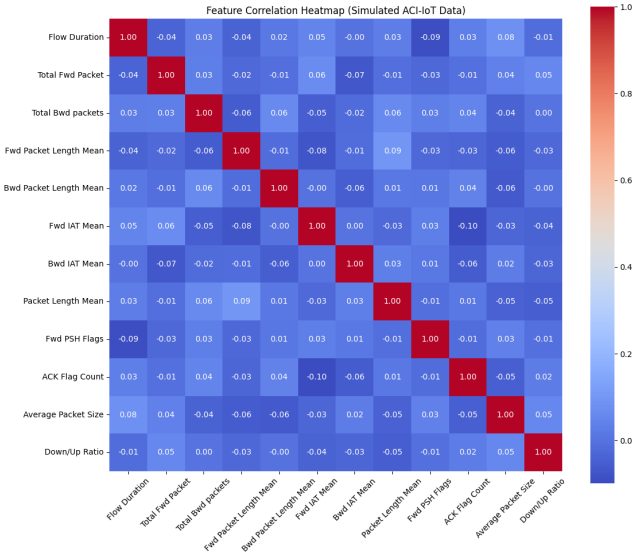


Figure 1. Feature correlation heatmap generated from the ACI-IoT traffic dataset. The visualization highlights linear relationships between key numerical features in the classification task. Most features show low pairwise correlation, indicating that the feature space is diverse and suitable for machine learning models.

3.1.1 Selected Features Overview

Research findings validated by domain expertise led experts to choose selected features as inputs for machine learning operations. The selection process focused on retaining features that delivered informative and non-redundant information spanning various aspects of IoT traffic behavior between flows, temporality and protocol governing elements. Table 1 summarizes the key features selected for modeling, along with a brief description and their role in classification. These features were chosen to maximize the representational power of the model while minimizing noise and redundancy. The choice of attack indicators was determined by features that provide clear interpretations and their correspondence to known attack patterns within IoT traffic. ACK Flag Count and Fwd PSH Flags display control actions in TCP connections, while Fwd IAT Mean records timing irregularities that occur during scanning or flooding attacks. As previously explained, the chosen features, standardization, and variable transformation create a fundamental platform for model development

across subsequent research sections.

3.2 Dataset Preprocessing

Data processing involved multiple structured steps before im-plementing machine learning models for earthquake magni-tude forecasting purposes. Clearing and transforming global seismic records from 2023 required detailed preparation to obtain valuable patterns [25]. The research employed this preprocessing methodology for data preparation.

Handling Missing and Incomplete Records: The first re-view process detected absent data while confirming complete seismic records availability. Data fields for magnitude, depth, and geographic coordinates required replacement because data imputation and removal operations depended on record frequency and level of missing data. Statistical validity and temporal data structures remained intact through the impu-tation methods that used mean or median substitution and occasional forward-fill techniques to address non-essential data fields.

Format and Structure Consistency: A standardized tabular format was created to harmonize diverse seismic data sources initially present in various formats. The standardized time stamps followed the format of ISO 8601, while the categori-cal attributes received normalization through a standardizing process. All numerical data received consistent data typifica-tion to enable compatibility with machine learning operations and prevent type coercion problems.

Feature Transformation and Unit Standardization: The model needed specific fields to undergo structural changes for operational purposes. The depth values were converted to kilometers as standard units, while magnitude values received two decimal point rounds to maintain uniform presentation. The date-based characteristics from data were extracted to generate temporal features during the engineering process for temporal pattern analysis. The data retention of latitude and longitude values in decimal degrees format allowed spatial pattern recognition without modifications.

Outlier Detection and Filtering: The analysis removed extreme outlier data points through IQR filtering together with domain expertise knowledge because they showed un-reasonable magnitude or depth measurements (e.g., negative quantities or values above geological limits). The process identified and eliminated extreme outliers to minimize dis-tortions during modeling and boost performance assessment stability.

Normalization and Scaling: A process of feature scaling became essential to enhance understanding and achieve speedier convergence rates with KNN and SVR, which depend on distance calculations. The consistent variables depth, magnitude and gap underwent Min-Max scaling or Z-score normalization based on particular algorithm demands. Processing the raw seismic records with these methods made creating a machine-learning-appropriate structured dataset possible. The pipeline improved model reliability for earthquake mag-nitude prediction by dealing with inconsistent data, removing outliers, and resolving scaling problems.

4. RESULTS

A thorough assessment of several machine learning models was conducted for IoT network traffic classification. The main objective is determining which models exhibit the best performance when dealing with the complications found in IoT traffic data sets. The study incorporates classical algorithms for speed and interpretability combined with an ensemble voting model that utilizes DL, RF, and XGB. The collection of heterogeneous models combines their strengths with weakness compensation mechanisms, resulting in a more robust and scalable intelligent traffic monitoring solution.

4.1 Evaluation Metrics

We use standard classification metrics for model assessment, including precision, recall, and F1-score. When applied to IoT security, the metrics perform crucial roles because incorrect pessimistic predictions (attacked systems not detected) and incorrect optimistic predictions (false alerts) have meaningful financial consequences. Precision shows the percentage of genuine positive predictions from all model-generated positive predictions. High precision helps minimize false alarms because it is vital for systems that should minimize automated shutdowns and unnecessary alert signals. Recall determines a model's capability to detect all important positive cases. Elevated recall performance in intrusion detection systems means even less noticeable and rare attacks will remain detected without failure. The combination of precision and recall forms the F1-score through its harmonic mean calculation. The tool proves beneficial when working with datasets that present an imbalance between classes because benign data typically surpasses attack traffic within IoT environments.

Let TP denote true positives, FP false positives, and FN false negatives. The formulas are:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{2}$$

$$\text{F1-score} = \frac{2\,\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{3}$$

The metric calculation method used micro-average aggregation to combine contributions from every class in the assessment process. Using this method, overall performance assessment includes every part of the data distribution without being unduly influenced by class frequency patterns, which is suitable for IoT traffic distributions.

4.2 Machine Learning Models

Performance evaluation focused on well-established classifiers for assessing outcomes using simple linear and probabilistic decision systems.

4.2.1 Naive Bayes

The Naive Bayes algorithm uses Bayes' Theorem under an unrealistic condition of independent features. The model finds the most probable class by evaluating posterior probabilities and selects the class with the greatest value according

Table 1. Summary of Selected Features for IoT Traffic Classification

Feature Name	Description	Category
Flow Duration	Total duration of the network flow in microseconds	Temporal Feature
Total Fwd Packet	Total number of packets sent in the forward direction	Traffic Volume
Total Bwd Packets	Total number of packets sent in the backward direction	Traffic Volume
Fwd Packet Length Mean	Mean size of forward packets	Packet Size Statistics
Bwd Packet Length Mean	Mean size of backward packets	Packet Size Statistics
Fwd IAT Mean	Mean inter-arrival time between forward packets	Temporal Feature
Bwd IAT Mean	Mean inter-arrival time between backward packets	Temporal Feature
Packet Length Mean	Mean packet size across the flow	Packet Size Statistics
Fwd PSH Flags	Number of PSH flags in forwarding packets	Protocol Flags
ACK Flag Count	Number of ACK flags observed in the flow	Protocol Flags
Average Packet Size	Average size across all packets in the flow	Aggregate Feature
Down/Up Ratio	Ratio of download bytes to upload bytes	Behavioral Feature

to the model:

$$\hat{z} = \arg\max_{V_m} P(V_m) \prod_{i=1}^{n} P(y_i \mid V_m). \tag{4}$$

The formula includes prior probability of class V_m as $P(V_m)$ along with the conditional probability of feature y_i under class condition V_m. The Naive Bayes model achieved good results while detecting basic attacks of high frequency but delivered insufficient accuracy on complex traffic that violated its independence assumption.

4.2.2 Stochastic Gradient Descent

Using stochastic optimization, the SGD Classifier trains linear model structures. The method implements iterative updates to reduce a chosen loss function that frequently uses logistic loss in binary classification issues:

$$\theta_{t+1} = \theta_t - \eta \nabla L(\theta_t). \tag{5}$$

Here, θ is the parameter vector, η is the learning rate, and L is the chosen loss function. Its memory efficiency and speed make it suitable for online learning on streaming IoT data, but its linear nature limits performance in highly non-linear class boundaries.

4.2.3 Linear Discriminant Analysis

The goal of LDA consists of transforming data into a space with fewer dimensions which optimizes the differentiation between categories. The model bases its assumption on linked class covariances combined with features that follow normal distributions. The prediction function used during processing consists of:

$$\delta_m(\mathbf{y}) = \mathbf{y}^T \Sigma^{-1} \mu_m - \frac{1}{2} \mu_m^T \Sigma^{-1} \mu_m + \log P(V_m). \tag{6}$$

The calculation includes mean vectors from class k through μ_k together with shared covariance matrix Σ and prior class probabilities expressed through $P(V_m)$. The problem areas for LDA include its inability to perform effectively in cases involving overlapping class distributions and class imbalance.

4.2.4 Quadratic Discriminant Analysis

QDA makes LDA better by allowing different covariance matrices for each class to enable modeling of non-linear boundaries. The function for QDA discriminant analysis takes the form:

$$\delta_m(\mathbf{y}) = -\frac{1}{2} \log |\Sigma_k| - \frac{1}{2} (\mathbf{y} - \mu_m)^T \Sigma_k^{-1} (\mathbf{y} - \mu_k) + \log P(V_m). \tag{7}$$

Introducing this technique enhances expressiveness yet it creates higher variability that affects smaller datasets. The results of this study demonstrated QDA delivered better performance than LDA except in attacks where it detected noisy samples incorrectly.

4.2.5 Proposed Ensemble Voting Model (DL + RF + XGB)

A combination voting algorithm was implemented to mitigate single classifier flaws through the strategic utilization of three different learning approaches:

- Deep Learning (DL) provides non-linear modeling capacity and excels at capturing hierarchical data representations. It is particularly effective for detecting complex attack patterns that may not be explicitly represented in the feature set.
- Random Forest (RF) brings ensemble stability and resistance to overfitting, offering better generalization by averaging multiple decision trees trained on bootstrapped data.
- Extreme Gradient Boosting (XGB) focuses on correcting errors through successive learning iterations, offering high accuracy and fine-grained decision boundaries, especially effective on structured data like flow-based traffic features.

The combination of predictions from the three models occurs through weighted majority voting. The prediction $p_i^{(j)}$ from the model which uses the weight w_j represents each instance i. The ensemble obtains its final prediction through:

$$\hat{y}_i = \arg\max_k \sum_{j=1}^{3} w_j \mathbb{I}\left(p_i^{(j)} = k\right). \tag{8}$$

The deep neural network architecture includes:

- Batch normalization for faster convergence and input stability;

- Dense layers with 64 and 16 neurons activated by ReLU;
- Dropout layers (rate = 0.5) to prevent overfitting;
- A softmax output layer to produce class probabilities.

The network received training through an Adam optimizer while applying categorical cross-entropy loss and it stopped training automatically based on validation loss stabilization. All performance metrics demonstrated excellent results through the ensemble. All detection metrics for this network surpassed 99% precision and recall and F1-score and surpassed traditional models during both collective evaluation and distinct class recognition performance. The ensemble modeled attacks effectively while dealing with traffic imbalance because it identified even uncommon attack categories that other detection systems missed. Edge and fog computing frameworks would benefit most from this outcome since they require both precise operation and flexible adaptability capabilities. The ensemble showcases attributes which make it a candidate for future smart security systems by delivering consistent generalization results without repetitive retrainings.

4.3 Results Analysis

The following section includes complete performance testing of various machine learning methods which classify IoT network traffic among normal and harmful types. The correct identification of IoT network traffic remains crucial due to today's complex communication systems which requires it to protect system integrity and support both service quality maintenance and real-time intrusion detection. The experimental design conducts learning method tests by including analysis of the proposed ensemble voting solution that unites XGBoost (XGB) with Random Forest (RF), Deep Learning (DL) and other classifiers to exploit their respective capabilities. Multiple evaluation metrics serve to provide strict benchmarking of the developed models. The accuracy metric enables researchers to determine the number of proper predictions among all responses. Precision emphasizes the true positive instances to all positive predictions to show how accurately the system detects bona fide cases particularly for security systems. Recall as a measure indicates how well the model detects all positive cases it should identify. The F1-score combines precision and recall into one measurement benefiting ratios with unequal classes between positive and negative instances. Hamming Loss serves as one complementary metric to penalize incorrect predictions. Additional metrics are Matthews Correlation Coefficient (MCC), which provides balanced results on class-imbalanced data, Cohen's Kappa, which addresses chance agreement, Balanced Accuracy, which averages class sensitivity, and Log Loss, which penalizes confident incorrect predictions for probabilistic classifiers.

The main classification metrics appear in Table 2. All major classification metrics reached near-perfect values when using ensemble voting methodology because they each reached levels close to 0.9985 including accuracy, precision, recall and F1-score. The model demonstrates its effectiveness in handling diverse traffic conditions with high reliability when it reaches metrics values approaching 0.9985 globally. The ensemble voting approach makes the model robust due to its integration of deep learning pattern recognition methods with RF and XGB decision-tree interpretability and gradient-boosting performance. The ensemble model leads all metrics in core classification numbers from the provided table because of its strong ability to detect frequent and rare patterns in IoT traffic data. The model successfully balances sensitivity and precision according to its high F1-score which creates optimal conditions in an anomaly detection system for attack detection without false alarms. SGD and Naive Bayes models demonstrate solid baseline results so they find utility in edge-device deployment due to their minimal resource requirements. The ensemble model demonstrates superior performance to both SGD and Naive Bayes at their respective task. The linear and quadratic assumptions of LDA and QDA restrict their performance since high-dimensional non-Gaussian traffic distributions do not align well with the assumptions. The flexibility advantage QDA has with its non-linear boundaries does not compensate for its weakness processing noisy or imbalanced real-time network data.

Table 2. Classification Metrics: Accuracy, Precision, Recall, and F1-score

Model	Accuracy	Precision	Recall	F1-score
Ensemble (DL + RF + XGB)	0.9985	0.9985	0.9985	0.9985
SGD Classifier	0.9340	0.9437	0.9340	0.9358
LDA	0.8181	0.8655	0.8181	0.8188
QDA	0.7695	0.9436	0.7695	0.8057
Naive Bayes	0.9351	0.9330	0.9351	0.9285

The additional performance indicators in Table 3 serve to enhance our knowledge about classifier operation. The performance indicators provide evaluation of model confidence, generalization ability and adaptability to imbalanced classes. The ensemble model's robust nature is supported through evaluation metrics that include MCC and Kappa since these metrics perform well under class imbalance conditions. The strong relationship between predicted classes and actual classes can be observed in the nearly perfect MCC score value of 0.998. The low Hamming Loss score signifies that classifications contain a minimal number of incorrect predictions which stands vital for applications needing reliable performance. A side-by-side bar chart of main metrics from all models appears in Figure 2 for visual enhancement of these findings. The visual diagram lets users instantly examine all models through a uniform graphical format.

The ensemble model presents superior performance in every metric according to the bar chart visual depiction as classical models show conflicting results. The high precision rating from QDA undermines its ability to correctly identify attack cases which leads to a weakened F1-score indicative of conservative prediction behaviors with lowered sensitivity. All models present their classification confusion matrices through Figure 3 for per-class precision analysis. The matrices reveal which particular attack methods are properly identified or

Table 3. Additional Evaluation Metrics: Hamming Loss, MCC, Cohen's Kappa, Balanced Accuracy, and Log Loss

Model	Hamming loss	MCC	Kappa	Balanced accuracy	Log loss
Ensemble (DL + RF + XGB)	0.0015	0.9980	0.998	0.9021	1.6843
SGD Classifier	0.0660	0.9149	0.9139	0.7535	0.2851
LDA	0.1819	0.7732	0.7664	0.6705	1.2223
QDA	0.2305	0.7621	0.7216	0.7866	7.9399
Naive Bayes	0.0649	0.9154	0.9141	0.8908	1.3462

confused by the models. Subfigure 3a shows that the ensemble model delivers near-perfect classification performance because it features negligible off-diagonal elements proving the model successfully distinguishes all classes. The classification methods displayed substantial confusion in other subfigures since they experienced difficulties separating low-volume attack patterns with comparable characteristics. The Naive Bayes classifier demonstrates confusion in its detection of benign flows and stealthy attack patterns because of its inability to model conditional dependencies. SGD classifier produces equally distributed results although it provides incorrect classifications when analyzing complex attack types that overlap with one another. ROC curve analysis was used to evaluate both the generalization capability and the distinctness between classes of the models. Although Figure 4 presents the ROC curves for all models, it illustrates how each model differentiates between classes at several detection thresholds. The ensemble model shows nearly perfect ROC curve performance (subfigure 4a) through its full range of classes containing an Area Under the Curve value of 1.00 which proves its discriminatory ability. The classical models perform worse than the ensemble model in terms of AUC measurements since the AUC values show greater variation when classes are unbalanced or traffic features overlap. The ensemble model demonstrates superior ROC performance compared to its individual classical models thus proving its capability to analyze IoT traffic precisely. The ensemble model demonstrates strong performance in multi-class IoT traffic classification according to evidence presented from tables and charts and statistical results. The model demonstrates robustness and accuracy together with precision which makes it suitable for smart city network systems along with industrial IoT monitoring and edge-based cybersecurity applications.

The evaluation of discriminative power and robustness for each classification model required Precision-Recall (PR) curve generation because these curves suit performance assessment on imbalanced IoT traffic datasets.

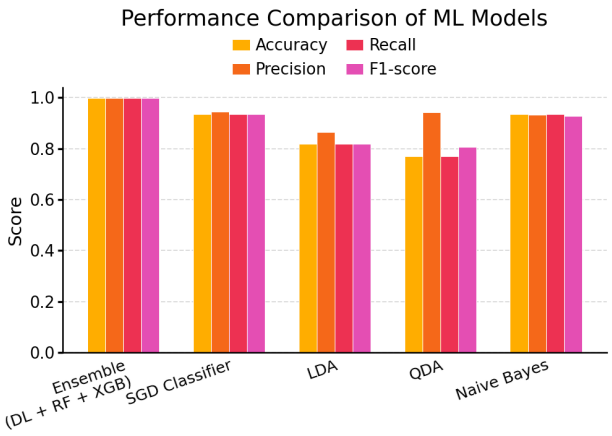


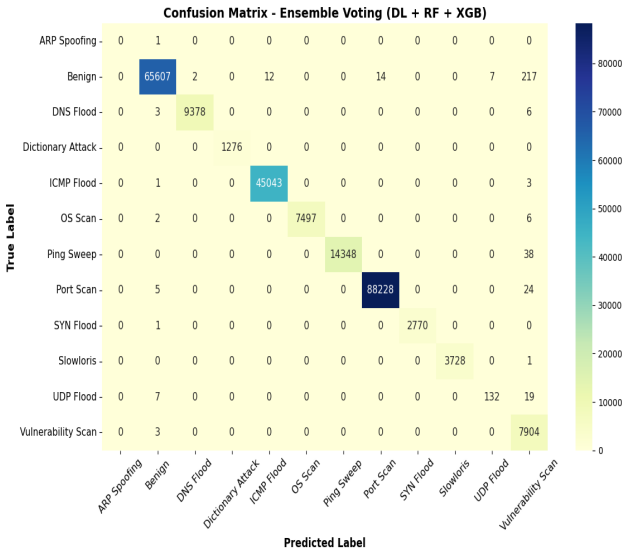
Figure 2. Bar chart comparing Accuracy, Precision, Recall, and F1-score across ML models.

PR curves be-came essential when analyzing security applications since they show how precision relates to recall in detecting posi-tive and negative cases while ROC curves focus on overall classification ability. Figure 5 displays the multiclass PR curves for all evaluated models. The ensemble model (sub-figure 5a) shows practically flawless performance across all classes by positioning its curves close to the top-right corner for both high precision and recall measurements. The PR performance of LDA and QDA fluctuates substantially across different classes because of both unbalanced and partly in-tersecting data distributions. Naive Bayes and SGD achieve better outcomes in particular classes although their performance remains inconsistent when compared to the ensemble structure.

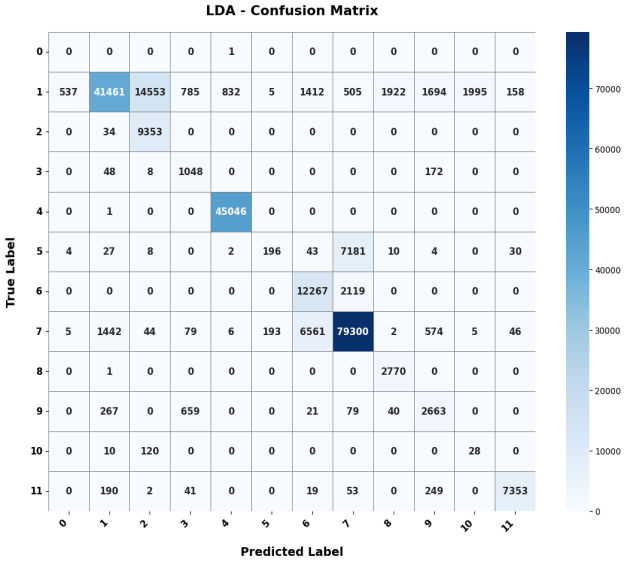
5. DISCUSSION

This research proves that IoT traffic classification systems will succeed by combining multiple algorithms rather than in-dividual approaches. Research findings demonstrate that the ensemble voting model established more than performance excellence by defining itself as an operational model for adap-tive and scalable network security. The ensemble system optimizes several advantages of its Deep Learning, Random Forest and XGBoost components through an integration that reduces each component's limitations. This fusion's delib-erate nature created a solution that the experimental results substantiate. The ensemble model surpassed classical models LDA and QDA, which provided interpretive boundaries, and Naive Bayes, which delivered swift classification processes since it displayed exceptional performance across multiple traffic situations. The ensemble displayed excellent results in data imbalance management and subtle class separation through its F1-score and MCC values approaching perfec-tion. Confusion matrix analysis demonstrated the ensemble model's dominance since it delivered precise, confident re-sults with minimal false results that IoT security environments strictly require.

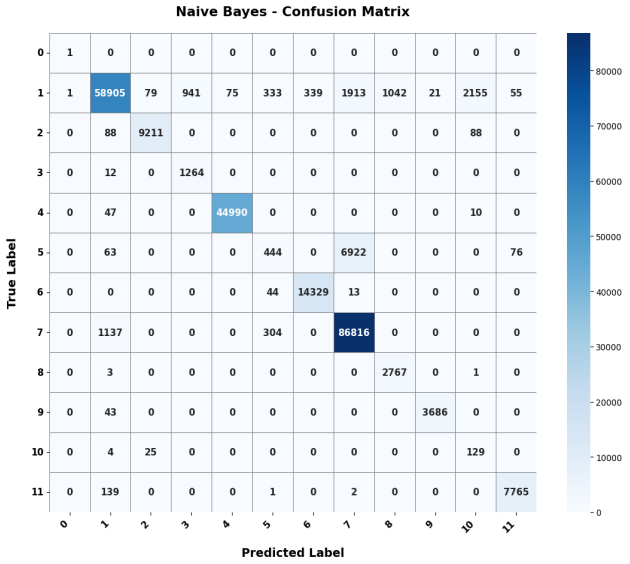
These results gain additional strength from the wide range of validation metrics employed in the study. The model as-sessment benefits from metrics including Hamming Loss and Balanced Accuracy, and conventional accuracy measure-ments because they establish a comprehensive understanding of the model's outcomes. The multi-factor evaluation showed traditional models presented themselves well only in specific assessment criteria yet demonstrated weakness in diverse network scenarios. Data presented in ROC curves showed definitive proof that the ensemble model remains reliable in its behavioral performance because of its superior statistical strength. The model consistently achieved high AUC scores throughout all examination classes due to its ability to iden-tify minor traffic distinctions under multiple conditions. QDA and LDA models lost effectiveness because of class boundary intersections or infrequent attack situations.



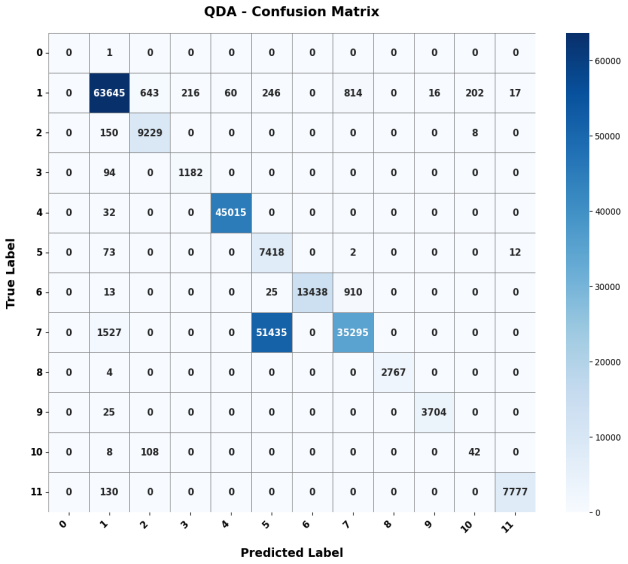
(a) Ensemble (DL + RF + XGB)



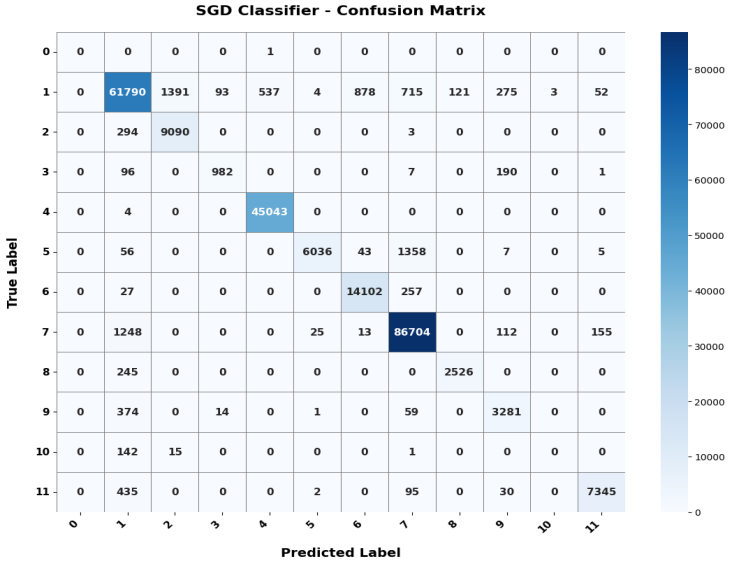
(b) LDA classifier



(c) Naive Bayes classifier

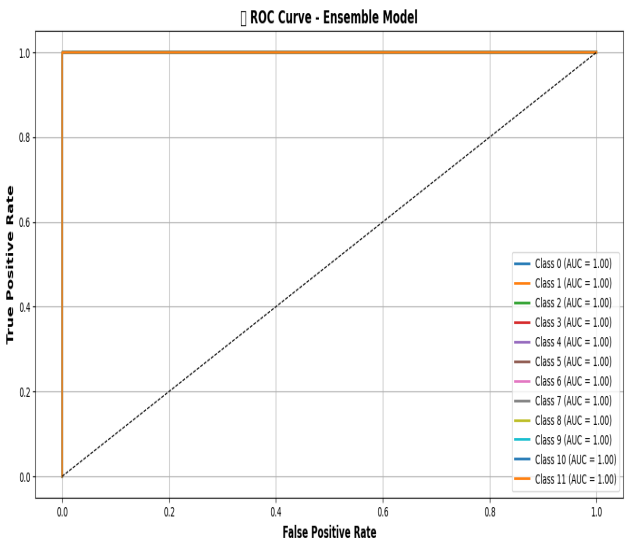


(d) QDA classifier

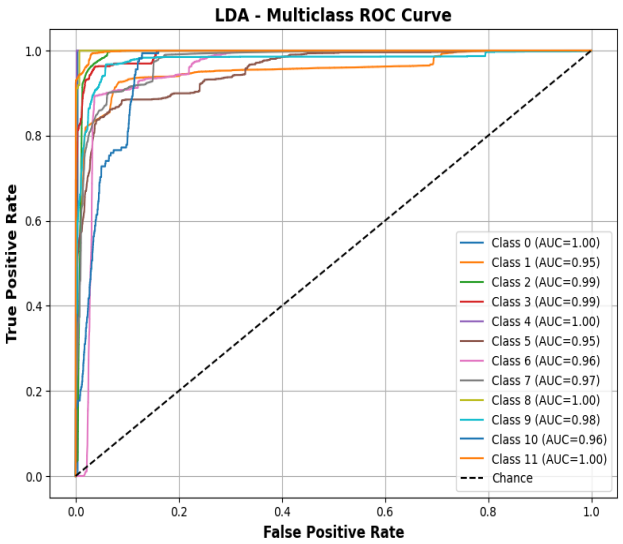


(e) SGD classifier

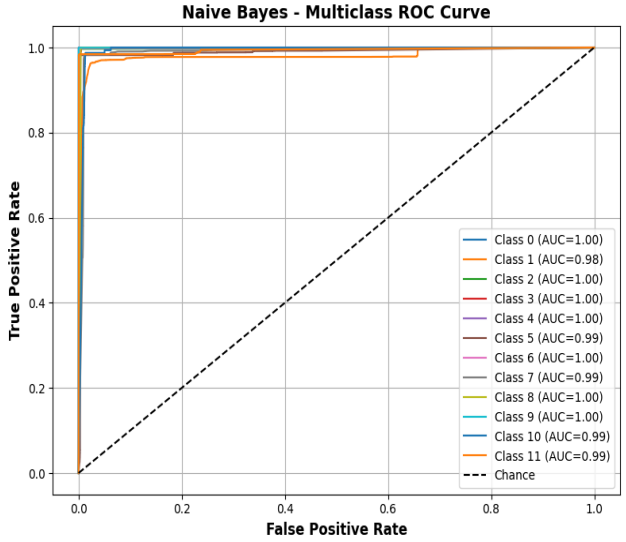
Figure 3. Confusion matrices for all evaluated classifiers.



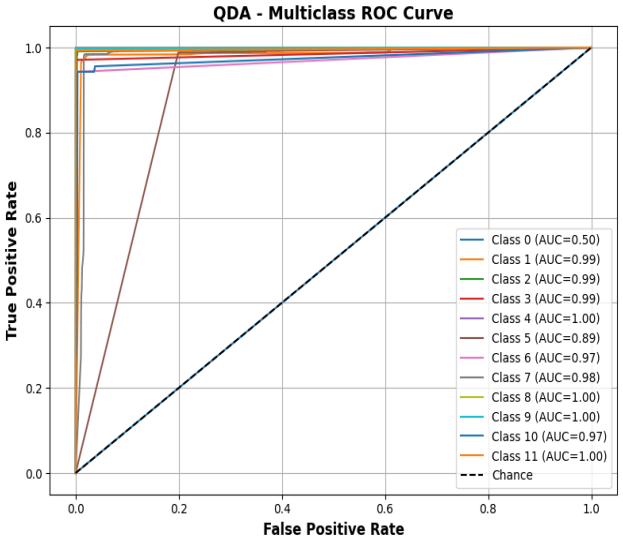
(a) Ensemble model



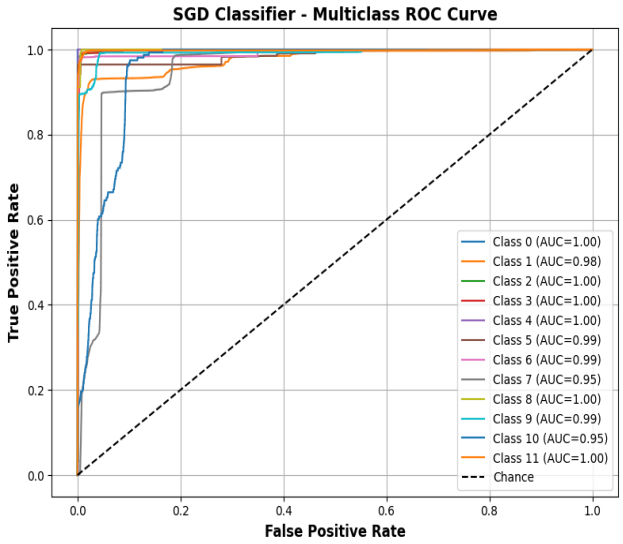
(b) LDA classifier



(c) Naive Bayes classifier

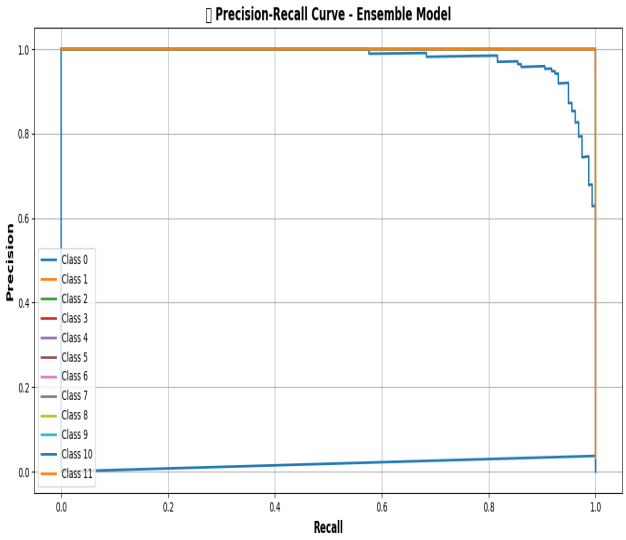


(d) QDA classifier

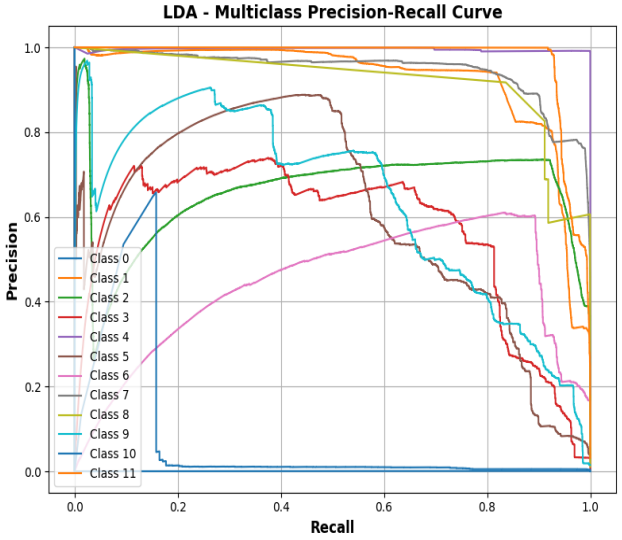


(e) SGD classifier

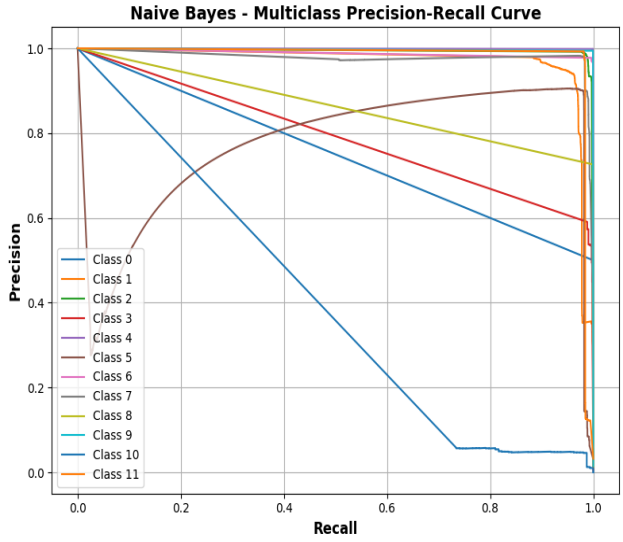
Figure 4. Multiclass ROC curves for all evaluated classifiers.



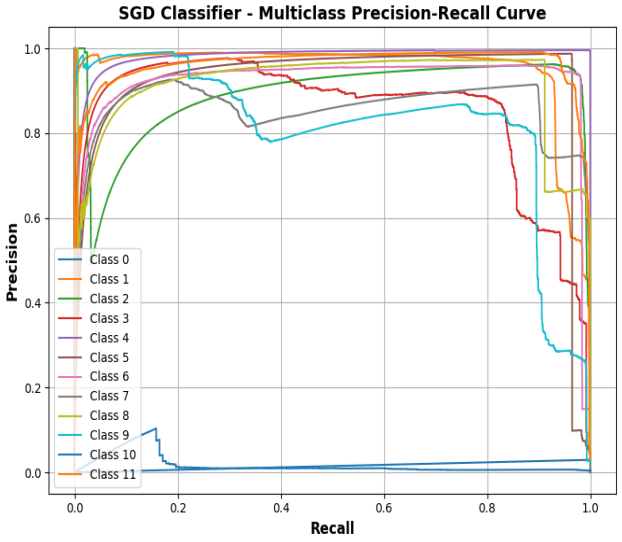
(a) Ensemble model



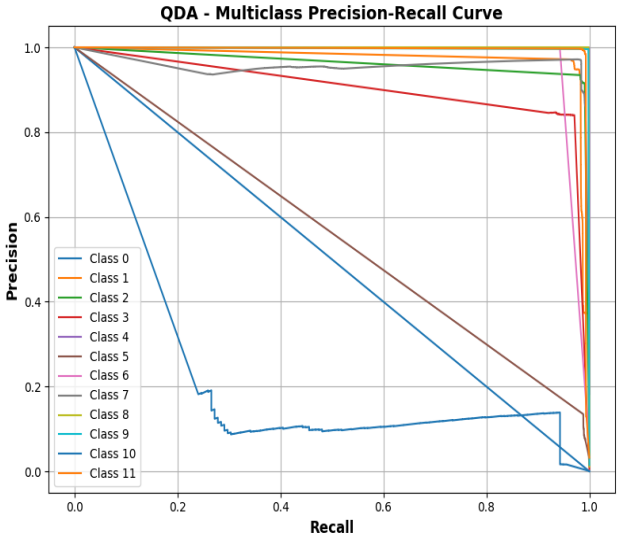
(b) LDA classifier



(c) Naive Bayes classifier



(d) SGD classifier



(e) QDA classifier

Figure 5. Multiclass Precision-Recall curves for all evaluated classifiers.

Future IoT security structures gain guidance from the results obtained from this benchmarking initiative. The ensemble architecture provides both scalability and adaptability because it can automatically respond to traffic profile changes and attack patterns without needing substantial retraining. The ability to adapt is essential because IoT networks continue their expansion into healthcare and manufacturing industries and autonomous systems, which require the prevention of devastating system outages and security violations. Moreover, the research affirms the value of thorough data preparation. The preprocessing series, which included normalization and feature correlation assessment followed by choice-based feature selection, operated as a fundamental factor for model stability and clarity. Broad and deep contextual properties characterize the selected flow-level features, providing rich conditions for pattern detection without unneeded information redundancy. The research shows that future IoT traffic classifiers should consolidate individual experiments through ensemble-based systems. This proposed system provides accuracy, deployment readiness, and resistance to disruptions. Intelligent hybrid solutions like this will serve as essential infrastructure for creating secure, future-proof network environments because of growing IoT traffic complexity and increasing attackers' sophistication.

6. CONCLUSION AND FUTURE WORK

Researchers analyzed the rising need to identify Internet of Things (IoT) network traffic by developing a new framework incorporating deep learning, Random Forest, and XGBoost algorithms through weighted voting algorithms. The model received evaluation using a flow-level dataset from the Army Cyber Institute (ACI) that showed a variety of network behaviors between benign and malicious activities. A complete series of data preparation steps were used beforehand for model training that involved removing anomalies and standardizing features, introducing category identifiers, and omitting features that showed weak correlations. The data preprocessing step protected the reliability and valid representation of training information. The experiments used multiple performance measures such as accuracy, precision, recall, F1-score, Hamming loss, Matthews correlation coefficient (MCC) and balanced accuracy to evaluate how the model operated. All evaluation aspects showed that the ensemble model delivered superior performance beyond traditional LDA, QDA, Naive Bayes, and SGD classifiers. The nonlinear nature and the multi-dimensionality of IoT traffic patterns demonstrate the effectiveness of ensemble learning through the obtained experimental results. Attribute evaluation through feature analysis and correlation heatmap established the diverse and abundant nature of selected features, leading to excellent generalization performance during evaluation. The ensemble architecture used different learning-biased models as components to counteract overfitting and achieve robust performance in unstable, high-variability environments.

Future research will focus on different possible strategies beyond current findings. The first extension includes adding real-time traffic streaming capabilities to the classification pipeline for efficient on-site device inference throughout edge systems. The proposed system design would optimize performance for time-critical systems, including anomaly detection systems in residential homes and industrial process controls. The ensemble system would benefit from incorporating various deep learning frameworks, which include CNNs and RNNs, to boost the recognition of temporal patterns in IoT traffic. The model adaptability can be enhanced through unsupervised and semi-supervised learning methods when the data is sparse or unbalanced. Research dealing with adversarial traffic effects on models should be conducted together with investigations into adversarial training, transfer learning, and federated learning methods to strengthen models while maintaining secure deployments. The ensemble approach is a resilient, scalable solution for intelligent traffic categorization in IoT networks. The research demonstrates new standards because multiple classifier integration produces accurate results and flexibility for real-time deployment in Internet of Things network security systems.

REFERENCES

[1] W. Jiang, "Cellular traffic prediction with machine learning: A survey," *Expert Systems with Applications*, vol. 201, p. 117163, 2022.

[2] G. Sripriyanka and A. Mahendran, "Mirai botnet attacks on IoT applications: Challenges and controls," In *International Conference on Information Systems and Management Science*, Springer, pp. 49–67, 2021.

[3] Y. Xiong, S. Dong, R. Liu, F. Shi, and X. Jing, "IoT network traffic classification: A deep learning method with fourier transform-assisted hyperparameter optimization," *Frontiers in Physics*, vol. 11, p. 1273862, 2023.

[4] N. A. Hung, "Traffic modeling and anomaly detection for internet of things," Ph.D. dissertation, Shibaura Institute of Technology, 2021.

[5] M. Snehi and A. Bhandari, "Introspecting diverse IoT-traffic analysis methods in smart environments and prospects," In *IEEE International Conference on Data Science and Information System (ICDSIS)*, IEEE, pp. 1–5, 2022.

[6] Z. K. Maseer, R. Yusof, S. A. Mostafa, N. Bahaman, O. Musa, and B. A. S. Al-Rimy, "DeepIoT.IDS: Hybrid deep learning for enhancing IoT network intrusion detection," *Computers, Materials and Continua*, vol. 69, no. 3, pp. 3946–3967, 2021.

[7] M. Pons, E. Valenzuela, B. Rodríguez, J. A. Nolazco-Flores, and C. Del-Valle-Soto, "Utilization of 5g technologies in IoT applications: Current limitations by interference and network optimization difficulties—a review," *Sensors*, vol. 23, no. 8, p. 3876, 2023.

[8] V. D. Jangampet, S. R. Pulyala, and A. G. Desetty, "Optimized alternating graph-regularized neural network for cyber security threats detection in internet of things," *International Journal of Information Security*, vol. 2, no. 1, 2023.

[9] M. Nicholson, R. Agrahari, C. Conran, H. Assem, and J. D. Kelleher, "The interaction of normalisation and clustering in sub-domain definition for multi-source

transfer learning based time series anomaly detection," *Knowledge-Based Systems*, vol. 257, p. 109894, 2022.

[10] M. Habiba, M. R. Islam, S. M. Muyeen, and A. B. M. S. Ali, "Edge intelligence for network intrusion prevention in IoT ecosystem," *Computers and Electrical Engineering*, vol. 108, p. 108727, 2023.

[11] G. Chakrapani, G. S. Rao, and K. V. Krishna, "Securing IoT networks: IoT device identification and classification through network traffic analysis using machine learning," *Turkish Journal of Computer and Mathematics Education*, vol. 14, no. 3, pp. 1021–1028, 2023.

[12] D. Alekseeva, N. Stepanov, A. Veprev, A. Sharapova, E. S. Lohan, and A. Ometov, "Comparison of machine learning techniques applied to traffic prediction of real wireless network," *IEEE Access*, vol. 9, pp. 159 495–159 514, 2021.

[13] M. Hammoudeh, J. Pimlott, S. Belguith, G. Epiphaniou, T. Baker, A. S. M. Kayes, B. Adebisi, and A. Bounceur, "Network traffic analysis for threat detection in the internet of things," *IEEE Internet of Things Magazine*, vol. 3, no. 4, pp. 40–45, 2021.

[14] H. Nguyen and R. F. Kashef, "TS-IDS: Traffic-aware self-supervised learning for IoT network intrusion detection," *Knowledge-Based Systems*, vol. 279, p. 110966, 2023.

[15] J. P. Kushwaha, S. Bhadauria, and S. Tapaswi, "Multimethod stacked feature selection approach based IDS for IoT networks," *Procedia Computer Science*, vol. 230, pp. 564–573, 2023.

[16] R. Lazzarini, H. Tianfield, and V. Charissis, "A stacking ensemble of deep learning models for IoT intrusion detection," *Knowledge-Based Systems*, vol. 279, p. 110941, 2023.

[17] E. C. P. Neto, S. Dadkhah, R. Ferreira, A. Zohourian, R. Lu, and A. A. Ghorbani, "CICIoT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment," *Sensors*, vol. 23, no. 13, p. 5941, 2023.

[18] H. Nguyen-An, T. Silverston, T. Yamazaki, and T. Miyoshi, "IoT traffic: Modeling and measurement experiments," *IoT*, vol. 2, no. 1, pp. 140–162, 2021.

[19] S. B. M. A. T. S. D. V. B. B. I. Neelakandan, M. A. Berlin, S. Tripathi, V. B. Devi, I. Bhardwaj, and N. Arulkumar, "IoT-based traffic prediction and traffic signal control system for smart city," *Soft Computing*, vol. 25, no. 18, pp. 12 241–12 248, 2021.

[20] P. Khandait, N. Hubballi, and B. Mazumdar, "IoTHunter: IoT network traffic classification using device specific keywords," *IET Networks*, vol. 10, no. 2, pp. 59–75, 2021.

[21] O. Salman, I. H. Elhajj, A. Chehab, and A. Kayssi, "A machine learning based framework for IoT device identification and abnormal traffic detection," *Transactions on Emerging Telecommunications Technologies*, vol. 33, no. 3, p. e3743, 2022.

[22] A. I. Owusu and A. Nayak, "An intelligent traffic classification in sdn-IoT: A machine learning approach," In *IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*, IEEE, pp. 1–6, 2020.

[23] R. Kumar, M. Swarnkar, G. Singal, and N. Kumar, "IoT network traffic classification using machine learning algorithms: An experimental analysis," *IEEE Internet of Things Journal*, vol. 9, no. 2, pp. 989–1008, 2021.

[24] S. P. Khedkar and R. AroulCanessane, "Machine learning model for classification of IoT network traffic," In *Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)*, IEEE, pp. 166–170, 2020.

[25] A. P. Joshi and B. V. Patel, "Data preprocessing: The techniques for preparing clean and quality data for data analytics process," *Oriental Journal of Computer Science and Technology*, vol. 13, no. 2–3, pp. 78–81, 2021.