IRSTI 81.93.29

10.51889/2959-5894.2025.92.4.013

G. Aksholak[1*], R. Magazov[1]

[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan
[*]*e-mail: gaksholak@gmail.com*

APPLICATION OF MACHINE LEARNING METHODS TO ANALYZE MALICIOUS NETWORK TRAFFIC

Abstract

The increasing deployment of Internet of Things (IoT) devices has made networks more vulnerable to malicious activities, necessitating effective traffic classification methods. This study investigates machine learning-based approaches to classify network traffic using the CTU-IoT-Malware-Capture-3-1conn.log.labeled dataset, which includes over 150,000 records labeled as benign or malicious. Key numeric features such as packet counts, data volumes (bytes), protocol types, and source IP frequencies were selected to enhance the models' predictive capabilities. Preprocessing involved normalization with StandardScaler to ensure equal contribution of features to model predictions. Four machine learning algorithms were evaluated: Logistic Regression, Random Forest, Support Vector Machines (SVM), and Gradient Boosting. The experiments utilized a 5-fold cross-validation framework to ensure robustness and reliability. Gradient Boosting outperformed other models with a mean accuracy of 99.13%, followed by Random Forest at 99.11%. To address class imbalance, SMOTE was applied, significantly improving the recall of minority classes. This study demonstrates the potential of machine learning for improving IoT network security by identifying anomalous behaviors in real-world attack scenarios. The results underline the importance of preprocessing, feature selection, and evaluation in achieving high detection accuracy.

Keywords*:* IoT security, malicious traffic detection, machine learning, feature selection, SMOTE, anomaly detection, cybersecurity.

Г.И. Ақшолақ[1], Р.С. Мағазов[1]
[1]Әл-Фараби атындағы Қазақ Ұлттық Университеті, Алматы қ., Қазақстан

ЗИЯНДЫ ЖЕЛІЛІК ТРАФИКТІ АНЫҚТАУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІН ҚОЛДАНУ

Аңдатпа

IoT құрылғыларының көбеюі желілерді зиянды әрекеттерге осал етті, бұл трафикті жіктеудің тиімді әдістерін қолдануды талап етеді. Бұл зерттеу CTU-IoT-Malware-Capture-3-1conn деректер жинағын пайдалана отырып, желілік трафикті жіктеудің машиналық оқытуға негізделген тәсілдерін қарастырады. Деректер қоры қатерсіз немесе зиянды деп белгіленген 150 000-нан астам жазбаларды қамтиды. Модельдердің болжамды мүмкіндіктерін кеңейту үшін пакеттер саны, деректер көлемі (байттар), Протокол түрлері және дереккөздердің IP мекенжай жиіліктері сияқты негізгі сандық сипаттамалар таңдалды. Деректер модель болжамдарына өнімділіктің тең үлесін қамтамасыз ету үшін StandardScaler көмегімен алдын ала қалыпқа келтірілді. Машиналық оқытудың төрт алгоритмі бағаланды: логистикалық регрессия, кездейсоқ орман, тірек векторлық әдіс (SVM) және градиенттік үдеу. Тәжірибелер сенімділікті қамтамасыз ету үшін 5 есе кросс-тексеру жүйесі қолданылды. Градиенттік үдеу басқа модельдерге қарағанда жоғарғы дәлдікті көрсетті (99,13%), содан кейін кездейсоқ орман (99,11%). Деректердегі теңгерімсіздікті жою үшін SMOTE алгоритмі қолданылды, бұл азшылықтардың кері байланысын едәуір жақсартты. Бұл зерттеу нақты шабуыл сценарийлеріндегі қалыптан тыс мінез-құлықты (anomalous behaviors) анықтау арқылы заттардың интернет желісінің қауіпсіздігін жақсарту үшін машиналық оқытудың әлеуетін көрсетеді. Нәтижелер жоғары анықтау дәлдігіне қол жеткізу үшін алдын ала өңдеудің, өнімділікті таңдаудың және бағалаудың маңыздылығын көрсетеді.

Түйін сөздер*:* Интернет заттарының қауіпсіздігі, зиянды трафикті анықтау, Машиналық оқыту, мүмкіндіктерді таңдау, SMOTE, аномалияларды анықтау, киберқауіпсіздік.

Г.И. Акшолак[1], Р.С. Магазов[1]
[1]Казахский Национальный Университет имени аль-Фараби, г.Алматы, Казахстан

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА ВРЕДОНОСНОГО СЕТЕВОГО ТРАФИКА

Аннотация

Растущее число устройств Интернета вещей сделало сети уязвимыми для вредоносной деятельности, что потребовало использования эффективных методов классификации трафика. В этом исследовании рассматриваются подходы, основанные на машинном обучении, для классификации сетевого трафика с использованием набора данных CTU-IoT-Malware-Capture-3-1conn. База данных содержит более 150 000 записей, помеченных как доброкачественные или вредные. Чтобы расширить возможности прогнозирования моделей, были выбраны основные числовые характеристики, такие как количество пакетов, объем данных (байты), типы протоколов и частоты IP-адресов источников. Данные были предварительно нормализованы с помощью StandardScaler, чтобы обеспечить равную долю производительности в прогнозах модели. Были оценены четыре алгоритма машинного обучения: логистическая регрессия, случайный лес, метод опорных векторов и градиентное ускорение. Эксперименты 5-кратная система перекрестного тестирования использовалась для обеспечения надежности. Градиентное ускорение показало более высокую точность, чем другие модели (99,13%), за которыми следует случайный лес (99,11%). Алгоритм SMOTE использовался для устранения дисбаланса данных, что значительно улучшило обратную связь меньшинств. Это исследование демонстрирует потенциал машинного обучения для повышения безопасности Интернета вещей путем выявления аномального поведения (anomalous behaviors) в реальных сценариях атак. Результаты подчеркивают важность предварительной обработки, выбора производительности и оценки для достижения высокой точности обнаружения.

Ключевые слова: безопасность Интернета вещей, обнаружение вредоносного трафика, машинное обучение, выбор функций, SMOTE, обнаружение аномалий, кибербезопасность.

Introduction

The integration of Internet of Things (IoT) technology across various industrial sectors has brought considerable advances in automated processes and network connectivity. At the same time, this proliferation of IoT devices has exposed networks to new security risks. A common characteristic of many IoT deployments is insufficient built-in security, which makes them vulnerable to exploitation through attack vectors such as Distributed Denial of Service (DDoS), data exfiltration, and port scanning.

Current threat statistics paint a concerning picture. The AV- TEST Institute records over 450,000 new malware and un- wanted applications daily [1]. The 2025 Astra Security Report documents 560,000 daily malware detections [2], indicating a significant rise. This surge underscores the urgent need for automated methods that can detect threats intelligently. Prior studies [3], [4] emphasize that handling these challenges requires robust anomaly detection to identify malicious traffic quickly.

Anomaly detection involves defining patterns of “normal” behavior based on business operations and identifying deviations that may indicate potential threats [5]. Manual monitoring becomes practically impossible in IoT environments due to the massive volume and variety of data traffic these networks generate. Machine learning offers a solution by automating the detection of anomalies, which helps organizations identify and respond to attacks more quickly. Previous research has demonstrated the value of this approach – for example, Rana showed how real-time anomaly detection can effectively counter threats like DoS attacks [6].

The aim of the present study is to assess the capability of several supervised machine learning classifiers to differentiate between benign and malicious IoT traffic. Logistic Regression, Random Forest, Support Vector Machine, and Gradient Boosting models are examined using a labeled dataset representing realistic IoT attack scenarios. The evaluation framework includes data normalization, correlation analysis, and class balancing, allowing the assessment of model behavior under conditions that reflect real-world intrusion detection requirements.

This research provides several key contributions:

- A comparative evaluation of four supervised ML algorithms for IoT malicious traffic classification using real- world network data
- An empirical analysis of feature importance and the effect of normalization on model accuracy
- The integration of SMOTE to address severe class imbalance, significantly improving recall for minority attack classes
- A robust methodology that can be extended toward lightweight, real-time IoT intrusion detection systems

The remainder of this paper is organized as follows: Section II reviews literature review and research gaps; Section III presents the proposed methodology; Section IV details the experimental results; Section V discusses findings and limitations; and Section VI concludes with future research directions.

Literature review

Detecting anomalies in network traffic plays a central role in cybersecurity. When systems can identify deviations from normal behavior, they can spot new types of threats that signature-based methods might miss. This capability has made anomaly detection an active research area, particularly as machine learning techniques have matured.

A. Machine Learning-Based Intrusion Detection

Almutairi et al. explored machine learning methods for IDS in IoT networks [7]. They assessed various ML algorithms, including SVMs, Random Forests, and J48 Decision Trees, finding that Random Forests provided the best balance be- tween performance and computational efficiency. Similarly, Hussain et al. identify challenges in resource-constrained IoT environments, including the need for lightweight and real-time detection systems [8].

As Iglesias and Zseby note [9], efficient feature selection not only reduces computational costs but also enhances detection accuracy. Disha and Waheed demonstrated the effectiveness of the Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique in reducing dimensionality while maintaining high classification accuracy [10]. Fernandes et al. provided a comprehensive survey on network anomaly detection, highlighting the importance of feature engineering for model efficiency [11].

Wang et al. explored encrypted malicious traffic detection using flow-based machine learning approaches [12]. Their work highlighted the effectiveness of Random Forest classifiers for achieving an optimal balance between detection accuracy and computational cost, making them suitable for large-scale IoT network environments.

B. Deep Learning and Hybrid Approaches

To overcome the limitations of static ML models, several studies explored deep learning for capturing temporal and spatial dependencies in network data. Radford et al. highlighted the suitability of Long Short-Term Memory (LSTM) for sequence-based anomaly detection in network traffic, achieving significant performance improvements with AUC scores of 0.84 on benchmark datasets [13]. Fotiadou et al. demonstrated the effectiveness of Convolutional Neural Networks (CNNs) and LSTMs for analyzing sequential traffic data [14].

Balyan et al. proposed a hybrid intrusion detection model combining Enhanced Genetic Algorithm (EGA) and Particle Swarm Optimization (PSO) with an improved Random Forest (IRF) method [15]. Their model effectively tackled data imbalance issues by enhancing minor data samples, achieving high accuracy (98.979%) for binary classification using the NSL-KDD dataset.

Fu et al. introduced Whisper, a system leveraging frequency- domain features for real-time malicious traffic detection, achieving robustness against diverse attack types [16]. Complementarily, Koumar et al. proposed NetTiSA, a compact feature set optimized for high-speed network traffic classification [17].

C. Feature Engineering and Data Balancing

Feature selection and data balancing remain major determinants of IDS performance. Alduailij et al. presented a method for DDoS attack detection in cloud computing using Mutual Information and Random Forest Feature Importance techniques [18]. Their results showed that Random Forests and

Gradient Boosting outperformed other methods in accuracy and feature relevance. Subbiah et al. employed Boruta feature selection with a Grid Search Random Forest to develop an IDS for wireless sensor networks, achieving 99% accuracy [19]. Several studies emphasized the role of Synthetic Minority Oversampling Technique (SMOTE) to mitigate class imbalance, which often causes models to favor benign traffic. This preprocessing method has consistently improved recall for minority attack categories, as confirmed by Al-Amri et al. [20].

D. Research Gap

Although prior works have achieved promising results, several limitations persist. Many models rely on simplified or filtered datasets, failing to represent the full diversity of IoT network traffic. Deep learning-based approaches, while accurate, are computationally intensive, making them unsuitable for real-time or edge deployments. Few studies conduct systematic evaluations comparing classical and ensemble ML models under balanced, standardized preprocessing conditions. The impact of feature scaling and class balancing techniques has not been consistently quantified across models.

To bridge these gaps, the present study conducts a comprehensive comparative analysis of four supervised ML algorithms – Logistic Regression, SVM, Random Forest, and Gradient Boosting–on the CTU-IoT-Malware-Capture-3-1 dataset. The proposed framework emphasizes feature normalization, correlation analysis, and class rebalancing, providing new insights into optimizing ML-based intrusion detection for heterogeneous IoT environments.

Research methodology

The proposed system follows a multi-layered architecture consisting of data acquisition, preprocessing, feature selection, and machine learning model evaluation. Each component is designed to handle specific aspects of IoT network traffic classification, as illustrated in Fig. 1.

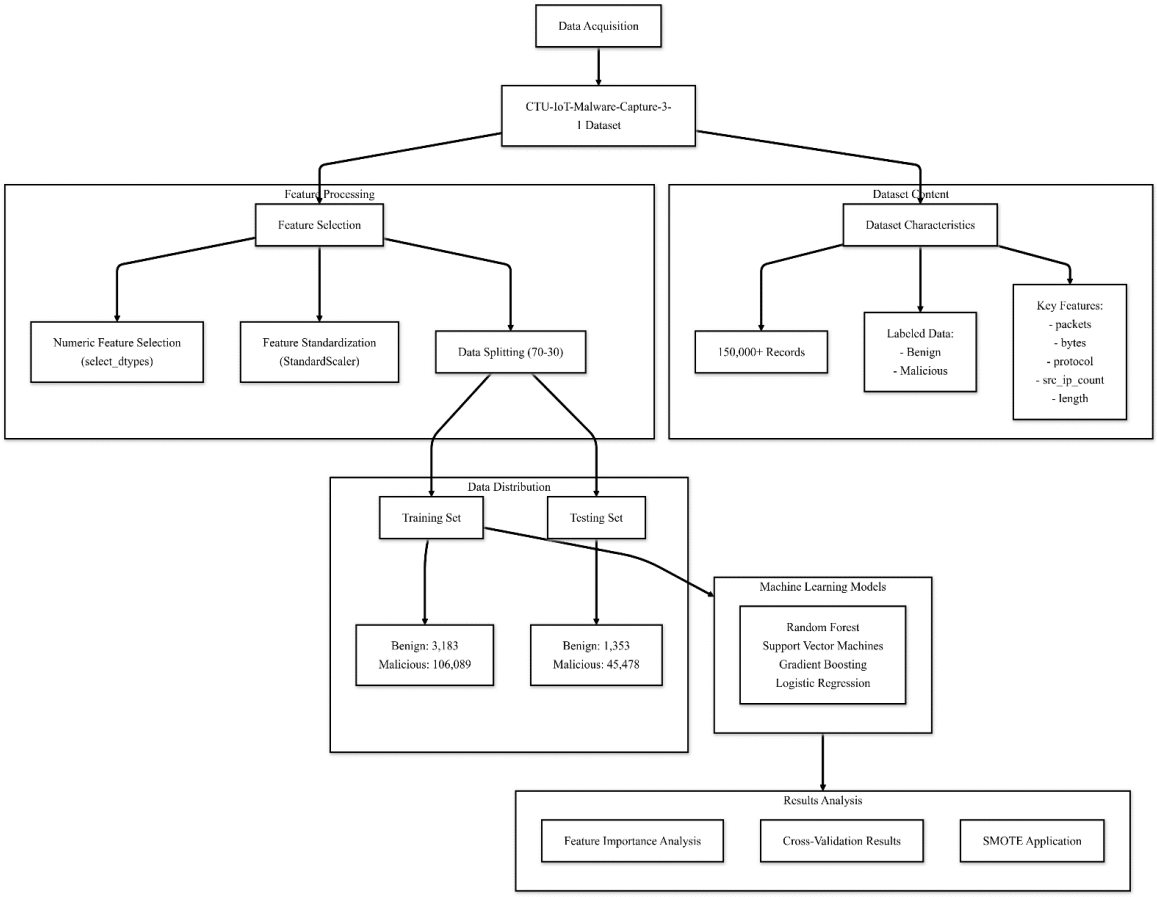

Fig. 1. Architecture of the developed system

A. Data Description

The CTU-IoT-Malware-Capture-3-1 dataset [21], a part of the IoT-23 dataset, is a comprehensive labeled collection of network traffic data designed for analyzing malicious activities in IoT environments. It contains over 150,000 records of network connections, each labeled as either Benign (normal traffic) or Malicious (malicious traffic), enabling clear differentiation for machine learning tasks.

This dataset includes key features such as:

- packets: Number of packets in a connection, where high counts can indicate Distributed Denial of Service (DDoS) attacks.
- bytes: Total data volume, often suggesting potential data exfiltration in cases of large values.
- protocol: Protocol type (e.g., TCP, UDP, ICMP), aiding in understanding connection behavior.
- src_ip_count: Frequency of requests from a single IP, which may point to scanning activities.
- length: Packet size, helpful for identifying anomalies in traffic patterns.

B. Data Preprocessing

Data preprocessing ensures consistency and comparability across all features. The following steps were performed:

1) Numeric Filtering: Only numeric columns were retained using the select_dtypes function to prevent categorical encoding bias.

2) Standardization: All numeric features were normalized via StandardScaler from scikit-learn to achieve zero-mean and unit-variance scaling:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

where, μ is the mean and σ is the standard deviation. This process prevents features with large ranges from dominating model training.

3) Data Splitting: The dataset was divided into 70% training and 30% testing subsets.

4) Balancing with SMOTE: To mitigate severe class imbalance (≈3k benign vs 106k malicious), the Synthetic Minority Oversampling Technique (SMOTE) was applied to equalize both classes. This step significantly improved recall for minority samples and stabilized learning across folds.

C. Feature Selection and Correlation Analysis

Feature selection is a critical step in optimizing machine learning models. It involves identifying the most relevant attributes to enhance model performance, interpretability, and computational efficiency. We implemented specific methods to preprocess the dataset and ensure that features were suitable for machine learning tasks.

Numeric Feature Selection

Using the select_dtypes method from the pandas library, we filtered numeric features from the dataset:

numeric_dataset = dataset.select_dtypes(include='number')

This ensures that only numeric data is used, simplifying input for machine learning algorithms and avoiding errors during training. Key numeric features included:

- packets: Indicates potential DDoS attacks due to high counts.
- bytes: Suggests data exfiltration when values are unusually large.
- src_ip_count: High frequencies from a single IP may indicate scanning activity.

These features provide a direct correlation with network behaviors and serve as strong indicators of malicious activity.

D. Machine Learning Models and Evaluation

The study evaluates several machine learning models, each chosen for its specific strengths in detecting network anomalies:

- Random Forest (RF): Known for its robustness and ability to handle high-dimensional data effectively.
- Support Vector Machines (SVM): Effective in binary classification with non-linear data distributions.
- Gradient Boosting (GB): Provides iterative improvements, optimizing performance for complex datasets.
- Logistic Regression (LR): A simple yet effective baseline model for binary classification, offering probabilistic insights into predictions.

Each model was trained using 5-fold cross-validation to en sure robustness and to prevent overfitting. Evaluation metrics included accuracy, precision, recall, F1-score, and AUC (Area Under Curve), defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

Results of the study

This section presents a detailed analysis of the results, including visualizations and comparisons in accordance with the research methodology. Key observations and trends are contextualized by comparing them with relevant prior studies.

A. Feature Correlation Analysis

Figure 2 illustrates the correlation of features with malicious traffic. Features such as id.orig_p and id.resp_p exhibit the highest correlations, indicating their importance in detecting anomalies. In contrast, the missed_bytes feature shows minimal correlation, suggesting limited utility for prediction purposes.

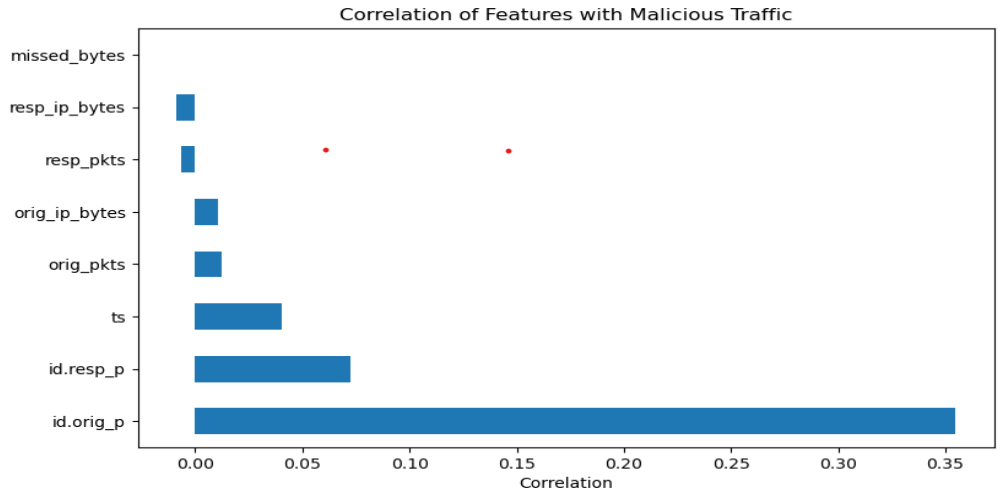

Fig. 2. Correlation of Features with Malicioul Traffic

B. Impact of Feature Scaling

Fig. 3 compares model performance with and without feature scaling. Logistic Regression and SVM demonstrated significant performance improvements post-scaling, achieving higher precision and recall. Random Forest and Gradient Boosting, being tree-based models, were less affected by scaling.

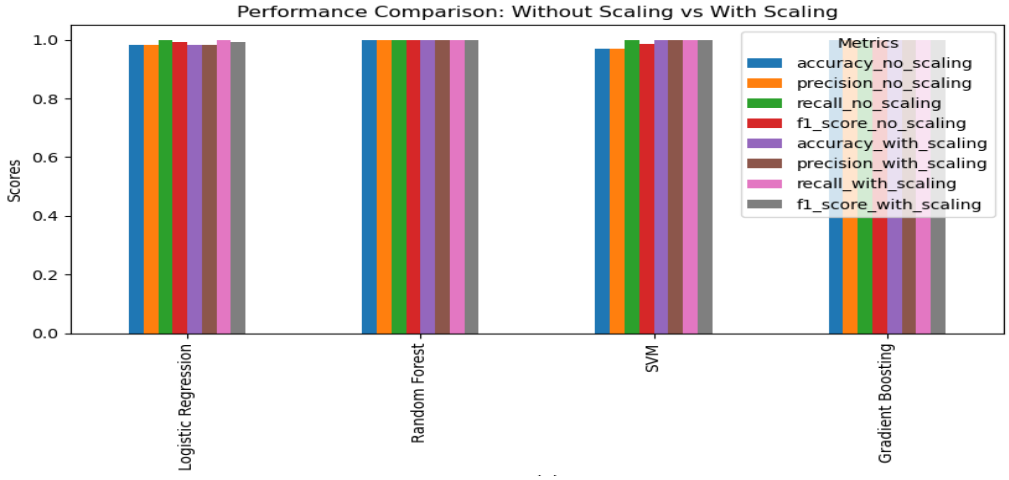

Fig. 3. Model performance with and without feature scaling

Feature scaling has proven essential for models sensitive to feature ranges, such as Logistic Regression and SVM. Similar findings in previous works validate the benefit of preprocessing in achieving balanced and accurate predictions.

C. SMOTE Application Results

Class imbalance was addressed using SMOTE, balancing the number of benign and malicious samples. Before SMOTE, the dataset had 3183 benign and 106,089 malicious samples. Post-SMOTE, both classes had equal representation (106,089 samples each), significantly improving model recall.

Fig. 4 highlights the effect of SMOTE on class distribution, balancing the originally skewed dataset. Model performance, particularly recall and F1-score, improved significantly after addressing class imbalance.

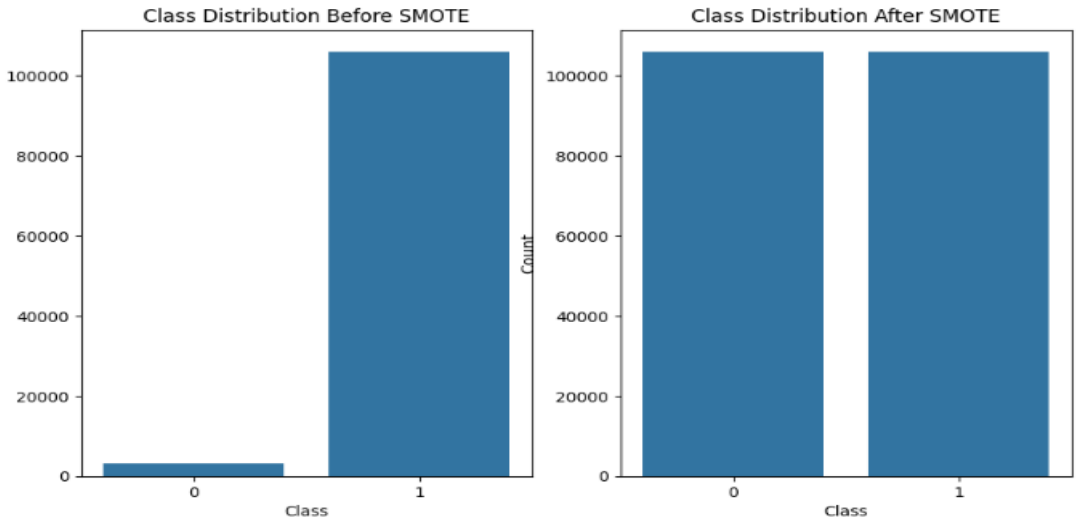

Fig. 4. Effect of SMOTE on Class Distribution

D. Cross-Validation Results

Cross-validation ensured reliable evaluation, preventing overfitting and confirming model robustness across different data splits. Using 5-Fold Cross-Validation, the models were evaluated for stability and consistency. Table 1 displays the result of the accuracy of four algorithms.

Table 1. Accuracy results of the LR, RF, SVM and Gradient Boosting

5 Fold	*LR*	*RF*	*SVM*	*GB*
1	*98.55%*	*96.03%*	*97.09%*	*96.03%*
2	*98.69%*	*99.67%*	*97.09%*	*99.79%*
3	*97.89%*	*99.91%*	*97.09%*	*99.91%*
4	*97.93%*	*99.97%*	*97.09%*	*99.96%*
5	*98.61%*	*99.95%*	*97.09%*	*99.94%*
Mean Accuracy	*98.34%*	*99.11%*	*97.09%*	*99.13%*

The results, shown in Fig. 5, demonstrate that Gradient Boosting achieved the highest mean accuracy score of 99.13%, followed closely by Random Forest with an accuracy of 99.11%. The Logistic Regression algorithm yielded an accuracy of 98.34%, while SVM achieved a slightly lower accuracy of 97.09%.

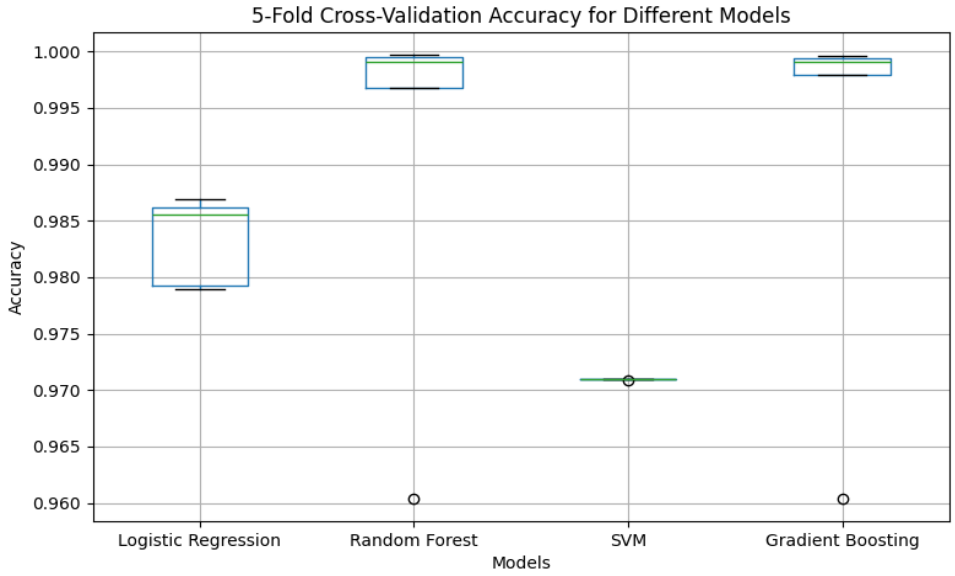

Fig. 5. Distribution of 5-fold cross-validation accuracy

These findings indicate that Gradient Boosting and Random Forest are particularly effective in handling complex datasets and providing robust classification performance, whereas Logistic Regression serves as a strong baseline method. SVM, while slightly less accurate, remains consistent in its performance.

E. Computational Cost Evaluation

To assess the feasibility of deploying the evaluated machine learning models in resource-constrained IoT environments, measurements were conducted for both training time and inference latency for each classifier using the same prepro cessing pipeline (StandardScaler and SMOTE). The results are summarized in Fig. 6.

Training Time. Logistic Regression achieved the fastest training time (1.40~s), followed by Random Forest (5.46 ~s) and Gradient Boosting (19.08~s). SVM required 488.46 s, confirming its impracticality for large-scale or real-time IoT deployments.

Inference Latency. Logistic Regression also achieved the lowest inference time (1.30 ~ms). Gradient Boosting and Random Forest produced moderate delays (24.84 ~ms and 80.81 ~ms, respectively), while SVM exhibited prohibitively high latency (5136.24 ~ms).

Prediction Throughput. Logistic Regression reached the highest throughput (35.95M predictions/s), followed by Gradient Boosting (1.89M/s) and Random Forest (0.58M/s). SVM yielded only 9,118 predictions/s.

Accuracy–Efficiency Trade-off. Gradient Boosting pro vides the best balance between accuracy and computational efficiency. Random Forest offers competitive performance at lower cost, while Logistic Regression remains suitable for lightweight edge-level IoT devices. SVM demonstrated excessive overhead and is unsuitable for real-time IoT intrusion detection.

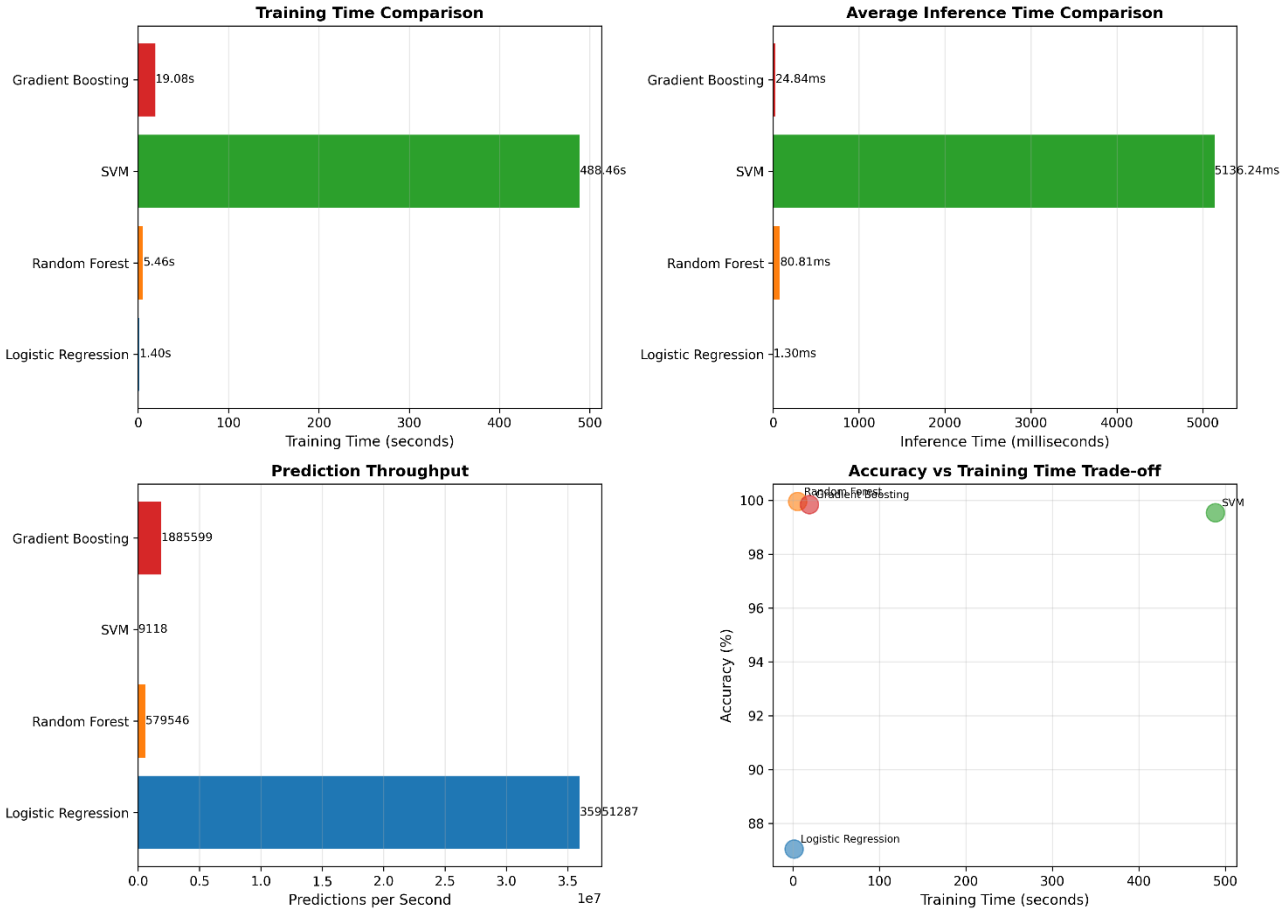

Fig. 6. Computational cost analysis comparing (a) training time, (b) average inference time, (c) prediction throughput and (d) accuracy-efficiency trade-off

Discussion

The findings of this study underscore the potential of machine learning algorithms in detecting malicious IoT network traffic. By leveraging a well-labeled dataset, we demonstrated the effectiveness of preprocessing steps such as feature scaling and SMOTE to address challenges like imbalanced data and feature dominance.

5-fold cross-validation ensured robust model evaluation, reducing the likelihood of overfitting. As shown in Table 1, Gradient Boosting exhibited minimal variance across folds, indicating its stability for this classification task.

Features such as id.orig_p and id.resp_p showed strong correlations with malicious traffic, corroborating findings from studies emphasizing the role of port-related features in identifying network anomalies.

Class imbalance, a common challenge in IoT datasets, was effectively mitigated using SMOTE. This approach not only balanced the dataset but also significantly improved recall for the minority class, aligning with the objective of enhancing anomaly detection.

The results validate the applicability of machine learning models in IoT security. However, real-time implementation remains a challenge due to computational constraints. Future research could explore lightweight models or hybrid approaches combining machine learning with deep learning techniques for real-time detection.

The high performance of ensemble models suggests their utility in developing scalable anomaly detection systems for IoT networks. Future work could incorporate additional contextual features, such as temporal patterns, to improve detection capabilities further.

Exploring advanced data augmentation techniques beyond SMOTE could help enhance model performance on imbalanced datasets. The study primarily focused on binary classification, overlooking multi-class scenarios that are often encountered in complex IoT environments. The dataset, while comprehensive, may not fully represent all potential attack vectors in real-world networks, limiting the generalizability of the findings.

Conclusion

This research highlights the potential of machine learning algorithms in detecting malicious IoT network traffic. By leveraging a well-curated dataset and employing advanced preprocessing techniques, including feature selection and normalization, we developed effective models for binary traffic classification. Among the evaluated algorithms, Gradient Boosting demonstrated superior performance, achieving an accuracy of 99.13%. The application of SMOTE addressed class imbalance, significantly enhancing recall for minority classes and improving the overall robustness of the models.

The study establishes a comprehensive framework for malicious traffic detection, emphasizing the critical role of preprocessing and model evaluation through cross-validation. The findings contribute to IoT security by offering scalable solutions for real-time traffic monitoring and anomaly detection. Future research should focus on incorporating deep learning techniques and real-time implementation to further enhance detection capabilities and address evolving cyber threats. This study serves as a foundation for advancing IoT network security, bridging the gap between academic research and industrial applications.

This work contributes to the field by providing a robust methodology for malicious traffic detection in IoT environments. It highlights the significance of preprocessing steps, feature importance analysis, and the integration of machine learning models tailored for imbalanced datasets. The findings not only validate the efficacy of ensemble models in IoT security but also provide a foundation for further exploration, such as incorporating real-time detection and hybrid approaches that combine machine learning with deep learning methods.

By addressing the critical challenge of class imbalance and demonstrating the effectiveness of advanced machine learning techniques, this study offers valuable insights for both researchers and practitioners working to enhance IoT network security.

References

[1] AV-TEST Institute, "Malware Statistics," 2025. Available online: https://www.av-test.org/en/statistics/malware/ (accessed on 17 June 2025). (in English)

[2] R. Goyal, 30+ Malware Statistics You Need To Know In 2025, Available online: https://www.getastra.com/blog/security-audit/malware-statistics/ (accessed on 1 February 2025). (in English)

[3] Nassif, A. B., Talib, M. A., Nasir, Q., & Dakalbab, F. M. (2021). Machine learning for anomaly detection: A systematic review. Ieee Access, 9, 78658-78700. https://doi.org/10.1109/access.2021.3083060

[4] Alshammari, A., & Aldribi, A. (2021). Apply machine learning techniques to detect malicious network traffic in cloud computing. Journal of Big Data, 8(1), 90. https://doi.org/10.1186/s40537-021-00475-1

[5] Quiroz-Vázquez, C., IBM. 2025. Anomaly detection in machine learning: Finding outliers for optimization of business functions. Available online: https://www.ibm.com/think/topics/machine-learning-for-anomaly-detection. (accessed on 17 June 2025). (in English)

[6] Rana, S. (2019). Anomaly Detection in Network Traffic using Machine Learning and Deep Learning Techniques. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 10(2), 1063-1067. https://doi.org/10.17762/turcomat.v10i2.13626

[7] Almutairi, Y. S., Alhazmi, B., & Munshi, A. A. (2022). Network intrusion detection using machine learning techniques. Advances in Science and Technology Research Journal, 16(3), 193-206. https://doi.org/10.12913/22998624/149934.

[8] Hussain, F., Abbas, S. G., Shah, G. A., Pires, I. M., Fayyaz, U. U., Shahzad, F., ... & Zdravevski, E. (2021). A framework for malicious traffic detection in IoT healthcare environment. Sensors, 21(9), 3025. https://doi.org/10.3390/s21093025

[9] Iglesias, F., & Zseby, T. (2015). Analysis of network traffic features for anomaly detection. Machine Learning, 101, 59-84. https://doi.org/10.1007/s10994-014-5473-9.

[10] Disha, R. A., & Waheed, S. (2022). Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique. Cybersecurity, 5(1), 1. https://doi.org/10.1186/s42400-021-00103-8

[11] Fernandes, G., Rodrigues, J. J., Carvalho, L. F., Al-Muhtadi, J. F., & Proença, M. L. (2019). A comprehensive survey on network anomaly detection. Telecommunication Systems, 70, 447-489. https://doi.org/10.1007/s11235-018-0475-8

[12] Wang, Z., Fok, K. W., & Thing, V. L. (2022). Machine learning for encrypted malicious traffic detection: Approaches, datasets and comparative study. Computers & Security, 113, 102542. https://doi.org/10.1016/j.cose.2021.102542

[13] Radford, B. J., Apolonio, L. M., Trias, A. J., & Simpson, J. A. (2018). Network traffic anomaly detection using recurrent neural networks. Available online: https://arxiv.org/abs/1803.10769 (accessed on 5 July 2025). (in English)

[14] Fotiadou, K., Velivassaki, T. H., Voulkidis, A., Skias, D., Tsekeridou, S., & Zahariadis, T. (2021). Network traffic anomaly detection via deep learning. Information, 12(5), 215. https://doi.org/10.3390/info12050215

[15] Balyan, A. K., Ahuja, S., Lilhore, U. K., Sharma, S. K., Manoharan, P., Algarni, A. D., ... & Raahemifar, K. (2022). A hybrid intrusion detection model using ega-pso and improved random forest method. Sensors, 22(16), 5986. https://doi.org/10.3390/s22165986.

[16] Fu, C., Li, Q., Shen, M., & Xu, K. (2021, November). Realtime robust malicious traffic detection via frequency domain analysis. In Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (pp. 3431-3446). (in English) https://doi.org/10.1145/3460120.3484585.

[17] Koumar, J., Hynek, K., Pešek, J., & Čejka, T. (2024). NetTiSA: Extended IP flow with time-series features for universal bandwidth-constrained high-speed network traffic classification. Computer Networks, 240, 110147. (in English) https://doi.org/10.1016/j.comnet.2023.110147.

[18] Alduailij, M., Khan, Q. W., Tahir, M., Sardaraz, M., Alduailij, M., & Malik, F. (2022). Machine-learning-based DDoS attack detection using mutual information and random forest feature importance method. Symmetry, 14(6), 1095. (in English) https://doi.org/10.3390/sym14061095.

[19] Subbiah, S., Anbananthen, K. S. M., Thangaraj, S., Kannan, S., & Chelliah, D. (2022). Intrusion detection technique in wireless sensor network using grid search random forest with Boruta feature selection algorithm. Journal of Communications and Networks, 24(2), 264-273. https://doi.org/10.23919/JCN.2022.000002

[20] Al-amri, R., Murugesan, R. K., Man, M., Abdulateef, A. F., Al-Sharafi, M. A., & Alkahtani, A. A. (2021). A review of machine learning and deep learning techniques for anomaly detection in IoT data. Applied Sciences, 11(12), 5320. https://doi.org/10.3390/app11125320

[21] Stratosphere Laboratory. A labeled dataset with malicious and benign IoT network traffic. Available online: https://www.stratosphereips.org/datasets-iot23. (accessed on 20 December 2024). (in English)