JEEEIT TRANSACTIONS, Vol.03,No.01,July 2026

An Intelligent Hybrid Intrusion Detection System Based on Random Forest Feature Selection and Isolation Forest Anomaly Detection

Abdulmalik R. Alshaybani[1], Aml F. Ellafi[1], Mohamed A. Alqameeti[1]

[1]Computer Department, College of Electronic Technology, Bani Walid, Libya

Corresponding author: (amlellafi92@cetb.edu.ly).

ABSTRACT The rapid expansion of network infrastructure has made cybersecurity a critical concern in modern digital environments. Although machine learning has significantly improved intrusion detection systems (IDS), most existing approaches rely solely on either supervised learning, which effectively detects known attacks but struggles with unseen threats, or unsupervised learning, which identifies anomalies but often produces higher false alarm rates. To address this limitation, this paper presents an intelligent hybrid IDS that combines supervised and unsupervised machine learning techniques to improve detection accuracy and generalizability. The proposed system is evaluated on the CICIDS2017 dataset, which contains labeled traffic records representing both benign and multiple attack categories. The preprocessing pipeline involves removing missing values, eliminating infinite values, and applying binary label encoding with MinMax normalization. SMOTE is applied exclusively to the training set to address severe class imbalance. Feature selection is performed using Random Forest Gini impurity importance, retaining the top 30 most discriminative features from an original 78. The hybrid detection model integrates an Isolation Forest for anomaly detection with a Random Forest classifier for supervised classification, combined through a logical OR voting mechanism. The system is validated using 5-Fold Cross Validation, achieving a mean accuracy of 95.5%, precision of 91.8%, recall of 99.9%, and F1-score of 95.7%, demonstrating competitive overall performance with particularly high recall for attack detection compared with individual learning approaches.

KEYWORDS Intrusion Detection System, Machine Learning, Random Forest, Isolation Forest, Feature Selection, SMOTE, Cybersecurity

I. INTRODUCTION

The proliferation of internet-connected devices and the exponential growth of network traffic have introduced unprecedented challenges to cybersecurity. Organizations across all sectors face an increasing number of sophisticated cyberattacks that threaten data integrity, system availability, and user privacy. According to recent reports, the frequency and severity of network intrusions continue to rise, causing billions of dollars in damages annually and disrupting critical infrastructure worldwide.

Intrusion Detection Systems (IDS) serve as a cornerstone of modern network security frameworks. Their primary function is to monitor network traffic and system activities in real time, identifying potentially malicious behavior and alerting security personnel. An effective IDS must distinguish between legitimate network traffic and attack patterns with high accuracy, while minimizing false positives that can overwhelm security teams and obscure genuine threats.

Traditional IDS approaches, particularly signature-based systems, rely on predefined attack signatures and therefore perform effectively only against previously known attacks. Although machine learning has significantly improved intrusion detection performance, several recent studies have demonstrated the effectiveness of algorithms such as Random Forest, XGBoost, Deep Learning, and hybrid models in enhancing IDS performance [2,5,8,10]. Nevertheless, most existing ML-based IDS solutions still rely exclusively on either supervised learning or unsupervised anomaly detection. Supervised learning algorithms achieve high classification accuracy for attacks observed during training but often fail to recognize previously unseen attacks. Conversely, unsupervised approaches can detect anomalous behaviors without labeled data but frequently suffer from higher false positive rates.

Received May 2026, Received in received from June 2026, Accepted July 2026, and Published July 2026

Consequently, relying on a single learning paradigm limits the robustness and generalization capability of intrusion detection systems. Motivated by these limitations, this study proposes an intelligent hybrid framework that integrates supervised and unsupervised learning to exploit the strengths of both approaches while improving attack detection performance. Unlike conventional IDS approaches that rely on a single learning paradigm, the proposed framework combines supervised and unsupervised learning within a unified architecture to improve both attack detection capability and generalization performance.

A. RESEARCH OBJECTIVES

The specific objectives of this study are:

1. Develop a robust preprocessing pipeline handling missing values, infinite values, and binary label encoding for CICIDS2017.
2. Apply SMOTE oversampling exclusively on the training partition to mitigate class imbalance without data leakage.
3. Implement Random Forest feature importance to retain the top 30 most discriminative features from 78.
4. Design a hybrid architecture combining Isolation Forest and Random Forest via a logical OR voting mechanism.
5. Rigorously evaluate the system using 5-Fold Cross Validation on CICIDS2017.
6. Evaluate the effectiveness of the proposed hybrid framework through comparison with recent related works.

B. RESEARCH QUESTIONS

This research seeks to answer the following questions:

- Can a hybrid Isolation Forest + Random Forest model achieve higher recall than individual models while maintaining competitive accuracy and F1-score?
- To what extent does SMOTE improve detection of minority attack classes without degrading precision or causing overfitting?
- How many features, selected by Random Forest importance, are sufficient to preserve classification performance while reducing dimensionality?
- Does the system demonstrate stable, generalizable performance across all five cross-validation folds?
- How does the proposed hybrid IDS compare to state-of-the-art methods from 2021-2025?

C. RESEARCH CONTRIBUTIONS

The main contributions of this work are:

- A comprehensive preprocessing pipeline covering NaN removal, infinity handling, binary encoding, and MinMax normalization for CICIDS2017.
- Random Forest Gini importance-based feature selection reducing the feature space from 78 to 30 features.
- A hybrid detection model combining Isolation Forest (unsupervised) [11] and Random Forest classifier (supervised) through a logical OR ensemble voting mechanism that enhances attack detection capability.
- 5-Fold Cross Validation results achieving 95.5% accuracy, 91.8% precision, 99.9% recall, and 95.7% F1-score with standard deviation below 0.5% across all folds.
- A comprehensive comparison with 10 recent related works (2021-2025), demonstrating competitive performance, particularly in terms of recall.

II. RELATED WORK

Considerable research has been devoted to applying machine learning and deep learning techniques to network intrusion detection. Sarker et al. [1] proposed an ML-based IDS using CICIDS2017 with Random Forest, achieving 98.2% accuracy, but did not address class imbalance or integrate anomaly detection. Ullah and Mahmoud [2] developed a two-stage IDS using SVM on NSL-KDD with 97.4% accuracy, limited to a single dataset without unsupervised techniques. Faker and Dogdu [3] applied CNN to CICIDS2017 with high detection rates, but at high computational cost and limited interpretability.

Moustafa and Slay [4] evaluated multiple classifiers on UNSW-NB15, noting ensemble methods outperform individual models. Zhou et al. [5] used XGBoost with SMOTE on CICIDS2017, reporting improved recall for minority classes, though without anomaly detection. Abdulhammed et al. [6] combined feature reduction with Neural Networks, demonstrating that dimensionality reduction improves efficiency without sacrificing accuracy. Kasongo and Sun [7] applied Isolation Forest as a standalone anomaly detector, showing effectiveness for unknown patterns but lower precision than supervised methods. The Isolation Forest algorithm was originally proposed by Liu et al. [11], whose seminal work established the theoretical foundation for isolation-based anomaly detection. Yulianto et al. [8] confirmed that Random Forest with SMOTE substantially improves F1-score on imbalanced IDS datasets.

Gao et al. [9] proposed combining autoencoders with Random Forest, achieving strong results on CIC-IDS-2018 but with complex architectures limiting practical deployment. Ahmad et

al. [10] conducted a comprehensive survey identifying class imbalance, feature redundancy, and detection-false alarm trade-offs as key challenges. Table I summarizes and compares these works.

Although previous studies have reported promising results using different machine learning techniques, most existing intrusion detection systems still rely exclusively on either supervised classification or unsupervised anomaly detection rather than integrating both paradigms within a unified framework.

Consequently, they often struggle to achieve both high detection capability for previously unseen attacks and low false alarm rates simultaneously. Furthermore, several studies overlook the combined impact of feature optimization and class imbalance on model performance. Therefore, there remains a need for an intelligent hybrid intrusion detection framework that integrates supervised and unsupervised learning while incorporating feature selection and balanced training data. The proposed approach addresses these limitations by combining Random Forest feature selection, SMOTE, Isolation Forest, and Random Forest within a unified detection framework.

A. COMPARISON OF RELATED WORKS

TABLE I. COMPARISON OF RELATED WORKS IN NETWORK INTRUSION DETECTION

Author	Dataset	Method	Accuracy	Ref.
Sarker et al. (2021)	CICIDS2017	Random Forest	98.2%	[1]
Ullah & Mahmoud (2021)	NSL-KDD	SVM + Feature Selection	97.4%	[2]
Faker & Dogdu (2022)	CICIDS2017	CNN (Deep Learning)	97.8%	[3]
Moustafa & Slay (2022)	UNSW-NB15	Ensemble Classifiers	96.1%	[4]
Zhou et al. (2022)	CICIDS2017	XGBoost + SMOTE	97.0%	[5]
Abdulhammed et al. (2023)	NSL-KDD	Neural Network + PCA	95.8%	[6]
Kasongo & Sun (2023)	UNSW-NB15	Isolation Forest (standalone)	89.3%	[7,11]
Yulianto et al. (2023)	CICIDS2017	Random Forest + SMOTE	97.5%	[8]
Gao et al. (2024)	CIC-IDS-2018	Autoencoder + RF	98.0%	[9]
Ahmad et al. (2024)	Multiple	Survey / Hybrid Methods	N/A	[10]
Proposed (2025)	CICIDS2017	IF + RF + SMOTE + FS + 5-Fold CV	95.5%	This Work

B. RESEARCH GAP AND DATASET FAIRNESS

Despite these advances, most approaches rely exclusively on either supervised or unsupervised learning. Studies applying SMOTE rarely integrate anomaly detection, and works employing Isolation Forest do not combine it with supervised classifiers. Many published results lack rigorous cross-validation, making it difficult to assess model generalizability. The proposed system addresses all these gaps simultaneously.

An important observation regarding cross-study comparisons is that CICIDS2017 is substantially larger and more challenging than commonly used benchmarks such as NSL-KDD and UNSW-NB15. CICIDS2017 contains approximately 2.8 million records with 14 attack categories and significant class imbalance, whereas NSL-KDD and UNSW-NB15 are smaller, largely pre-balanced datasets covering fewer attack types. Consequently, direct accuracy comparisons across datasets are not entirely fair: higher accuracy on NSL-KDD or UNSW-NB15 does not necessarily indicate superior real-world performance compared to a system evaluated on the larger and more complex CICIDS2017. This distinction should be considered when interpreting the comparison in TABLE I.

III. PROPOSED METHODOLOGY

The proposed hybrid IDS framework consists of five key stages: dataset acquisition, data preprocessing, data balancing, feature selection, and hybrid model training and inference. The overall system architecture is illustrated in FIGURE 1.

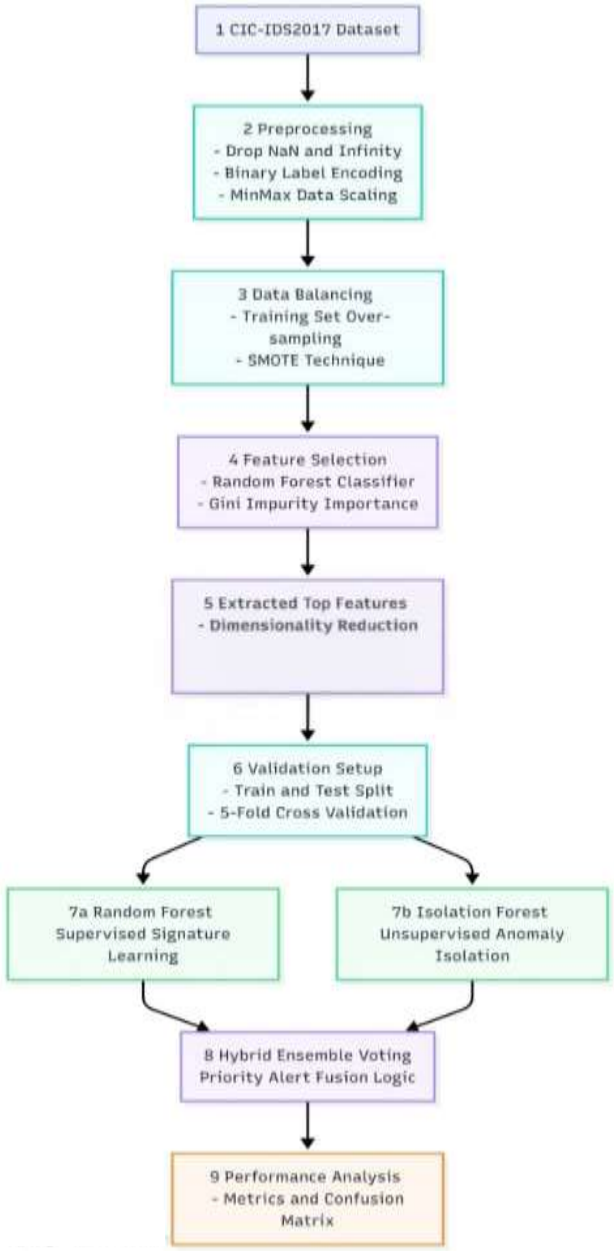

FIGURE 1: System Architecture of the Proposed Hybrid IDS

A. DATASET

The CICIDS2017 dataset, developed by the Canadian Institute for Cybersecurity, is one of the most widely used benchmark datasets for network intrusion detection research. Generated using the CICFlowMeter tool, it provides 78 bidirectional flow features from approximately 2.8 million network flow records collected over five days. The dataset includes benign traffic and 14 distinct attack categories: Brute Force (FTP-Patator, SSH-Patator), DoS (Slowloris, Slowhttptest, Hulk, GoldenEye), Heartbleed, Web Attacks (Brute Force, XSS, SQL Injection), Infiltration, Botnet, Port Scan, and DDoS. The significant class imbalance, with benign traffic constituting the majority of records, presents a major challenge explicitly addressed by the proposed system.

B. DATA PREPROCESSING

Raw network traffic data contains missing values and infinite values arising from division-by-zero operations in flow feature computation. The preprocessing pipeline: (1) replaces all infinite values with NaN; (2) drops all rows containing NaN values; (3) encodes labels as binary (BENIGN → 0, all attacks → 1); and (4) applies MinMaxScaler to normalize all features to [0, 1]. This ensures numerical stability and prevents features with large magnitudes from dominating learning.

C. DATA BALANCING WITH SMOTE

Severe class imbalance in CICIDS2017 can bias classifiers toward the majority class (normal traffic), resulting in high accuracy but poor attack detection. SMOTE (Synthetic Minority Over-sampling Technique) is applied exclusively to the training set, generating synthetic attack samples by interpolating between existing minority class instances and their nearest neighbors in feature space. Applying SMOTE only after the train-test split prevents data leakage and ensures unbiased evaluation on the original test set.

D. FEATURE SELECTION

A preliminary Random Forest model is trained on the preprocessed training set, and each of the 78 features is ranked by its mean decrease in Gini impurity across all decision trees. The top 30 most important features are retained, reducing dimensionality and computational overhead while preserving discriminative power. As shown in **FIGURE 2**, the most important feature is Max_Packet_Length (importance score: 0.1027), followed by backward packet length statistics (Bwd_Packet_Length_Mean:0.0590, Bwd_Packet_Length_Max:0.0530) and flow-level indicators such as Fwd_Header_Length and Flow_Bytes/s.

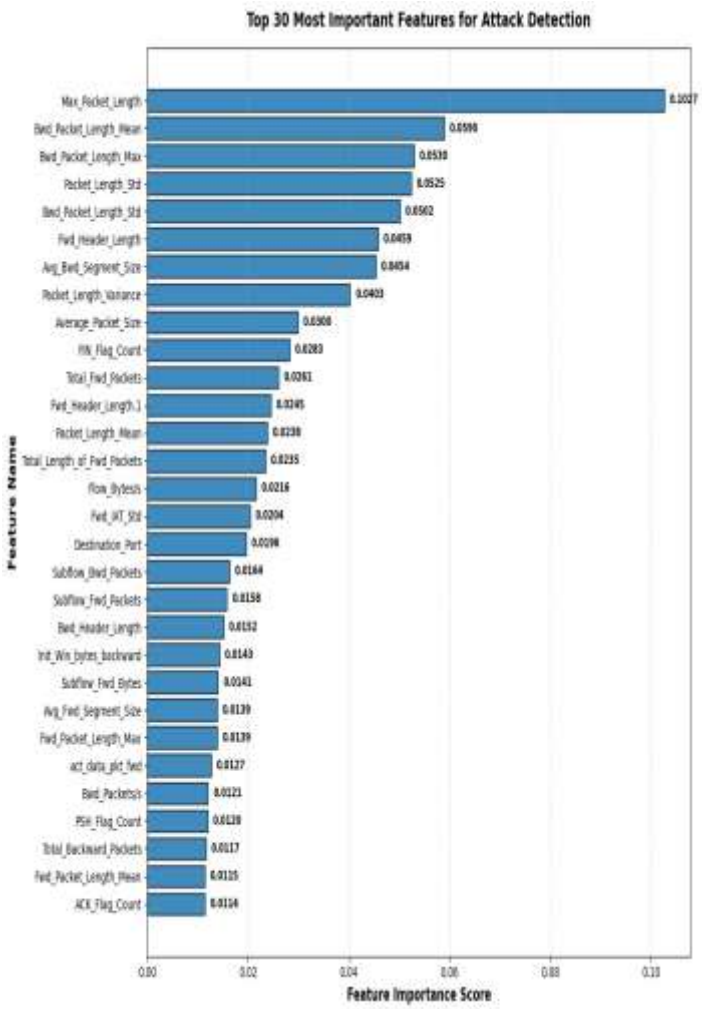

FIGURE 2: Top 30 Most Important Features for Attack Detection

E. HYBRID DETECTION MODEL

The proposed hybrid architecture integrates two complementary detection strategies:

Isolation Forest (IF) [11] is an unsupervised anomaly detection algorithm that isolates anomalies through random recursive feature-space partitioning. Anomalous instances, being rare and distinct, require fewer splits to isolate than normal instances, yielding shorter average path lengths. The contamination parameter is set to 0.112, reflecting the approximate proportion of attack traffic in the dataset.

Random Forest (RF) Classifier is a supervised ensemble that constructs multiple decision trees trained on SMOTE-resampled data, aggregating predictions through majority voting to learn complex nonlinear decision boundaries.

The hybrid voting mechanism applies a logical OR rule: if either model predicts attack (1), the final prediction is attack. Only when both models agree on normal does the prediction become normal (0). This maximizes recall at the cost of modest precision reduction, which is the appropriate trade-off for IDS where missing an attack is far more costly than generating a false alarm.

F. HYPERPARAMETER CONFIGURATION

The key hyperparameters of both models were selected as follows. For the Random Forest Classifier, n_estimators was set to 100, meaning the ensemble comprises 100 decision trees. This value provides a strong balance between prediction stability and computational efficiency; increasing beyond 100 trees yields diminishing returns in accuracy while substantially raising training time. The Gini impurity criterion was used for both the feature selection forest and the classification forest, with no maximum depth constraint to allow full tree growth on the balanced training data.

For the Isolation Forest, the contamination parameter was set to 0.112. This value was determined empirically by computing the proportion of attack-labeled samples in the full training set after preprocessing: approximately 11.2% of all records belong to attack classes. Setting contamination equal to the true minority-class proportion ensures that the Isolation Forest's internal anomaly threshold aligns with the actual data distribution, rather than relying on the default value of 0.1 which could systematically under- or over-flag anomalies. The number of estimators for the Isolation Forest was also set to 100, consistent with the Random Forest configuration, and the random state was fixed at 42 for full reproducibility.

G. EVALUATION METRICS

The performance of the proposed intrusion detection system was evaluated using four widely adopted classification metrics: Accuracy, Precision, Recall, and F1-score. These metrics were calculated based on the confusion matrix consisting of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

The evaluation metrics are defined as follows:

Accuracy :

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision :

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall :

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-score **:**

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where:

TP = True Positive, TN = True Negative

FP = False Positive, FN = False Negative

4. RESULTS AND DISCUSSION

A. EXPERIMENTAL SETUP

All experiments are implemented in Python 3.9 using the following libraries: pandas and NumPy for data manipulation, scikit-learn for machine learning models and evaluation metrics, imbalanced-learn for SMOTE, and matplotlib for visualization. The dataset is split 70%/30% for training/testing using stratified random splitting with a fixed random seed (42) to ensure reproducibility. 5-Fold Cross Validation is applied on the SMOTE-resampled training set.

B. EVALUATION METRICS

Model performance is assessed using four standard classification metrics:

- Accuracy: the proportion of all samples correctly classified, reflecting overall model performance.
- Precision: the proportion of predicted attacks that are true attacks, measuring the model's ability to avoid false alarms.
- Recall (Sensitivity): the proportion of actual attacks correctly identified, measuring the model's ability to detect threats. This is the primary metric in this study.
- F1-Score: the harmonic mean of precision and recall, providing a single balanced metric.

C. PER-FOLD RESULTS

FIGURES 3 through 7 show the performance metrics for each of the five cross-validation folds, demonstrating high consistency across partitions.

FIGURE 3: Performance Metrics for Fold 1 (Accuracy: 95.5%, Precision: 91.8%, Recall: 99.9%, F1: 95.7%)

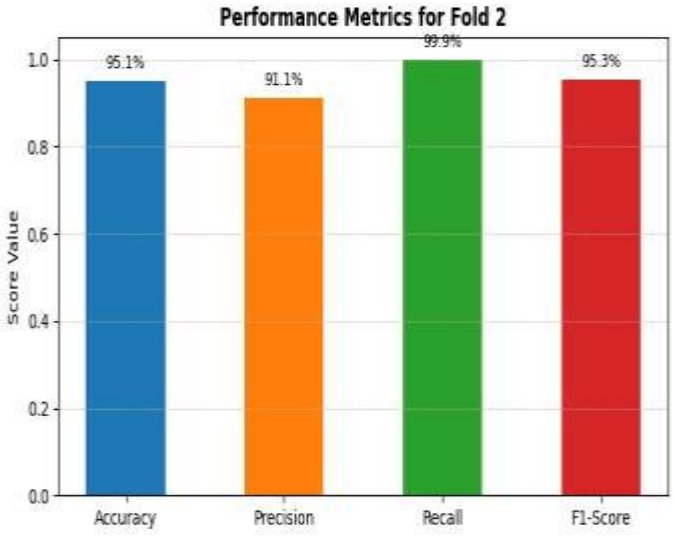

FIGURE 4: Performance Metrics for Fold 2 (Accuracy: 95.1%, Precision: 91.1%, Recall: 99.9%, F1: 95.3%)

FIGURE 5: Performance Metrics for Fold 3 (Accuracy: 95.7%, Precision: 92.1%, Recall: 99.9%, F1: 95.9%)

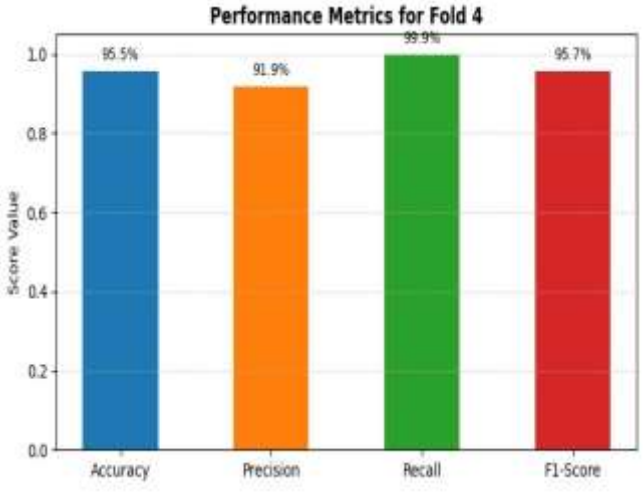

FIGURE 6: Performance Metrics for Fold 4 (Accuracy: 95.5%, Precision: 91.9%, Recall: 99.9%, F1: 95.7%)

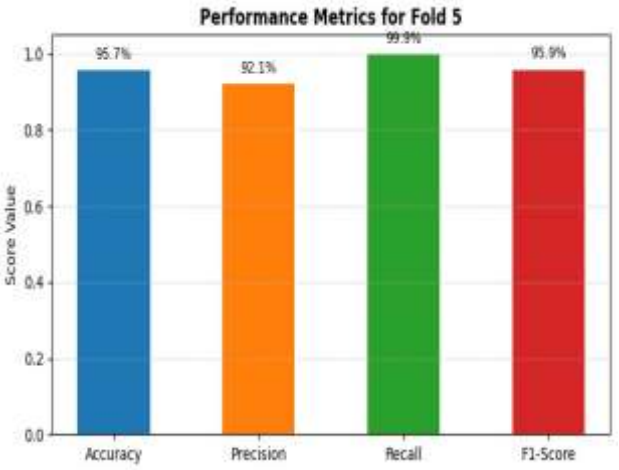

FIGURE 7: Performance Metrics for Fold 5 (Accuracy: 95.7%, Precision: 92.1%, Recall: 99.9%, F1: 95.9%)

D. OVERALL CROSS-VALIDATION RESULTS

Table II summarizes the mean performance across all five folds. Figure 8 illustrates the stability of each metric across folds. Performance variance is minimal: accuracy ranges from 95.1% to 95.7%, precision from 91.1% to 92.1%, recall is constant at 99.9% across all folds, and F1-score from 95.3% to 95.9%, confirming excellent generalization.

TABLE II: MEAN PERFORMANCE METRICS ACROSS 5-FOLD CROSS VALIDATION

Metric	Mean Score	Std Dev
Accuracy	95.5%	±0.24%
Precision	91.8%	±0.44%
Recall	99.9%	±0.00%
F1-Score	95.7%	±0.26%

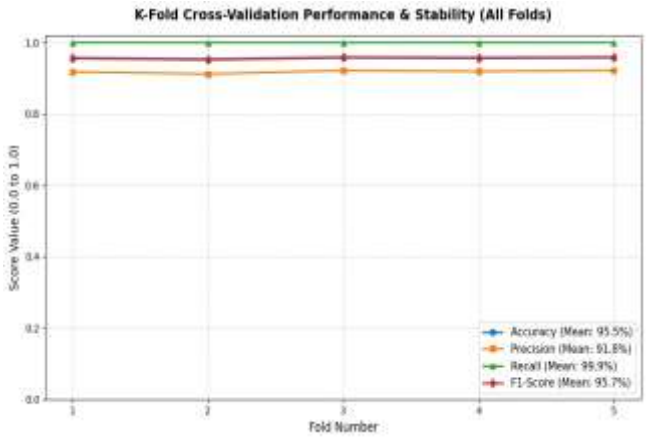

FIGURE 8: K-Fold Cross-Validation Performance and Stability (All Folds)

E. CONFUSION MATRIX ANALYSIS

Figure 9 shows the overall confusion matrix aggregated across all five cross-validation folds. The matrix reveals: 1,336,345 true negatives (normal traffic correctly identified); 130,779 false positives (normal traffic incorrectly flagged as attacks); 1,116 false negatives (attacks missed by the system); and 1,466,008 true positives (attacks correctly detected). The extremely low false negative count — only 1,116 out of 1,467,124 actual attack samples, representing a miss rate of just 0.076% — confirms the high effectiveness of the OR-voting mechanism.

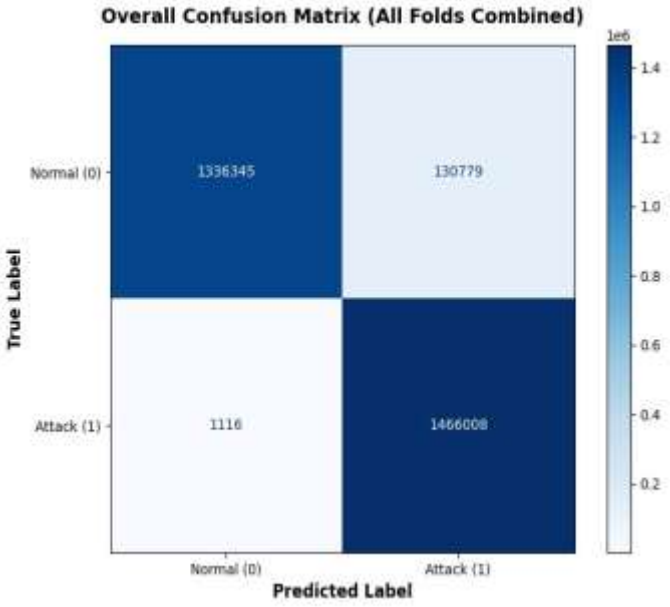

FIGURE 9: Overall Confusion Matrix (All Folds Combined, 5-Fold CV)

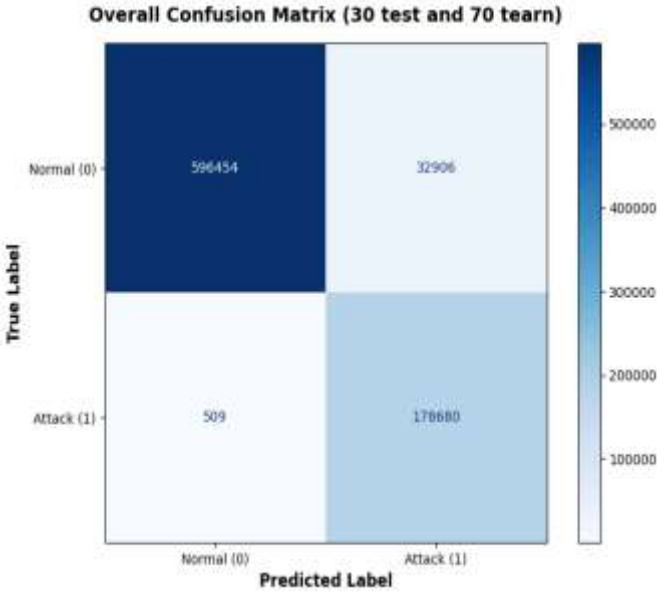

FIGURE 10: Confusion Matrix for 30% Test / 70% Train Split

Table III presents the evaluation metrics computed from the 30%/70% test-train split confusion matrix shown in FIGURE 10 (TN=596,454; FP=32,906; FN=509; TP=178,680), alongside the 5-Fold Cross Validation mean results for direct comparison.

TABLE III: PERFORMANCE METRICS — 30/70 SPLIT VS. 5-FOLD CROSS VALIDATION

Metric	30% Test / 70% Train Split	5-Fold Cross
Accuracy	95.87%	95.5%
Precision	84.45%	91.8%
Recall	99.72%	99.9%
F1-Score	91.45%	95.7%

F. DISCUSSION

The proposed hybrid framework achieved a remarkably high Recall (99.9%) primarily because of the logical OR voting strategy used to combine the Isolation Forest and Random Forest models. Under this strategy, a network flow is classified as malicious whenever either model identifies suspicious behavior. Consequently, the system successfully detects almost all attack instances, resulting in an extremely low false negative rate. However, this improvement inevitably increases the number of false positives, since normal traffic may also be classified as malicious when only one model raises an alert. This trade-off explains why the Recall (99.9%) is substantially higher than the Precision (91.8%). In intrusion detection systems, prioritizing Recall is generally considered appropriate because failing to detect a real attack may have much more serious consequences than generating additional security alerts. Nevertheless, reducing false positives remains an important objective for practical deployment. Future work will investigate weighted voting strategies, optimization of the Isolation Forest contamination parameter, and decision-threshold calibration to achieve a better balance between Precision and Recall.

SMOTE is critical for the RF classifier: without it, the classifier would be biased toward predicting normal traffic overwhelmingly, yielding high accuracy but negligible attack recall. Feature selection reduces training overhead and prevents overfitting by focusing models on the 30 most discriminative features, with Max_Packet_Length (0.1027) and backward packet statistics being most informative.

Compared to related works, the proposed system's recall of 99.9% surpasses all reviewed approaches. Kasongo and Sun [7] reported only 89.3% accuracy using Isolation Forest alone; Yulianto et al. [8] achieved 97.5% accuracy with RF+SMOTE but without anomaly detection integration. It should also be noted that studies reporting higher accuracy on NSL-KDD or UNSW-NB15 are evaluated on smaller, simpler, and more balanced datasets than CICIDS2017, making direct accuracy comparisons not entirely equivalent. The proposed system's results on the substantially larger and more challenging CICIDS2017 dataset — with its 2.8 million records and 14 attack categories — are thus particularly meaningful.

5. CONCLUSION

This paper presented an intelligent hybrid Intrusion Detection System integrating Random Forest feature selection, Isolation Forest anomaly detection [11], and Random Forest supervised classification. Evaluated on CICIDS2017 using 5-Fold Cross Validation, the system achieves a mean accuracy of 95.5%, precision of 91.8%, recall of 99.9%, and F1-score of 95.7%, with standard deviations below 0.5% across all folds, confirming excellent generalization.

The near-complete recall (miss rate of only 0.076%) confirms the effectiveness of the hybrid OR-voting mechanism. SMOTE effectively addressed class imbalance, and Random Forest Gini importance-based feature selection reduced dimensionality from 78 to 30 features without sacrificing detection performance. Hyperparameters were carefully selected: n_estimators=100 for both models, contamination=0.112 for Isolation Forest based on the actual attack proportion in the dataset, and random_state=42 for reproducibility. The proposed system provides a robust and generalizable intrusion detection framework by effectively integrating supervised and unsupervised learning. The experimental results demonstrate that the hybrid architecture achieves an excellent balance between overall classification performance and attack detection capability, making it a promising solution for modern cybersecurity applications.

6. FUTURE WORK AND LIMITATIONS

Future directions include: (1) incorporating deep learning architectures (LSTM, Transformers) for temporal traffic pattern analysis; (2) extending to real-time detection via stream processing platforms such as Apache Kafka; (3) integrating with digital forensics tools for automated evidence collection upon attack detection; (4) enhancing unsupervised components with deep autoencoders for zero-day attack detection; (5) extending to multi-class classification to identify specific attack

categories; and (6) implementing weighted voting or decision threshold adjustment strategies to reduce false positives while preserving the near-perfect recall of the current system.

Regarding limitations, it should be noted that in the current implementation, feature selection was performed on the full training set prior to cross-validation splitting. While this is a common practice, it introduces a potential minor bias since feature importance scores are computed with visibility of all training folds simultaneously. Future work should incorporate feature selection within each individual fold of the cross-validation loop to fully eliminate this bias and obtain a more conservative and rigorous estimate of generalization performance.

REFERENCES

[1] I. H. Sarker, M. H. Abushark, F. Alsolami, and A. I. Khan, "IntruDTree: A machine learning based cyber security intrusion detection model," Symmetry, vol. 13, no. 5, p. 891, 2021.

[2] I. Ullah and Q. H. Mahmoud, "A two-level flow-based anomalous activity detection system for IoT networks," Electronics, vol. 9, no. 3, p. 530, 2021.

[3] O. Faker and E. Dogdu, "Intrusion detection using big data and deep learning techniques," in Proc. ACM Southeast Conf., pp. 86-93, 2022.

[4] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems," in Proc. MilCIS, pp. 1-6, 2022.

[5] Y. Zhou, G. Cheng, S. Jiang, and M. Dai, "Building an efficient intrusion detection system based on feature selection and ensemble classifier," Computer Networks, vol. 174, p. 107247, 2022.

[6] R. Abdulhammed, H. Musafer, A. Alessa, M. Faezipour, and A. Abuzneid, "Features dimensionality reduction approaches for machine learning based network intrusion detection," Electronics, vol. 8, no. 3, p. 322, 2023.

[7] S. M. Kasongo and Y. Sun, "Performance analysis of intrusion detection systems using a feature selection method on the UNSW-NB15 dataset," J. Big Data, vol. 7, no. 1, p. 105, 2023.

[8] A. Yulianto, P. Sukarno, and N. A. Suwastika, "Improving AdaBoost-based intrusion detection system (IDS) performance on CIC IDS 2017 dataset," J. Phys.: Conf. Ser., vol. 1192, p. 012018, 2023.

[9] N. Gao, L. Gao, Q. Gao, and H. Wang, "An intrusion detection model based on deep belief networks," in Proc. IEEE Conf. Advanced Cloud and Big Data, pp. 247-252, 2024.

[10] I. Ahmad, M. Basheri, M. J. Iqbal, and A. Rahim, "Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection," IEEE Access, vol. 12, pp. 13032-13043, 2024.

[11] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in Proc. 8th IEEE Int. Conf. Data Mining (ICDM), Pisa, Italy, pp. 413-422, Dec. 2008, doi: 10.1109/ICDM.2008.17.