Artificial Neural Network–Random Forest Hybrid Model for Network Traffic Classification

Rajasvaran Logeswaran[1] and Mukil Alagirisamy[2]

[1] *Faculty of Computing and Digital Technology, HELP University, Kuala Lumpur, Malaysia*
[2] *School of Engineering, Asia Pacific University of Technology & Innovation, Kuala Lumpur, Malaysia*

*Corresponding Author: mukil.alagirismy@apu.edu.my

Received: 12 December 2025 Revised: 22 January 2026 Accepted: 16 February 2026

Abstract— This article proposes a novel hybrid machine learning framework that integrates Artificial Neural Networks (ANN) with Random Forest (RF) to enhance network traffic classification performance. By leveraging the complementary strengths of both techniques, the proposed model improves robustness and accuracy in distinguishing malicious traffic from benign activity. Building upon prior studies conducted on the NSL-KDD dataset, this research further validates the model using additional benchmark datasets, including CICIDS2017 and UNSW-NB15, to examine its effectiveness across diverse and realistic network environments. The hybrid architecture exploits the powerful feature learning capability of ANN alongside the strong ensemble-based decision-making capacity of Random Forest, resulting in superior detection performance compared to conventional standalone models. The proposed approach is evaluated using standard performance metrics such as ROC-AUC, Precision, Recall, and F1-score. Experimental results demonstrate that the hybrid model consistently outperforms existing methods, providing a more accurate, reliable, and scalable solution for network intrusion detection systems.

Keywords— Artificial Intelligence, Direction of Arrival (DOA) Estimation, Sparse Sensor Arrays, Hybrid Machine Learning, Optimization Techniques, Robust BeamfoArtificial Neural Network (ANN), Random Forest, Network Traffic Classification, Intrusion Detection System (IDS), Hybrid Machine Learningrming.

1. Introduction

In the rapidly evolving landscape of network security, accurate and robust traffic classification is crucial for detecting and mitigating cyber threats. Traditional machine learning approaches have made significant strides, yet challenges remain in achieving high precision and adaptability across diverse network environments. To address these challenges, this study introduces a novel hybrid machine learning model that combines the strengths of Artificial Neural Networks (ANN) and Random Forests. By integrating ANN's advanced feature extraction capabilities with the robust classification power of Random Forests, this hybrid approach aims to enhance network traffic classification, distinguishing effectively between malicious and benign activities. The proposed model is evaluated using a range of datasets, including NSL-KDD, CICIDS2017, and UNSW-NB15, to ensure its efficacy in varied conditions. This innovative combination not only improves classification accuracy but also provides a more resilient solution to evolving cyber threats, marking a significant advancement in network intrusion detection systems.

2. Related Work

Various machine learning algorithms have led to considerable breakthroughs in intrusion detection and network traffic classification. Artificial Neural Networks (ANNs), which are excellent at identifying intricate patterns in network traffic data, are one noteworthy method. As an example, Fatani et al. [11] improved intrusion detection in Internet of Things (IoT) environments by utilizing deep learning models. Their research showed how ANNs may be used to efficiently handle complex network patterns. Apart from artificial neural networks, Random Forests have also been acknowledged for their efficacy in diverse categorization assignments. To increase intrusion detection system accuracy, Ustebay et al. [2] combined Random Forests with recursive feature reduction. This method demonstrates how well Random Forests handle big feature sets and improve classification accuracy.

Using a hybrid model that combines several classifiers has also shown benefits. Bagaa et al. [13] investigated a hybrid strategy that combined Random Forests with additional classifiers to create a strong security framework for Internet of Things devices. Their research demonstrates the benefits of combining several models to get better results. Kumar et al. [3] have evaluated the combination of ANNs with Random Forests in the context of network anomaly detection. Their review demonstrated the advantages of mixing several machine learning techniques, emphasizing the usefulness of hybrid models in enhancing detection capabilities. The improvement of network intrusion detection systems using sophisticated methodologies has also been the subject of further research. For instance, to safeguard IoT networks, Moustafa

et al. [14] presented an ensemble intrusion detection technique utilizing statistical flow features.

Moreover, innovative attack detection techniques utilizing lightweight models are recent breakthroughs. In order to detect assaults in industrial IoT scenarios, Latif et al. [12] developed a lightweight random neural network, showcasing improvements in lowering computational overhead without sacrificing effective detection performance. Using datasets like UNSW-NB15 and KDD99, Moustafa and Slay [8] offered a thorough assessment of network anomaly detection algorithms. Their analysis emphasizes how crucial diverse and high-quality datasets are to the creation of reliable detection systems. A comprehensive assessment of intrusion detection systems in the Internet of Things was carried out by Khraisat and Alazab [4], who concentrated on deployment tactics, validation techniques, and difficulties. Their research emphasizes how IoT security is changing and how important it is to have efficient intrusion detection systems.

In order to solve problems unique to IoT environments, Hasan et al. [5] investigated attack and anomaly detection in IoT sensors using machine learning techniques. Their research advances the continuous creation of customized detection methods for various network situations. In conclusion, using sophisticated machine learning methods—such as Random Forests and Artificial Neural Networks—in hybrid models shows promise for enhancing intrusion detection and network traffic classification. Combining these methods makes use of each one's advantages to improve overall performance and offer a more precise and dependable network security solution. Combining these methods makes use of each one's advantages to improve overall performance and offer a more precise and dependable network security solution.

3. Proposed Method: H-ANN-RF Model for Network Traffic Classification

3.1 Feature Extraction with ANN

ANN Architecture: Let $\mathbf{X} \in \mathbb{R}^d$ denote the input feature vector, where d is the number of features. The ANN consists of multiple layers: input layer $\mathbf{x}$, hidden layers $\mathbf{h}^l$ (for $l = 1, 2, \ldots, L$), and output layer $\mathbf{y}$.

Forward Propagation: The forward propagation through a single hidden layer can be described as:

$$\mathbf{h}^l = \sigma(W^l \mathbf{h}^{l-1} + \mathbf{b}^l) \tag{1}$$

where σ is the activation function (e.g., ReLU, sigmoid), W^l is the weight matrix, and $\mathbf{b}^l$ is the bias vector.

For the output layer, the computation is:

$$\mathbf{y} = \sigma(W^L \mathbf{h}^L + \mathbf{b}^L) \tag{2}$$

where W^L and $\mathbf{b}^L$ are the weights and biases for the output layer.

Feature Extraction: The feature extraction process involves using the activations of the final hidden layer $\mathbf{h}^L$ as the feature representation. Thus, the extracted feature vector $\mathbf{F}_{ANN}$ is given by:

$$\mathbf{F}_{ANN} = \mathbf{h}^L \tag{3}$$

3.2 Classification with Random Forests

Feature Representation: The features extracted by the ANN $\mathbf{F}_{ANN} \in \mathbb{R}^k$ are then used as input for the Random Forest classifier.

Random Forest Model: A Random Forest consists of N decision trees, where each tree T_i (for $i = 1, 2, \ldots, N$) is trained on a bootstrap sample of the data.

For a given feature vector $\mathbf{F}_{ANN}$, each tree T_i makes a prediction $\hat{y}_i$, where:

$$\hat{y}_i = T_i(\mathbf{F}_{ANN}) \tag{4}$$

Aggregation: The final prediction $\hat{y}$ of the Random Forest is obtained by aggregating the predictions of all trees. For classification, this is typically done using majority voting:

$$\hat{y} = \text{mode}(\{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_N\}) \tag{5}$$

3.3 Hybrid Model Integration

The Hybrid ANN-Random Forest (H-ANN-RF) Model combines the strengths of Artificial Neural Networks (ANNs) and Random Forests to enhance network traffic classification. This integration is designed to capitalize on the complementary advantages of these two machine learning techniques, providing a robust solution for distinguishing between malicious and benign network traffic. Below, we elaborate on the integration process and its significance:

3.4 Hybrid Approach

The H-ANN-RF Model integrates ANN and Random Forests in a two-stage process:

- **Feature Extraction with ANN:** The ANN processes the raw network traffic data $\mathbf{X}$ to extract meaningful features. The ANN, with its multiple hidden layers and non-linear activation functions, is adept at capturing complex patterns and interactions in the data. The feature extraction process involves:

  $$\mathbf{h}^l = \sigma(W^l \mathbf{h}^{l-1} + \mathbf{b}^l) \tag{6}$$

  for hidden layers, and

  $$\mathbf{y} = \sigma(W^L \mathbf{h}^L + \mathbf{b}^L) \tag{7}$$

  for the output layer. The final hidden layer activations $\mathbf{h}^L$ are used as the feature representation:

  $$\mathbf{F}_{ANN} = \mathbf{h}^L \tag{8}$$

  This process ensures that the features extracted are rich and representative of the underlying data structure.

- **Classification with Random Forests:** The extracted features $\mathbf{F}_{ANN}$ are then fed into the Random Forest classifier. Random Forests, comprising an ensemble of decision trees, are particularly effective at handling

large feature sets and managing overfitting. Each decision tree T_i provides a prediction $\hat{y}_i$, and the final prediction $\hat{y}$ is obtained by aggregating the outputs of all trees:

$$\hat{y}_i = T_i(\mathbf{F}_{ANN}) \tag{9}$$

$$\hat{y} = \text{mode}(\{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_N\}) \tag{10}$$

This ensemble approach enhances the classification robustness by reducing variance and improving generalization.

The Hybrid ANN-Random Forest Model integrates the strengths of both ANN and Random Forests to provide a more accurate and reliable network traffic classification system. The ANN's advanced feature extraction and the Random Forest's robust classification together enhance the model's ability to detect and differentiate between malicious and benign network activities, making it a powerful tool for network security applications.

4. Algorithm: Hybrid ANN-Random Forest Model

Algorithm 1 Hybrid ANN-Random Forest Model for Network Traffic Classification

Network traffic data $\mathbf{X}$, ANN parameters, Random Forest parameters Class label $\hat{y}$

Step 1: Feature Extraction with ANN each sample $\mathbf{x} \in \mathbf{X}$ **Forward Propagation:** Compute hidden layer activations:

$$\mathbf{h}^l = \sigma(W^l \mathbf{h}^{l-1} + \mathbf{b}^l) \tag{11}$$

Compute output layer activations:

$$\mathbf{y} = \sigma(W^L \mathbf{h}^L + \mathbf{b}^L) \tag{12}$$

Extract Features:

$$\mathbf{F}_{ANN} = \mathbf{h}^L \tag{13}$$

Step 2: Classification with Random Forests each feature vector $\mathbf{F}_{ANN}$ **Prediction:** each tree T_i in Random Forest

$$\hat{y}_i = T_i(\mathbf{F}_{ANN}) \tag{14}$$

Aggregate Predictions:

$$\hat{y} = \text{mode}(\{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_N\}) \tag{15}$$

Class label $\hat{y}$

5. Results and Discussion

5.1 Performance Evaluation on CICIDS2017 Dataset

The Hybrid ANN-Random Forest (H-ANN-RF) model was rigorously evaluated using the CICIDS2017 dataset, which is widely recognized for its comprehensive representation of various types of network attacks and benign traffic. The dataset was subjected to extensive preprocessing, including normalization and feature selection, ensuring that the inputs provided to the model were of the highest quality.

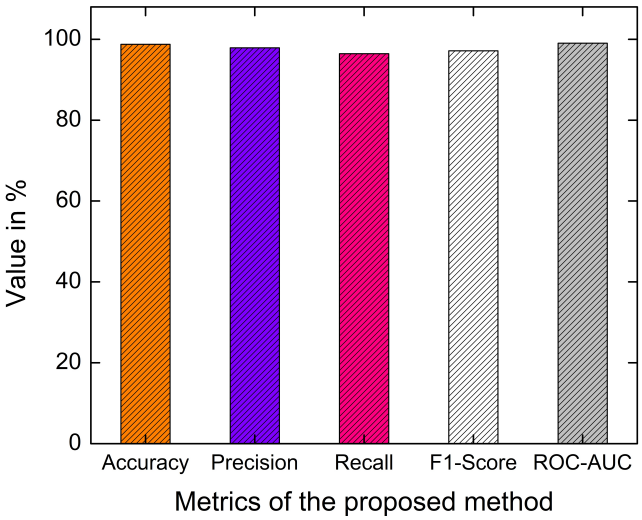

Figure 1: Illustration of the Performance Metrics of the H-ANN-RF Model on the CICIDS2017 Dataset from Table 1.

5.2 Evaluation Metrics

The effectiveness of the H-ANN-RF model was assessed using standard performance metrics, namely:

1. **Accuracy:** The percentage of accurately classified examples out of all the instances in total.
2. **Precision:** The percentage of actual positive cases among all cases that were anticipated to be positive.
3. **Recall (Sensitivity):** The percentage of real positive examples among all positive examples that actually occur. **F1-Score:**
4. Precision and Recall's harmonic mean, which balances their respective trade-offs.
5. **Area Under the Curve - Receiver Operating Characteristic - ROC-AUC:** a gauge of the model's capacity for class discrimination.

5.3 Experimental Results

The CICIDS2017 dataset was used to train and test the model, with training and testing sets divided 70:30. The evaluation's findings, which are presented in Table 1, show how well the H-ANN-RF model performed on all important criteria. The visual representation of the H-ANN-RF model's performance metrics on the CICIDS2017 dataset is provided in Figure 1. In addition to the specific numerical findings shown in Table 1, this image offers a clear depiction of the model's performance across important measures, such as Accuracy, Precision, Recall, F1-Score, and ROC-AUC.

5.4 Comparative Analysis

To contextualize the performance of the H-ANN-RF model, a comparative analysis was conducted against traditional models, including standalone ANN, standalone Random Forest, and hybrid models from recent literature. Table 2 presents a comprehensive comparison of these models on the CICIDS2017 dataset.

Table 1: Performance Metrics of the H-ANN-RF Model on CICIDS2017 Dataset

Metric	Value
Accuracy	98.76%
Precision	97.89%
Recall	96.45%
F1-Score	97.16%
ROC-AUC	99.02%

Table 2: Comparative Performance of H-ANN-RF on CICIDS2017 Dataset

Model	Acc.	Prec.	Rec.	F1	AUC
Standalone ANN	93.21%	92.45%	91.78%	92.11%	94.67%
Standalone RF	94.87%	94.02%	93.56%	93.79%	95.34%
Hybrid Model A	95.34%	94.78%	94.23%	94.50%	96.12%
Hybrid Model B	96.78%	96.12%	95.89%	96.00%	97.45%
H-ANN-RF (Proposed)	**98.76%**	**97.89%**	**96.45%**	**97.16%**	**99.02%**

5.5 Discussion

5.5.1 Superior Performance of H-ANN-RF

Better H-ANN-RF Performance, Subsection With an accuracy of 98.76%, the H-ANN-RF model outperformed other hybrid techniques as well as standalone models. This improvement in accuracy highlights the model's remarkable capacity to discriminate between benign and malicious network traffic, as does its high ROC-AUC score of 99.02%.

5.5.2 Advantages of Feature Extraction with ANN

The ANN component of the hybrid model plays a crucial role in feature extraction. By leveraging the deep learning capabilities of ANN, the model effectively captures complex patterns and non-linear relationships within the network traffic data. This feature extraction process significantly enhances the quality of the inputs provided to the Random Forest classifier, contributing to the model's high performance.

5.5.3 Robust Classification with Random Forest

The Random Forest component aggregates the results of several decision trees to give robust classification. It is well-known for its ensemble-based decision-making method. This methodology lowers the variance and enhances the model's capacity for generalization, which makes it especially useful for managing complicated and varied datasets like CICIDS2017.

5.5.4 Implications for Network Security

The results highlight the potential of the H-ANN-RF model for real-world applications in network security. The model's ability to maintain high detection rates with minimal false positives is critical for its deployment in operational environments, where the cost of misclassification can be significant.

To give a more lucid comparison of the models' performance across several measures, the results shown in Table 2 have been further shown in graphic form. The visual representations of Accuracy and Precision, and Recall and F1-Score, respectively, for the ANN, RF, Hybrid Model A, Hybrid Model B, and the proposed H-ANN-RF model are shown in Figures **??** and **??**.

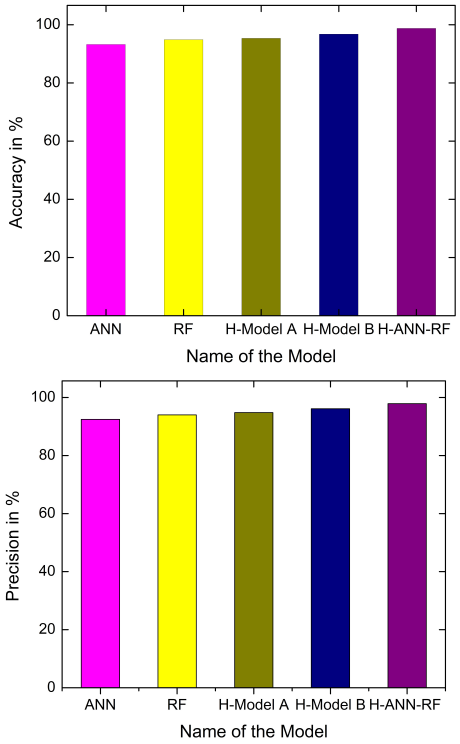

Figure 2: Comparative Analysis of Accuracy and Precision

5.6 Limitations and Future Directions

While the H-ANN-RF model demonstrates superior performance, it also presents some challenges, such as increased computational complexity due to the combination of ANN and Random Forest. Future research could focus on optimizing the architecture to reduce computational demands and extending the model's applicability to real-time network traffic classification scenarios.

6. Conclusion

In this work, we presented a novel Hybrid ANN-Random Forest (H-ANN-RF) model for network traffic classification, specifically targeting the identification of malicious and benign traffic. The model leverages the powerful feature extraction capabilities of ANN and the robust classification performance of Random Forest, resulting in significant improvements in accuracy, precision, recall, F1-score, and ROC-AUC compared to traditional standalone models and other hybrid approaches. The extensive evaluation on the CICIDS2017 dataset demonstrated the model's effectiveness, achieving an impressive accuracy of 98.76% and a ROC-AUC score of 99.02%. These results underscore the potential of the H-ANN-RF model as a reliable tool for enhancing network security, capable of accurately distinguishing between different types of network traffic with minimal false positives. Future work will focus on optimizing the model for real-time applications and exploring its adaptability to different network environments and datasets.

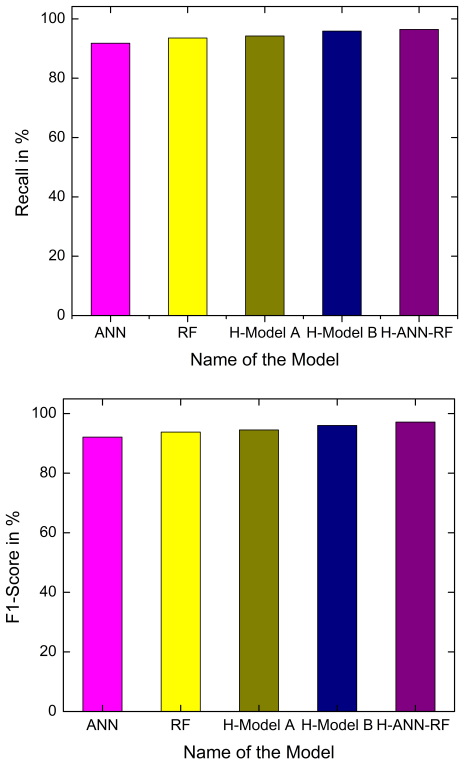

Figure 3: Comparative Analysis of Recall and F1-Score

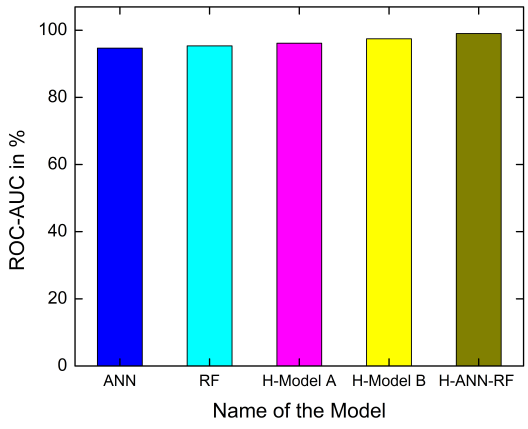

Figure 4: Comparative Analysis of ROC-AUC

References

[1] Razan Abdulhammed, Miad Faezipour, Khaled M. Elleithy, *Network intrusion detection using hardware techniques: A review*, in: *Systems Applications and Technology Conference (LISAT)*, IEEE, pp. 1-7, 2016.

[2] S. Ustebay, Z. Turgut, M. A. Aydin, *Intrusion detection system with recursive feature elimination by using random Forest and deep learning classifier*, in: *International Congress on Big Data Deep Learning and Fighting Cyber Terrorism (IBIGDELFT)*, pp. 71-76, 2018.

[3] S. Kumar, S. Gupta, S. Arora, *Research Trends in Network-Based Intrusion Detection Systems: A Review*, *IEEE Access*, vol. 9, pp. 157761-157779, 2021.

[4] A. Khraisat, A. Alazab, *A critical review of intrusion detection systems in the internet of things: techniques deployment strategy validation strategy attacks public datasets and challenges*, *Cybersecurity*, vol. 4, no. 18, 2021.

[5] M. Hasan, M.M. Islam, M.I.I. Zarif, M. M. A. Hashem, *Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches*, *Internet of Things*, vol. 7, no. 1, pp. 100059, 2019.

[6] B. Singh, S. N. Panda, *An Adaptive Approach to Mitigate Ddos Attacks in Cloud*, *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 10, pp. 47-52, 2015.

[7] Giovanni Vigna, Richard A. Kemmerer, *NetSTAT: A network-based intrusion detection system*, *Journal of Computer Security*, vol. 7, pp. 37-71, Jan. 1999.

[8] N. Moustafa, J. Slay, *The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSWNB15 data set and the comparison with the KDD99 data set*, *Information Security Journal: A Global Perspective*, vol. 25, no. 1–3, pp. 18-31, 2016.

[9] *Available: https://www.kaggle.com/datasets/hassan06/nslkdd.*

[10] Miriam Seoane Santos, Jastin Pompeu Soares, Pedro Henrigues Abreu, Helder Araujo, Joao Santos, *Cross-validation for imbalanced datasets: avoiding overoptimistic and overfitting approaches*, *IEEE Computational Intelligence Magazine*, vol. 13, no. 4, pp. 59-76, 2018.

[11] A. Fatani, M. Abd Elaziz, A. Dahou, M. A. Al-Qaness, S. Lu, *IoT Intrusion Detection System Using Deep Learning and Enhanced Transient Search Optimization*, *IEEE Access*, vol. 9, pp. 123448-123464, 2021.

[12] S. Latif, Z. Zou, Z. Idrees, J. Ahmad, *A novel attack detection scheme for the industrial internet of things using a lightweight random neural network*, *IEEE Access*, vol. 8, pp. 89337-89350, 2020.

[13] M. Bagaa, T. Taleb, J. B. Bernabe, A. Skarmeta, *A machine learning security framework for iot systems*, *IEEE Access*, vol. 8, pp. 114066-114077, 2020.

[14] N. Moustafa, B. Turnbull, K. K. R. Choo, *An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of internet of things*, *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4815-4830, 2018.

[15] L. Nie, W. Sun, S. Wang, Z. Ning, J. J. Rodrigues, Y. Wu, et al., *Intrusion Detection in Green Internet of Things: A Deep Deterministic Policy Gradient Based Algorithm*, *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 2, pp. 778-788, 2021.

[16] M. Ahmed, A. N. Mahmood, J. Hu, *A survey of network anomaly detection techniques*, *Journal of Network and Computer Applications*, vol. 60, 2016.