IJEMD-CSAI, Special Issue (2026) https://doi.org/10.54938/ijemdcsai.2026.04.2.660

IJEMD

International Journal of Emerging Multidisciplinaries: Computer Science and Artificial Intelligence

Research Paper

Journal Homepage: www.ojs.ijemd.com
ISSN (print): 2791-0164 ISSN (online): 2957-5036

A Two-Tier Hybrid Intrusion Detection System for IoT Networks

Wong Zoey [1], Yu Watanabe [1], Elena Loh Venyi [1], Wong Yong Jing [1], Siew Ryan [1], Pee Tze Hern [1], Teh Yu Xiang [1], Siva Raja Sindiramutty [1]

1. School of Computer Science, Taylor's University, Subang Jaya, 47500, Selangor Malaysia

1.0 Abstract

The rapid growth of Internet of Things (IoT) devices has made modern attacks more vulnerable to cyberattacks. Traditional signature-based Intrusion Detection Systems (IDS) are no longer enough to keep up with new and evolving threats. Although machine learning and deep learning have improved detection accuracy, many AI-driven IDS models still face major issues. They often struggle to detect zero-day attacks, produce high false-positive rates and perform poorly with imbalanced datasets. Some models are also too computationally heavy to run efficiently in real time. To address these weaknesses, this research proposes a two-tier hybrid IDS that uses a Random Forest model for quick initial detection and a Neural Network for deeper analysis of suspicious traffic. A confidence threshold of 0.8 is used to decide whether traffic should be accepted or sent for further inspection. Using the NSL-KDD dataset, the system includes preprocessing steps such as binary mapping and structured feature extraction to support both detection stages. Our comparative analysis shows that this hybrid approach can achieve better accuracy, fewer false alarms, and stronger detection of unknown attacks compared to existing Machine Learning / Deep Learning IDS methods. It is more practical for large, diverse IoT environments because it reduces computational load while maintaining strong detection capability. Overall, the proposed architecture provides a balanced and efficient solution that overcomes key limitations of existing IDS models and offers a pathway towards a more robust real-time IoT intrusion detection.

Keyword: Intrusion Detection System (IDS), Internet of Things (IoT), Hybrid Machine Learning, Random Forest, Neural Network

Email Addresses : 0366180@sd.taylors.edu.my, 0365660@sd.taylors.edu.my, 0366076@sd.taylors.edu.my, 0366658@sd.taylors.edu.my, 0365416@sd.taylors.edu.my, 0367602@sd.taylors.edu.my, teh.yuxiang@sd.taylors.edu.my, siva.sindiramutty@taylors.edu.my

2.0 Introduction & Problem Identification

2.1 Introduction to the Domain

An Intrusion Detection System (IDS) is a security system that monitors unauthorized access and abnormal network communications, notifying administrators to enhance the safety and reliability of a system. IDS can be categorized into signature-based detection models and anomaly-based detection models. The signature-based detection model detects attacks by comparing analyzed network traffic and log data with predefined malicious activity patterns. However, while it can detect known attack patterns, it is difficult for this model to identify new or unknown attacks. In contrast, the anomaly-based detection model identifies anomalies by discovering patterns and trends from large volumes of data. Therefore, it can handle unknown data. However, it also carries the risk of false positives.(Ansam Khraisat et al., 2019; Riskhan et al., 2025) To address these issues, it is necessary to use an IDS that provides automation. AI-driven IDS combines multiple machine learning approaches to achieve highly accurate detection. In particular, unsupervised learning is effective for the early detection of new or zero-day attacks. (Woo-Hyun Choi, 2024)Additionally, by using reinforcement learning, the IDS can continuously evaluate itself, automatically learn over time, and determine optimal responses. With the advancement of machine learning and AI, the capabilities of IDS are expected to further improve, enabling faster and more accurate anomaly detection. As cyber threats become increasingly complex and sophisticated, IDS is considered an essential and indispensable component for ensuring security.(Yakub Reddy, 2024) and (Fatima et.al. 2020)

2.2 Problem Identification

AI-powered Intrusion Detection Systems face the following challenges while achieving high-accuracy detection.

2.2.1 Vulnerability to Adversarial Attacks

An adversarial attack is an attack in which an attacker adds imperceptible noise that humans cannot notice, causing the model to misrecognize. This attack is designed to exploit the model's weaknesses. It confuses the model and leads to incorrect decisions. For example, if an adversarial attack occurs in a facial recognition system, the system may mistakenly identify the genuine user as someone else, or conversely, recognize an unauthorized person as the registered user. As a result, unauthorized logins may occur, leading to a decrease in reliability.(Sciforce, 2022; Sanjaya et al., 2025)

2.2.2 False Positive and False Negative

A False Positive occurs when an IDS incorrectly identifies normal access or communication as malicious, while a False Negative occurs when the IDS fails to detect an actual attack. A high rate of false positives increases the workload of system administrators and reduces trust in the IDS. Furthermore, AI-driven IDSs are more likely to miss attacks when there is insufficient training data for them, particularly in the case of zero-day or sophisticated attacks. In this way, both false positives and false negatives lead to higher operational costs and decreased reliability. In addition, although conventional AI-driven IDSs offer high

accuracy, they struggle to process large volumes of data in real time, making it difficult to detect attacks immediately. (Vincent Zibi Mohale & Ibidun Christiana Obagbuwa, 2025)

2..2.3 Imbalanced Dataset

An imbalanced dataset refers to a situation where the number of attack data samples is much smaller compared to normal communication data. This imbalance causes the AI model to learn in a biased manner, leading to skewed predictions. Models trained on imbalanced data have difficulty detecting anomalies, making it harder to identify new attacks such as zero-day attacks. In this way, imbalanced data lowers the accuracy of AI models.(Ahmed Abdelkhalek & Maggie Mashaly, 2023; Sindiramutty et al., 2024a)

2.3 Current solutions and their limitations

2.3.1 Solutions and Limitations to Adversarial Attacks

Currently, as solutions against Adversarial Attacks, the main methods are Adversarial Training and Defensive Distillation. Adversarial Training is a method that increases robustness to noisy data by using training data that contains noise. (Milvus, n.d.)This allows the system to detect anomalies even in noisy data. However, generating adversarial sample data requires repeatedly computing gradients, which greatly increases training time. In addition, training with adversarial samples tends to reduce accuracy on normal data. Furthermore, this method cannot defend against zero-day attacks. (Afnan Alotaibi & Murad A. Rossam, 2023)Defensive Distillation aims to improve the system's ability to continue operating stably even when subjected to Adversarial Attacks. As a result, it becomes less susceptible to small perturbations. However, Defensive Distillation requires training two models a teacher model and a student model which makes implementation difficult. Also, Defensive Distillation cannot defend against an attacker who knows this mechanism and creates new attacks.(GeeksforGeeks, 2025a)

2.3.2 Solutions and limitations to False Positive and False Negative

Currently, the main solutions for addressing False Positives and False Negatives are Ensemble Learning and Threshold Optimization. Ensemble Learning is a method that combines multiple learning models to make more accurate predictions. By combining multiple models, it improves prediction accuracy, robustness against noise, and flexibility. However, Ensemble Learning becomes more complex because it processes multiple models simultaneously. As a result, if errors increase, the computational cost rises and performance may decrease. (GeeksforGeeks, 2025b; Chaudhary et al., 2022)Threshold Optimization is the process of adjusting the threshold used to determine whether an output is positive or negative in order to find the optimal value. By optimizing the threshold, the prediction performance can be improved. However, adjusting the threshold is challenging if the threshold is set too high, more cases may be missed, and if it is set too low, false alarms may increase.(EVIDENTLY AI, 2025)

2.3.3 Solutions and Limitations to Imbalanced Dataset

Currently, the main solutions to the Imbalanced Dataset problem are the Synthetic Minority Over-sampling Technique (SMOTE) and Cost-Sensitive Learning (CSL). The Synthetic Minority Over-sampling Technique is a method that balances the dataset by generating new synthetic data for the minority

(abnormal) class. This helps improve the model's predictive performance. However, since synthetic data are created and used, the original data distribution may not be accurately reproduced, leading to a decrease in accuracy. Furthermore, if the original data contain noise, a large amount of noisy data may be generated, causing overfitting. (Cory Maklin, 2022)Cost-Sensitive Learning, on the other hand, is a technique that trains the model while considering the cost of each type of misclassification. This approach aims to minimize the total cost of misclassification errors. However, assigning different costs to different misclassifications increases the complexity of the model, which may lead to more false detections. In addition, cost-sensitive models heavily depend on the class distribution, reducing their adaptability to unseen data.(Ohad Volk & Gonen Singer, 2021; Brohi et al., 2020)

2.4 Recent real-world incidents

The first case is a practical study on adversarial attacks against AI-driven IDS. For example, GAN (Generative Adversarial Network) can be used as an attack method. Normally, GAN is used to create attack data. By including attack data in the training data, the detection ability of the model can be improved. However, GAN can also be used as an adversarial learning algorithm, which allows attackers to generate communication patterns that are difficult for IDS to detect. This shows that even AI-driven IDS may have difficulty detecting such attacks.(Afnan Alotaibi & Murad A. Rassan, 2023; Aldughayfiq et al., 2023)

The second case is the MOVEit Transfer zero-day attack carried out by the CL0P group. This attack occurred in 2023 when the group exploited a zero-day SQL injection vulnerability in MOVEit, causing data leaks and theft in various organizations, including political and healthcare institutions. Such zero-day attacks are difficult to detect using signature-based IDS. In addition, even AI-driven IDS may have difficulty detecting anomalies if the training data does not include data similar to the zero-day attack.(James Slaughter et al., 2023; Sindiramutty et al., 2024b)

3.0 Literature Review & Gap Analysis

3.1 Introduction to IDS in IoT Networks

The rapid growth of Internet of Things (IoT) devices has increased the risk of cyberattacks, requiring an adaptation of Intrusion Detection Systems (IDS) to secure the network. In this section, we will review the recent study on Intrusion Detection Systems (IDS) using Machine Learning and Deep Learning techniques, particularly for Internet of Things (IoT) networks. So that we can identify what the existing methods are, their effectiveness, and their limitations, which will help us identify research gaps and justify the need hybrid IDS combining Random Forest and Neural Network models.

3.2 Existing Research on ML/DL-Based IDS

3.2.1 ML-Based IDS Studies

Several studies have proposed different methods for improving IDS in IoT environments. Mohammed and Talib (Mohammed & Talib, 2024; Weiqi et al., 2024) reviewed common ML algorithms (Random Forest, Decision Tree, SWM, and KNN) in Intrusion Detection Systems. They discovered that ML-based IDS outperforms traditional signature-based systems and can improve detection accuracy compared to traditional systems. They evaluate how algorithms perform in identifying attacks such as Denial of Service (DoS), Probe, User to root (U2R), and Remote to Local (R2L) using benchmark datasets like KDDCup99 and NSL-KDD. The research emphasizes that Random Forest and SVM achieve higher detection accuracy, but because of dataset imbalance and overfitting remain challenges. However, they have problems in detecting zero-day attacks and often produce false positives. To overcome these challenges, future research suggests integrating Explainable AI (XAI) to make the IDS model more transparent and easier to understand. Therefore, integrating ML with other cybersecurity techniques, including encryption and secure communication, can also enhance general IoT security. Moreover, developing standardized benchmark datasets and evaluation frameworks will enable researchers to compare IDS models more effectively and improve the consistency of performance evaluations.

Building on these findings, Omitola and Wills (Nicho, Cusack, Girija, & Arachchilage, 2024; Ying et al., 2024) further discovered how classifier selection influences intrusion detection performance in IoT environments. They found out that ML classifiers can detect IoT sensor attacks if features are well-selected. That means when the input data is well pre-processed, like selecting the relevant attributes (packet size, signal strength, sensor readings, or frequency patterns), can improve the machine learning to be more accurate. It will help the classifier to distinguish between normal and abnormal sensor behaviours. Since IoT data can be noisy and dynamic, many ML models exhibit degraded performance. Therefore, IDS must be robust enough to deal with such noise; otherwise, they may not be suitable in highly dynamic IoT environments.

3.2.2 Hybrid and DL-Based IDS Studies

Naghipour, Salehpour, and Iranag (Salehpour, Balafar, & So, 2025; Annadurai et al., 2022) proposed an optimized IDS to be appropriate for resource-constrained IoMT situational contexts. The authors employed light-weight optimization methodologies that reduce computational costs but retain acceptable detection accuracy. While their approach works fine in small-scale configurations of IoT settings, it performed poorly when applied to high-traffic complex networks. This simply reflects the energy efficiency vs. quality of detection trade-off involved, especially in resource-limited computing environments.

Gavel, Raghuvanshi, and Tiwari (Wang Z, 2023; Almuayqil et al., 2022) presented a lightweight deep learning method that applies dynamic quantization to minimize memory usage in IoT devices. The model achieved competitive accuracy with faster response times but was plagued by high false-positive rates when handling real-world noisy data. Their work thus suggested that deep learning models require better feature selection or hybrid techniques that balance between accuracy and computational efficiency.

Muhammad (Alashjaee, 2025; Al-Musib et al., 2021) developed a hybrid IDS on feed-forward Neural Networks with traditional machine learning algorithms for IoT networks. It means they use Neural Networks to understand high-dimensional data, and the machine learning will detect the attacks. They discovered that the hybrid can detect a wide range of IoT attacks (DoS, Probe, U2R, R2L) and effectively handle high-dimensional, heterogeneous IoT data. The benefit of the hybrid approach allows combining the strengths of ML and DL techniques. They still faced challenges, like computationally intensive, requiring more processing power and memory, performance may degrade in resource-constrained IoT devices, and real-time detection could be affected due to high computational overhead.

3.2.3 Statistical and Federated Learning-Based IDS Studies

Walling (Kaushik, Sunil, 2025; Alwakid et al., 2023) used basic statistical techniques to choose the important features from the data and then applied ML classifiers to detect the abnormal behaviour in IoT devices. This approach is simple, fast and does not require training complexity. It also runs well even if the data has some noise or error. However, it has limited scalability, such as not scaling well with different types of IoT devices, and the accuracy will drop when it is highly heterogeneous IoT environments.

Hendaoui (Hendaoui, Meddeb, Trabelsi, Ferchichi, & Ahmed, 2025) created FLADEN, an anomaly detection system that uses federated learning. That means IoT devices train data locally without sending their raw data to a central server. The advantages of this is preserves data privacy while it works well across many IoT devices, since each device helps train the model. Therefore, communication overhead because it needs to share model updates, which can slow down the system, and needs a better way to combine all the models.

3.2.4 IoT Growth Trends

Statista (Number of Internet of Things (IoT) connections worldwide from 2022 to 2023, with forecasts from 2024 to 2034, 19) reports that the number of IoT devices will reach billions by 2034. This means IoT networks are growing extremely fast, becoming larger and more complex. However, security systems like IDS must be able to handle large amounts of data so that they can work well with different types of devices and handle large numbers of devices. IDS must be able to operate effectively so that it will not slow down large IoT networks.

3.3 Gap Analysis

Based on the reviewed literature, several gaps and limitations in the existing IoT IDS can be discerned:

Handling Zero-Day Attacks

- Indeed, Mohammed & Talib (Mohammed & Talib, 2024) proved that many ML-based IDS do not detect previously unseen attacks.
- Limitation: Most current models rely either on the known attack pattern or imbalanced datasets and result in false negatives.
- Gap: There is a need for models that can generalize better to unknown attacks.

Data Noise and Feature Quality

- Omitola & Wills (Nicho, Cusack, Girija, & Arachchilage, 2024)\ argued that good performance relies a lot on the selection of features. Poor data preprocessing will reduce the accuracy.
- Gap: IDS models have to be resilient in noisy and very dynamic IoT contexts.

Computational Efficiency

- Muhammad (Alashjaee, 2025) pointed out that hybrid ML/DL approaches improve detection but are computationally intensive.
- Gap: Resource-constrained IoT devices require lightweight solutions without sacrificing accuracy.

Scalability and heterogeneity

- According to Walling (Kaushik, Sunil, 2025), statistical feature selection reduces complexity but cannot handle highly heterogeneous IoT environments.
- According to the statistics shown by Statista (Number of Internet of Things (IoT) connections worldwide from 2022 to 2023, with forecasts from 2024 to 2034, 19), billions of IoT devices are projected, and a scalable IDS is essential.
- Gap: Existing IDS might not sustain performance for large diverse networks.

Privacy and Distributed Learning

- Hendaoui (Hendaoui, Meddeb, Trabelsi, Ferchichi, & Ahmed, 2025) proposed federated learning for privacy, though communication overhead and slow convergence remain challenges.
- Gap: There is a need for efficient privacy-preserving IDS solutions that minimize the network overhead.

3.4 Connecting Literature to the Proposed Solution

Our two-tier hybrid IDS design is directly motivated by the gaps identified in the reviewed literature.

Machine learning-based IDSs often fail to detect zero-day attacks due to limited generalization and imbalanced datasets, as shown by previous studies. Our system's neural network tackles this by learning non-linear, high-dimensional patterns, which leads to stronger recognition of unfamiliar behaviors.

Numerous studies have shown that there is a high rate of false-positives and limited robustness when using noisy or heterogeneous IoT traffic. Random Forest is a reliable first-stage filter due to its stability on structured features and lower noise sensitivity. The Neural Network can be optimized by directing only low-confidence samples to reduce unnecessary false alarms and computational overhead.

The third point is that IDS models that use deep learning are computationally heavy and unsuitable for real-time IoT use. Random Forest processes 80–90% of routine traffic quickly with our selective routing mechanism, while only a small portion undergoes deeper inspection, which improves efficiency without losing accuracy.

In earlier research, concerns about scalability and adaptability remained constant. The modular hybrid architecture is capable of being deployed at the edge or cloud level and can be enhanced to allow for privacy-preserving methods such as federated learning when necessary.

To summarize, the proposed two-tier hybrid IDS is designed to address the key shortcomings identified in current literature by improving detection accuracy, reducing false alarms, and enabling efficient real-time performance in diverse IoT environments.

4.0 Proposed Secure System

4.1 System Overview & Architecture

The introduction of this system uses intelligent triage to detect intrusions. Machine-learning methods have the potential to be faster but less accurate than traditional IDS models, leading to a clear trade-off, deep-learning models can be very accurate, but they are not fast enough for real-time use. A hybrid design is used in the proposed system to overcome this limitation.

The Random Forest classifier filters all network traffic before it goes through a dynamic confidence threshold set at 0.8, which is a major innovation, The ability to classify approximately 80–90% of routine traffic quickly and confidently is possible. A more powerful Neural Network is used to analyse the sample when the model's confidence falls below the threshold.

High detection accuracy is maintained while computational cost is significantly reduced using this selective routing. An IDS that achieves both efficiency and precision is the outcome, achieving a solution that is a balance between speed and accuracy, which are challenges in current systems.

4.2 Technical Specifications

Sequential triage logic is the basis of the proposed TS-HIDS' operation, Created with the intention of optimizing the balance between inspection depth and processing latency in highly trafficked IoT environments. There are three technical modules that make up the system architecture: the Data Preprocessing Unit, the High-Speed Filtration Tier and the Deep Inspection Tier.

4.2.1 Data Preprocessing and Transformation

The NSL-KDD benchmark dataset's raw network traffic data undergoes a rigorous preprocessing pipeline to ensure compatibility with machine learning algorithm

- Label Encoding:
  The raw dataset contains categorical string features such as protocol type (e.g., TCP, UDP), service type (e.g., HTTP, private), and flag (e.g., SF, S0). These categorical strings are converted into numeric vectors by the system using a label encoder, the classifiers can perform mathematical computation.
- Binary Classification Mapping:

Detecting binary anomalies is the main objective of this system, even though the dataset includes specific attack labels like Neptune, Smurf, and Satan. A mapping function aggregates all malicious labels into a single class

(1: Attack) and all benign traffic into another class (0: Normal).

- Data Splitting
  To ensure that the evaluation is accurate, the data is sorted into training (KDDTrain+) and testing (KDDTest+) sets for unseen traffic.

4.2.2 Tier 1 The High-Speed Random Forest Filter

An Ensemble Random Forest Classifier is the foundation of the first line of defence. With minimal latency, this model is designed to act as a high-speed filter that handles the majority of network traffic.

To achieve stable and fast predictions, the Tier 1 Random Forest is configured with 100 estimators, which combine the decisions of 100 trees. To prevent overfitting and maintain a lightweight inference, the maximum depth is limited to 20. When n_jobs = 0, the model can handle large traffic volumes in real-time by using all CPU cores for full parallel processing. Clear-cut cases, which are either normal or malicious, are quickly resolved in this tier, ensuring that only uncertain samples move on to the highest level. The heavier Tier 2 model is only accessible to uncertain samples.

4.2.3 The probabilistic Routing Protocol(The “Gatekeeper”)

The fundamental advancement of the TS-HIDS lies in its adaptive routing mechanism. In contrast to conventional IDS models that make a quick binary choice (either 0 or 1), our system assesses the statistical confidence of the Tier 1 prediction prior to confirming it.

The routing mechanism of the system operates on a straightforward probability principle. The Random Forest generates a confidence score P that ranges from 0 to 1, and this score is assessed against a predetermined confidence threshold $origin = 0.8$. When P Greater than or equal to 0.8 , the prediction is deemed high-confidence, and the Tier 1 result is immediately accepted. Conversely, if P is smaller than 0.8 , the traffic is classified as uncertain and automatically directed to Tier 2 for further evaluation. This decision-making approach enables the system to effectively distinguish between clear cases and those that are ambiguous.

This system guarantees that the intensive computations involved in Deep Learning are dedicated solely to the small percentage of traffic that poses classification challenges (around 15-20% of the data), while the bulk is processed without delay.

4.2.4 Tier 2 The Deep Inspection Neural Network

Traffic categorized as "Uncertain" by the routing protocol is directed to the Tier 2 Multi-Layer Perceptron (MLP).

The architecture employs a neural network comprised of two dense hidden layers: the first layer contains 64 neurons for feature extraction, and the second layer consists of 32 neurons for pattern analysis.

Activation Function: The ReLU (Rectified Linear Unit) activation function is employed in the hidden layers of the system. ReLU is selected for its efficiency in computation and its capability to address the "vanishing gradient" issue, enabling the model to capture non-linear patterns in intricate attack data.

The function of this tier is to act as the "Specialist." It examines the intricate, non-linear patterns in the unclear traffic (like slow-rate DoS attacks or disguised probe attempts) that the linear decision trees in Tier 1 failed to clarify with a high level of certainty.

4.3 Prototype Implementation

To assess the practicality of the proposed TS-HIDS architecture, a complete prototype was created and tested using the Python programming language. The implementation utilizes the Scikit-Learn machine learning library for developing models and Pandas for handling large-scale data.

Prototype Setup The simulation utilized the KDDTrain+ dataset for training the model and the KDDTest+ dataset for thorough evaluation. The prototype script mimics a real-time environment where traffic is processed, preprocessed (utilizing Label Encoding and Binarization), and assessed sequentially by the hybrid logic.

Key Logic of Implementation The primary innovation of this prototype is the "Triage" routing algorithm, which actively intercepts the predictions generated by the Random Forest. The code segment (Screenshot 4.1) below illustrates the routing logic that has been applied in our Python prototype.

```
confidence_threshold = 0.8

max_proba = np.max(rf_proba, axis=1)
uncertain_mask = max_proba < confidence_threshold

certain_indices = ~uncertain_mask
uncertain_indices = uncertain_mask

certain_count = certain_indices.sum()
uncertain_count = uncertain_indices.sum()
certain_percent = (certain_count / len(y_test)) * 100
uncertain_percent = (uncertain_count / len(y_test)) * 100

print(f"confirm data: {certain_count} ({certain_percent:.1f}%)")
print(f"not sure data: {uncertain_count} ({uncertain_percent:.1f}%)")
print(f"\n→ only {uncertain_percent:.1f}% data need deep thinking! \n")

print("-" * 60)
print("Second Pharse: Training Neural Network Model...")
print("-" * 60)

nn_model = MLPClassifier(
    hidden_layer_sizes=(64,32),
    activation ='relu',
    max_iter=100,
    random_state=42,
    verbose=True
)

print("\nStart Traing... Might be use 3-5min")
print("(You will saw the progress msg)\n")

start_time =time.time()

nn_model.fit(X_train, y_train)

nn_training_time = time.time() -start_time
print(f"\n✓ Neural Network model traing done! Total: {nn_training_time:.2f} s\n")

if uncertain_count > 0:
    print("Doing to the unsure data deep analytic")
    X_uncertain = X_test[uncertain_indices]
    y_uncertain = y_test[uncertain_indices]

    nn_predictions_uncertain = nn_model.predict(X_uncertain)
    nn_accuracy_uncertain = accuracy_score(y_uncertain, nn_predictions_uncertain)
    print(f"The accuracy of the unsure data: {nn_accuracy_uncertain*100:.2f}%\n")
else:
    print("There are no unsure data\n")

print("-"*60)
print("combining two result...")
print("-"*60)

final_predictions = rf_predictions.copy()

if uncertain_count > 0:
    final_predictions[uncertain_indices] = nn_predictions_uncertain
```

Screenshot 4.1

5.0 Experimental Setup/ Comparative Analysis

5.1 Comparative Analysis with Existing Solutions

A two-tier hybrid intrusion detection system combining random forest and neural networks for IoT networks is proposed in our project. In order to show the effectiveness of the performance and design of our proposal, the comparison against two existing security solutions, which are lightweight deep learning IDS for IoT (Wang Z, 2023) and a Feed-forward Neural Network (FFNN) with traditional machine learning algorithms (Alashjaee, 2025), (Humayun et.al. 2022).

5.1.1 Lightweight Deep learning IDS

The first existing security solution is a lightweight Deep learning IDS. Deep learning has become a technique to improve the performance of network intrusion detection effectively due to the complexity of attack environment. In recent years, many deep learning models have been employed and most of them are computationally complex and causes limitations to function effectively on IoT devices. Hence, it is important to focus on developing an IoT intrusion detection system including identifying network attacks in a shorter training time or develop an intrusion detection system with robust detection capabilities (Wang Z, 2023). To achieve the outlined goals, a lightweight intrusion detection method for IoT based on deep learning and dynamic quantization is introduced. This model was introduced to maintain high detection accuracy. A dynamic quantization method is applied to compress specified network structure units to produce a more streamlined intrusion detection model, reduce the memory footprint of the model and enable faster inference without significant loss in accuracy (Wang Z, 2023).

Table 1

Table 6 Multiclassification results of the CICIoT2023 dataset.

Model	Acurracy	Precision	Recall	F_1	Training time(s)	Inference time(s)
CNN	0.9221	0.9149	0.9221	0.9126	1515.4	7.2
RNN	0.9273	0.9124	0.9273	0.9150	717.8	8.5
LSTM	0.9275	0.9132	0.9275	0.9152	764.8	6.6
BiLSTM	0.9305	0.9133	0.9305	0.9173	792.6	8.0
DL-BiLSTM	0.9313	0.9180	0.9313	0.9194	708.4	6.4

(Wang Z, 2023)

Based on the diagram above, it is shown that the system achieved approximately 93-95% detection accuracy on IoT-based datasets (Wang Z, 2023). It identified multiple categories of network anomalies, such as DDoS, port scanning, and botnet activities. However, the approach suffers from a high false-positive rate, which occurs when the normal traffic behaviour is different from the training data. Therefore, by using the quantization method can successfully reduce computational cost and memory usage, which proves that the model is suitable for lightweight IoT environments, but it also causes a decrease in

precision when dealing with complex patterns. Although this method is suitable for lightweight environments, it still requires optimization to improve detection reliability and stability.

5.1.2 Feed-forward Neural Network (FFNN)

The second existing solution is an enhanced intrusion detection system that integrates a Feed-forward Neural Network (FFNN) with traditional machine learning algorithms to improve IoT network traffic. This FFNN model can be utilized to handle complex and high-dimensional IoT network traffic patterns. Its deep structure provides a better function in extraction, whereas XGBoost can enhance classification efficiency, making the combination for intrusion detection more powerful and useful (Alashjaee, 2025).

Table 2

Method / Parameter	Accuracy	Precision	Recall	F1-score
Proposed	99	97	98	97
FFNN	90.41	94.75	90.41	92.30
XGBoost	99.66	95.36	97.51	96.41
GA-mADAM-IIoT [16]	98.50	96.80	97.20	97
Modified GA + LSTM [17]	97.90	95.10	96.40	95.70

(Alashjaee, 2025)

Based on the diagram above, the model achieved an average accuracy rate of 96.8% from the UNSW-NB15 dataset and strongly performed in detecting attack types such as denial-of-service (DoS) and probing. However, the model faced some challenges like high computational complexity, low detection accuracy, and inefficiency when working with different types of attacks due to the deep structure of FFNN which may require a longer training time (Alashjaee, 2025). Moreover, when an imbalanced dataset is trained, it leads to overfitting and might be unable to generalize effectively to zero-day attack behaviour. Therefore, the limitations of the model will be found and making it less practical in a large IoT environment that requires low latency and adaptability for deployment.

5.1.3 Comparison with the existing solutions

In comparison, the proposed two-tier hybrid intrusion detection system approaches the strengths of both traditional machine learning and deep learning to diminish their weaknesses. For the Tier-1 Random Forest model, 87.6% of routine traffic able to classify with high confidence to reduce the computational workload, while the 12.4% uncertain samples are forwarded to Tier-2 for deeper analysis. In order to make the system suitable for real time IoT deployment, the selecting routing mechanism will ensure low latency and usage of resources efficiently. In comparison, the FFNN achieved high accuracy, high computational complexity and difficult to handle imbalanced dataset causing it less practical and unstable for IoT environments, while the hybrid system can increase efficiency by routing only small portion of data to the

tier-2 neural network system, which provides a better and stable performance under traffic loads. Moreover, the hybrid architecture can also improve on adaptability and stability. The proposed system can maintain performance even though it encounters different traffic patterns as the tier 1 can handle traffic patterns and tier 2 can detect anomalies to enhance robustness of the system, while the lightweight deep learning IDS and Feed Forward Neural Network relied heavily on the training data which might cause overfitting if the model unable to generalize unseen attacks. In conclusion, the proposed two-tier hybrid intrusion detection system demonstrates that it is more practical for IoT environments as it increases robustness of the system against traffic than both existing solutions.

5.2 Benchmarking for Comparative Analysis

The evaluation of benchmarks for comparative analysis can be carried out by using consistent datasets and data preprocessing procedures to indicate that the results were caused by different models instead of inconsistent data handling. For example, the goal of the benchmarking process is to measure performance accuracy by the evaluation metrics for comparison of the performance, increase the efficiency of computation by the model training, increase robustness and practical for IoT environments. Firstly, the types of performances such as classification and operational were evaluated for performance comparison to ensure a complete comparative analysis. Those key evaluation metrics include accuracy, precision, recall, F1-score, false positive rate, false negative rate, confusion matrix, and ROC-AUC score for evaluation. F1-score and recall are highly emphasized as most of the IoT datasets are imbalanced. Furthermore, the computational efficiency can show that measurements like inference latency, model throughput, CPU and memory utilization, and traffic percentage forwarded to Tier 2 in the hybrid model could perform well in real-time situations, which is a benefit to compute with the selecting routing mechanism used in the proposed system and also maintain a low latency intrusion detection in IoT environments. Moreover, it is important for a system to increase robustness and make it practical for IoT environments. For instance, when traffic patterns vary or unknown types of attack appear, both Tier-1 and 2 can help to prevent misclassification of traffic or attacks to maintain high detection accuracy.

5.3 Simulation of Hybrid System

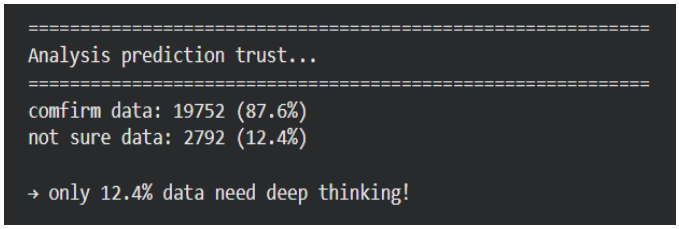


Figure 1.0

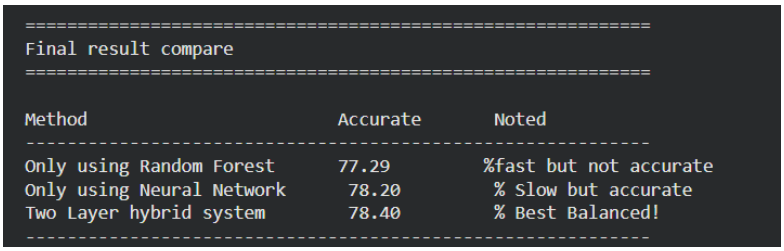


Figure 1.1

The dataset consists of 125973 training data and 22544 testing data. Firstly, the dataset will be inserted, then carried out data preprocessing through label encoding, normalization and binary mapping to differentiate between normal and attack traffic. Moreover, the tier 1 random forest was trained to classify all future test samples and produce accuracy and a high confidence score for each prediction. Those samples will be classified with a confidence value of 0.8 or higher. A confidence value of 0.8 or higher will be accepted immediately, whereas a lower confidence value will be transferred to tier 2 for deeper inspection using the neural network system. Furthermore, the data analysis prediction shows in figure 1.0 that 87.6% of the test samples are classified with high confidence by tier 1 random forest in the first phase, which means that only 12.4% of the data needs deep thinking by tier 2 neural network in the second phase. This kind of selecting routing mechanism can help to reduce the computational cost and improve system efficiency. In figure 1.1, it shows that the hybrid system achieved the highest accuracy which is 78.40% among the random forest which only achieved an accuracy of 77.29% and the neural network with an accuracy of 78.20%. This proved that the two-tier hybrid system offers a more balanced and effective intrusion detection system.

5.4 Technical Metrics of Two-Tier Hybrid IDS

Table 3

Method	Accuracy (%)
Random Forest	**77.29**
Neural Network	**78.20**
Hybrid System	**78.40**

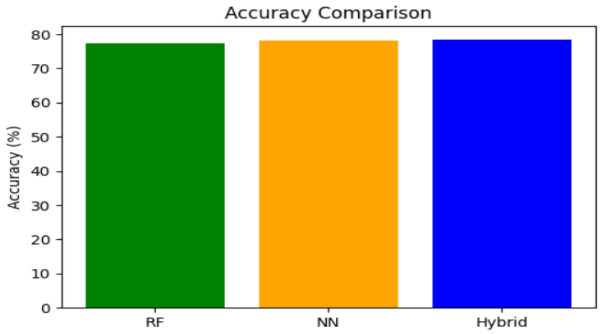


Figure 1.2

The table 3 and Figure 1.2 shows that the hybrid system achieved the highest accuracy in maintaining system efficiency.

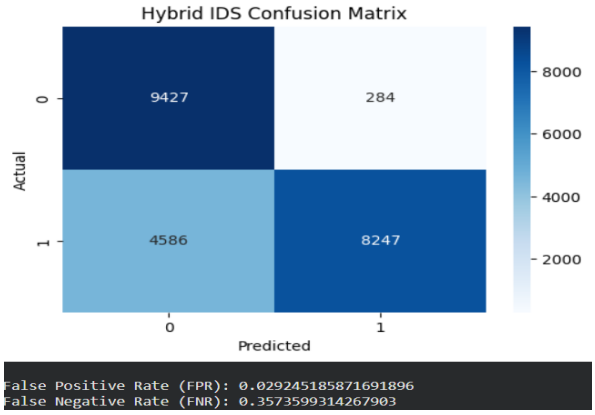
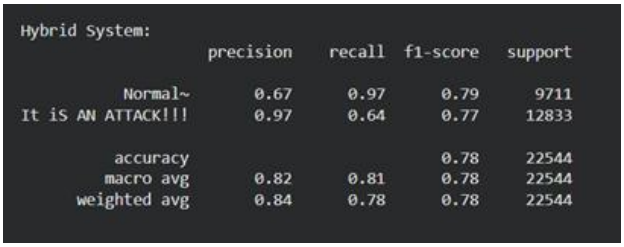


Figure 1.3

Furthermore, the hybrid IDS confusion matrix shows in Figure 1.3 that false positive rate is 0.029 and false negative rate is 0.357. This indicates that the lower the false positive rate, the higher trust of the system and those uncertain samples will cause the false negative rate to be higher.



Hybrid System:

	precision	recall	f1-score	support
Normal~	0.67	0.97	0.79	9711
It is AN ATTACK!!!	0.97	0.64	0.77	12833
accuracy			0.78	22544
macro avg	0.82	0.81	0.78	22544
weighted avg	0.84	0.78	0.78	22544

Figure 1.4

In addition, the hybrid model is classified into the normal class (0) and the attack class (1). Each class had defined its precision, recall and F1-score which are shown in Figure 1.4. Therefore, the overall accuracy of the hybrid system is 78%.

6.0 Implementation Challenges & Feasibility

6.1 Deployment Issues

The solution exhibits robust promise for IoT security, mainly by virtue of its selective routing mechanism and trade-off between speed and accuracy. However, implementing this system in real-world IoT environments is challenging. IoT environments exhibit a high diversity of devices containing various communication protocols and hardware specifications as well as operating systems. The proposed hybrid IDS must be deployed in an environment with numerous network architectures including edge, fog, and cloud. In addition, supporting different communication protocols, and how to adapt according to the security posture at each device will affect the data you are storing. Since TS-HIDS preprocesses network traffic using numerical encoding and feature extraction, any inconsistencies in both log formats or packet structure by IoT devices in comparison to other network devices may greatly distort models. Despite its speed-optimized features, Random Forest has difficulties with real-time intrusion detection, such as large volumes of data streams, which might overload the preprocessing unit. At times latency increases as a large percentage of traffic becomes “uncertain” and is sent to Tier 2. If the pipeline for feature extraction is poorly optimized, real-time detection performance may not be obtained by the system. IDS must gather network traffic that may include personal data, user behavioral patterns, and sensitive files or communications, and employ encryption mechanisms for data-in-transit and data-at-rest; thus the computational overhead is also higher.

6.2 Scalability Challenges

Since most of the IoT devices have low power, limited RAM, and low-frequency CPU, this will make them unable to run a lightweight Random Forest model through micro-controllers or sensors directly. As a result, TS-HIDS would be primarily built around edge gateways or cloud servers themselves. These two methods add costs and logistical complexity. However, as the number of IoT nodes increases from tens to thousands, the number of packets will exponentially increase and the preprocessing unit would require encoding millions of rows of traffic consistently. This could result in saturated infrastructure without horizontal scaling (e.g., the establishment of multiple parallel nodes). New attacks mean that IDS models must undergo retraining regularly to stay accurate. But a Neural Network needs a lot of computational resources available in order to retrain. If large-scale retraining pipelines or frequent updates disrupt system availability and increase storage or bandwidth use, an appropriate model update mechanism is required but is complex to implement

6.3 Cost Considerations

The implementation of TS-HIDS has many hardware and infrastructure investments. There's investment to be made for edge servers or fog nodes to have processing distributed at scale and additional storage systems for large volumes of traffic logs, GPU-equipped machines for data retraining of a model, and high-speed network equipment for efficient flow of data. These needs raise starting deployment costs dramatically for companies not now having any machine learning-ready infrastructure. As well as the costs of initial deployment, TS-HIDS also incurs operational and maintenance expenses. This includes routine system retraining for models to guarantee their correctness, constant system monitoring and tuning

for performance maintenance, storage for storing long-term historical data of traffic, and maintenance through regular software updates and/or vulnerability patches. Since the system is hybrid in nature and needs to be maintained by two different models at the same time, this will increase the engineering burden and operation costs even more. In addition, false positives are not cheap per unit cost. TS-HIDS endeavors to keep false positives and false negatives at the minimum but misclassification carries economic costs. Excessive alerts can deplete security personnel's resources, and missed attacks can cause data breaches, service outages, or fines. So defining the optimum confidence threshold is very important, which can involve a meticulous analysis of the trade-offs to achieve accuracy in detection and efficiency in operation.

6.4 Ethical and Legal Concerns

Intrusion detection systems (IDS) access every network communication and analyze them with significant ethical implications for user privacy. Organizations must think about how much user activity should be monitored, if it's known or agreed to by users. The retention period of logs also matters a lot since storing more records than needed could lead to privacy abuses in case of mismanagement or leakage of data. Random forests and neural network models can cause algorithmic biases, since these work largely with training datasets. While IoT traffic data are collected in virtual environments, the datasets they are trained on do not necessarily reflect the full range of real-world environments. Accordingly, the system can be biased on familiar traffic patterns and would classify unusual or rare behavior of devices as "rare." Such a bias would result in higher false alarm rates, outages of IoT devices, and unfair discrimination of many devices or protocols of different networks. Neural network models being "black boxes" are a difficult topic to explain; these are going to give rise to concerns of transparency and explainability. When security decisions are confusing, confidence in a system erodes. And this ambiguity also makes it more difficult to investigate for the purpose of forensics, where the reasoning behind flagging particular traffic as malicious will not always make the most sense to security teams. From an ethical perspective, organizations need to promote accountability of automated decisions through explainable AI technologies or provide additional documentation to back up model outputs.

6.5 Technical Feasibility

TS-HIDS works on a technical level because its core methods which are random forest and MLP can be deployed at the edge layer, making it a locally computationally efficient framework. The intelligent routing methodology improves the computational cost and, as shown in the performance evaluations carried out with the NSL-KDD dataset, it is feasible to achieve a notable improvement in detection performance in practice. Large-scale deployments could also require more processing to hold stable performance and strict real-time detection needs still pose challenging practicality in complex or high-traffic environments. Operationally, the system works seamlessly in the context of advanced IoT architectures which often use edge gateways and centralized security monitoring solutions. Firms which have medium technological capabilities can seamlessly adopt the system, thereby minimizing disruptions in their processes. Yet not for small institutions with fewer resources or knowledge base and who face challenges in dealing with distributed processing, model updates, and daily monitoring, making system adoption more challenging without the resources. Economically, an initial deployment of TS-HIDS is expensive both in hardware cost, data processing infrastructure, and maintenance of the model. Despite

this, the system also provides many long-term economic advantages including decreased false positives, reduced the burden on the security team, increased detection accuracy, and the prevention of costly security vulnerabilities. The bigger the network, the cheaper the system becomes, meaning that for medium and large deployments the system becomes more affordable.

7.0 Evaluation & Discussion

7.1 Strengths

Based on the proposed two-tier hybrid Intrusion Detection System, it demonstrates a strong potential for improving security in IoT environments by achieving a balance between detection accuracy, efficiency, and adaptability. Using the two-tier hybrid Intrusion Detection System provides a mutualistic effect. This is because by using this hybrid architecture, the Random Forest classifier performs feature filtering and preliminary classification, reducing noise and lowering computational overhead before forwarding data to the Neural Network for final decision-making. The combination between Random Forest and Neural Network provides a higher detection accuracy compared to single-model IDS due to the hybrid architecture. (What are the three types of IDS?, n.d.)

Other than that, the two-tier hybrid Intrusion Detection System promises good scalability as it can be deployed on distributed IoT gateways and edge-processing nodes allowing real-time traffic analysis without relying on cloud infrastructure. This is important for IoT networks with high heterogeneity and geographically distributed endpoints. It can also be very adaptive to new attacks as Neural Network components can learn from updated datasets over time.

Furthermore, it processes efficiently as Random Forest feature selection reduces computational load for Neural Network components, the overall computational cost is significantly lower than a deep-learning-only IDS. The balance between the performance and resource usage enables the system to detect a wide variety of attacks, including DDoS, port scanning, brutal force login attempts and botnet activity without overwhelming the hardware resources of the IoT gateway.

7.2 Weaknesses

However, the system is not without limitations. The hybrid IDS system requires higher processing power and memory than traditional lightweights IDS models, While IDS models are suitable the system may not run well on extremely resource constrained IoT sensors. This could limit the deployment as hardware capabilities are minimal.

Next, the system also highly depends on a labelled dataset for training. Since the model relies heavily on training data quality, the model may suffer if datasets do not reflect real-world traffic distribution. Retraining becomes very required if a new attack type evolves to maintain effectiveness. Organizations must plan for the operational overhead.

Other than that, latency may increase under high network traffic conditions since two classification types are performed sequentially, heavy traffic flows may increase processing time, that may cause the real-

time responsiveness to be affected. For time critical fields such as healthcare, this latency may lead to critical risks.

Furthermore, there is a risk of concept drift, where statistical properties of the target variable change over time, causing the model's prediction to become less accurate. Firmware updates, newly added IoT devices or changing of user activity patterns could lead to concept drift. This usually happens because of no continuous monitoring and model updates. Detection accuracy could decline. Maintenance of such systems also raises ethical considerations, involving data privacy and compliance with cybersecurity regulations if a network logs certain sensitive user information.

7.3 Comparison

Traditional signature-based IDS such as Snort and Suricata (Suricata vs Snort, n.d.) are considered industry standards in many organizations. These systems follow a rule-matching approach which compares incoming network traffic against known attack signatures. It is highly effective in detecting the previously identified threats but struggles with zero-day attacks and emerging malware because they rely on threat patterns that are already stored in the signature database.

The hybrid IDS addresses this gap by adopting a behaviour-based detection paradigm (ReasonLab, 2023). The proposed hybrid model overcomes this limitation by combining Random Forest to identify irregular statistical patterns and Neural Network that learns behavioural characteristics of network traffic. This two-tier model allows the system to recognise attacks such as DDoS, botnets, port scans and brute force attempts even when their signatures are unknown or intentionally disguised. Therefore, the hybrid IDS model not only can detect the known threats accurately but can also detect unknown and evolving attacks. This aligns with the NIST Cybersecurity Framework ((NIST), 2025), which emphasises the need for continuous and adaptive threat detection rather than reactive defence.

Compared with deep-learning-only IDS systems (B.V, 2021), the hybrid IDS system is a great choice as deep-learning-only IDS systems are normally very accurate but are often computationally expensive, difficult to retrain and impractical for constrained IoT hardware. The hybrid system is just right because it offers a high detection performance of deep learning while reducing the complexity through feature filtering using Random Forest.

Overall, the system shows compatibility with industry standards for modern IDS, in terms of adaptability, scalability and proactive defence. This architecture reflects where current cybersecurity direction, AI-driven detection and edge processing are becoming extremely important.

7.4 Conclusion of Strengths, Weaknesses and Comparison of two-tier hybrid intrusion Detection System

The two-tier hybrid intrusion Detection System offers a strong and well-balanced security solution for IoT environments. By integrating Random Forest and Neural Network components into a sequential architecture, the system archives improved detection accuracy, adaptability, and scalability. Compared with industry standard signature-based IDS systems, the hybrid system shows ability to detect both known

and unknown attacks and align with modern cybersecurity expectations emphasised by the NIST Cybersecurity Framework. However, the system also has weaknesses that must be improved, including higher resource demand compared with lightweight IDS, dependence on labelled data, retraining overhead and latency due to network traffic. These factors must be solved to ensure long-term reliability and operational efficiency. Despite these challenges, the hybrid IDS represents a promising advancement in IoT cybersecurity and provides viable bridge between traditional IDS and next generation AI-driven threat detection.

8.0 Conclusion and Future Works

This report looked at the ongoing security issues in IoT environments, where many current AI-based Intrusion Detection Systems (IDS) still struggle with zero-day attacks, false positives and heavy processing requirements. The hybrid IDS we proposed, using Random Forest for quick filtering and a Neural Network for deeper analysis, aims to offer a more balanced approach improving accuracy without demanding too much computational power. While the design shows potential, there are still several areas where the system can be strengthened to perform better in real-world situations.

Future improvements for the hybrid IDS could focus on making the system more flexible, scalable and practical for real-world use. Studies show that combining both signature-based and anomaly-based detection can lead to better accuracy and fewer false positives than using other methods alone, which supports our current model (WJARR, 2025). A key next step is testing the model using more recent and IoT-focused datasets, since older datasets like KDD CUP 1999 no longer reflect modern network traffic or attack behaviours (WJARR, 2025)

Another improvement involves integrating Continual Learning (CL). Unlike traditional models that need to be retrained from the beginning whenever new threats appear, Continual Learning allow the IDS to keep updating itself over time while still retaining knowledge of past attacks. This helps the system to react better to zero-day attacks and ongoing changes in attacker behaviour. (Guo, Li, Cheng, Yang & Gong, 2025)

Finally, we could also include Explainable AI (XAI) to make decisions of the Random Forest and Neural Network easier to understand. Since models can sometimes act like black boxes, meaning they generate outputs without clearly showing how each decision is made. Techniques such as Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) can break this down by showing which features influenced each alert, helping us to understand and confirm why something was flagged. Recent research also shows that using XAI can improve the transparency and practicality of IDS systems without affecting accuracy (Fatema, Dey, Anannya, Su & Mazumder, 2025). Together, these improvements would make hybrid IDS more reliable, adaptive and ready for modern cybersecurity challenges.

9.0 References

[1] (NIST), N. I. (2025). Framework for Improving Critical Infrastructure Cybersecurity (or relevant CSF version). Retrieved from NIST: https://www.nist.gov/cyberframework

[2] Alashjaee, A. M. (2025). Retrieved from Scientific Reports: https://doi.org/10.1038/s41598-025-20047-0

[3] Aldughayfiq, B., Ashfaq, F., Jhanjhi, N. Z., & Humayun, M. (2023). A Deep Learning Approach for Atrial Fibrillation Classification Using Multi-Feature Time Series Data from ECG and PPG. *Diagnostics*, *13*(14), 2442. https://doi.org/10.3390/diagnostics13142442

[4] Almuayqil, S. N., Humayun, M., Jhanjhi, N. Z., Almufareh, M. F., & Javed, D. (2022). Framework for improved sentiment analysis via random minority oversampling for user tweet review classification. *Electronics*, *11*(19), 3058. https://doi.org/10.3390/electronics11193058

[5] Al-Musib, N. S., Al-Serhani, F. M., Humayun, M., & Jhanjhi, N. (2021). Business email compromise (BEC) attacks. *Materials Today Proceedings*, *81*, 497–503. https://doi.org/10.1016/j.matpr.2021.03.647

[6] Alotaibi, A., & Rassam, M. A. (2023). *Adversarial machine learning attacks against intrusion detection systems: A survey on strategies and defense. Future Internet, 15*(2), 62. https://www.mdpi.com/1999-5903/15/2/62

[7] Alotaibi, A., & Rassam, M. A. (2023). *Adversarial machine learning attacks against intrusion detection systems: A survey on strategies and defense*. *Future Internet, 15*(2), 62. https://www.mdpi.com/1999-5903/15/2/62

[8] Alwakid, G., Gouda, W., Humayun, M., & Jhanjhi, N. Z. (2023). Deep learning-enhanced diabetic retinopathy image classification. *Digital Health*, *9*, 20552076231194942. https://doi.org/10.1177/20552076231194942

[9] Annadurai, C., Nelson, I., Devi, K., Manikandan, R., Jhanjhi, N., Masud, M., & Sheikh, A. (2022). Biometric Authentication-Based Intrusion Detection using Artificial intelligence internet of things in smart city. *Energies*, *15*(19), 7430. https://doi.org/10.3390/en15197430

[10] B.V, E. (2021). Complex Adaptive Systems Conference Theme: Big Data, IoT. Network Intrusion Detection System using Deep Learning, pp. 239-247.

[11] Brohi, S. N., Jhanjhi, N., Brohi, N. N., & Brohi, M. N. (2020). Key Applications of State-of-the-Art Technologies to Mitigate and Eliminate COVID-19.pdf. *TechRxiv*. https://doi.org/10.36227/techrxiv.12115596.v1

[12] Chaudhary, M., Gaur, L., Jhanjhi, N., Masud, M., & Aljahdali, S. (2022). Envisaging employee churn using MCDM and machine learning. *Intelligent Automation & Soft Computing*, *33*(2), 1009–1024. https://doi.org/10.32604/iasc.2022.023417

[13] Choi, W.-H., & Kim, J. (2024). Unsupervised learning approach for anomaly detection in industrial control systems. *Applied System Innovation, 7*(2), 18. https://www.mdpi.com/2571-5577/7/2/18

[14] Evidently AI. (2025). *Classification threshold*. Evidently AI. https://www.evidentlyai.com/classification-metrics/classification-threshold

[15] Fatema, K., Dey, S. K., Anannya, M., Su, C., & Mazumder, R. (2025). *Federated XAI IDS: An Explainable and Safeguarding Privacy Approach to Detect Intrusion Combining Federated Learning and SHAP. Future Internet, 17*(6), 234. https://doi.org/10.3390/fi17060234

[16] Fatima-tuz-Zahra, N. Jhanjhi, S. N. Brohi, N. A. Malik and M. Humayun, "Proposing a Hybrid RPL Protocol for Rank and Wormhole Attack Mitigation using Machine Learning," 2020 2nd International Conference on Computer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, 2020, pp. 1-6, doi: 10.1109/ICCIS49240.2020.9257607.

[17] GeeksforGeeks. (2025). How to use classification threshold to balance precision and recall

[18] GeeksforGeeks. (2025, July 23). *Defensive distillation*. In *Deep Learning*. Retrieved from https://www.geeksforgeeks.org/deep-learning/defensive-distillation/

[19] Guo, C., Li, X., Cheng, J., Yang, S., & Gong, H. (2025). *Continual Learning for Intrusion Detection Under Evolving Network Threats. Future Internet, 17*(10), 456. https://doi.org/10.3390/fi17100456

[20] Hendaoui, F., Meddeb, R., Trabelsi, L., Ferchichi, A., & Ahmed, R. (2025). FLADEN: Federated Learning for Anomaly DEtection in IoT Networks. Computers & Security.

[21] Humayun, M., Jhanjhi, N. Z., Niazi, M., Amsaad, F., & Masood, I. (2022). Securing drug distribution systems from tampering using blockchain. Electronics, 11(8), 1195.

[22] Kaushik, Sunil. (2025). Robust machine learning based Intrusion detection system using simple statistical techniques in feature selection. Scientific Reports, 3970.

[23] Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: Techniques, datasets and challenges. *Cybersecurity, 2*(1), 1–22.https://cybersecurity.springeropen.com/articles/10.1186/s42400-019-0038-7

[24] Maklin, C. (2022, May 14). *Synthetic Minority Over-sampling TEchnique (SMOTE)*. *Medium*. Retrieved from https://medium.com/@corymaklin/synthetic-minority-over-sampling-technique-smote-7d419696b88c

[25] Milvus. (n.d.). *What is adversarial training in deep learning?* Retrieved November 10, 2025, from https://milvus.io/ai-quick-reference/what-is-adversarial-training-in-deep-learning

[26] Mohammed, M. S., & Talib, H. A. (2024). Using Machine Learning Algorithms in Intrusion Detection Systems: A review. Tikrit Journal of Pure Science.

[27] Nicho, M., Cusack, B., Girija, S., & Arachchilage, N. (2024). Evaluating the cost of classifier discrimination choices for IoT sensor attack detection. International Journal of Computers and Applications, 956.

[28] Number of Internet of Things (IoT) connections worldwide from 2022 to 2023, with forecasts from 2024 to 2034. (19, November 2025). Retrieved from Statista: https://www.scribbr.com/citation/generator/folders/2o5lCHHg9bCsLILAvqusx2/lists/6Y59Cyg6muJeFxNsESmTfJ/

[29] ReasonLab. (2023). What is Behaviour-based Detection? Retrieved from reasonlabs: https://cyberpedia.reasonlabs.com/EN/behaviour-based%20detection.html

[30] Reddy, K. Y., & Lingam, G. S. (2024). Artificial intelligence in intrusion detection systems: Trends, frameworks, and future directions for cybersecurity. *International Journal of Intelligent*

Systems and Applications in Engineering, 12(17s), 949–. https://ijisae.org/index.php/IJISAE/article/view/7689

[31] Riskhan, B., Roua, A., Mahaba, B. S., Uswa, D., Gheisari, M., & Sindiramutty, S. R. (2025). Enhancing Diagnostic Accuracy for Breast Cancer Using Classical-Quantum Hybrid and Transfer Learning Technique. *Journal of Soft Computing and Data Mining, 6*(2), 41–56. https://penerbit.uthm.edu.my/ojs/index.php/jscdm/article/view/22061

[32] Salehpour, A., Balafar, M. A., & So, A. (2025). An optimized intrusion detection system for resource-constrained IoMT environments: enhancing security through efficient feature selection and classification. The Journal of Supercomputing.

[33] Sanjaya, F. N., Balakrishnan, S., Sindiramutty, S. R., & Narputro, P. (2025). XAI Technology in EEG Signal Based Disease Detection Models. *2025 International Conference on Metaverse and Current Trends in Computing (ICMCTC)*, 1–7. https://doi.org/10.1109/icmctc62214.2025.11196271

[34] Sciforce. (2022, September 7). *Adversarial attacks explained (And how to defend ML models against them)*. Medium. Retrieved from https://medium.com/sciforce/adversarial-attacks-explained-and-how-to-defend-ml-models-against-them-d76f7d013b18

[35] Sindiramutty, S. R., Jhanjhi, N. Z., Tan, C. E., Tee, W. J., Lau, S. P., Jazri, H., Ray, S. K., & Zaheer, M. A. (2024b). IoT and AI-Based Smart Solutions for the Agriculture Industry. In *Advances in computational intelligence and robotics book series* (pp. 317–351). https://doi.org/10.4018/978-1-6684-6361-1.ch012

[36] Sindiramutty, S. R., Jhanjhi, N., Tan, C. E., Lau, S. P., Muniandy, L., Gharib, A. H., Ashraf, H., & Murugesan, R. K. (2024a). Industry 4.0. In *Advances in logistics, operations, and management science book series* (pp. 342–405). https://doi.org/10.4018/979-8-3693-1363-3.ch013

[37] Slaughter, J., Gutierrez, F., & Imano, S. (2023, June 8). *MOVEit Transfer critical vulnerability (CVE-2023-34362) exploited as a 0-day.* Fortinet. https://www.fortinet.com/blog/threat-research/moveit-transfer-critical-vulnerability-cve-2023-34362-exploited-as-a-0-day

[38] Suricata vs Snort. (n.d.). Retrieved from Stamus Networks: https://www.stamus-networks.com/suricata-vs-snort

[39] Volk, O., & Singer, G. (2021, November 14). *Adaptive cost-sensitive learning in neural networks for misclassification cost problems* (arXiv:2111.07382). ArXiv. https://arxiv.org/abs/2111.07382

[40] Wang Z, C. H. (2023). A lightweight intrusion detection method for IoT based on. 34.

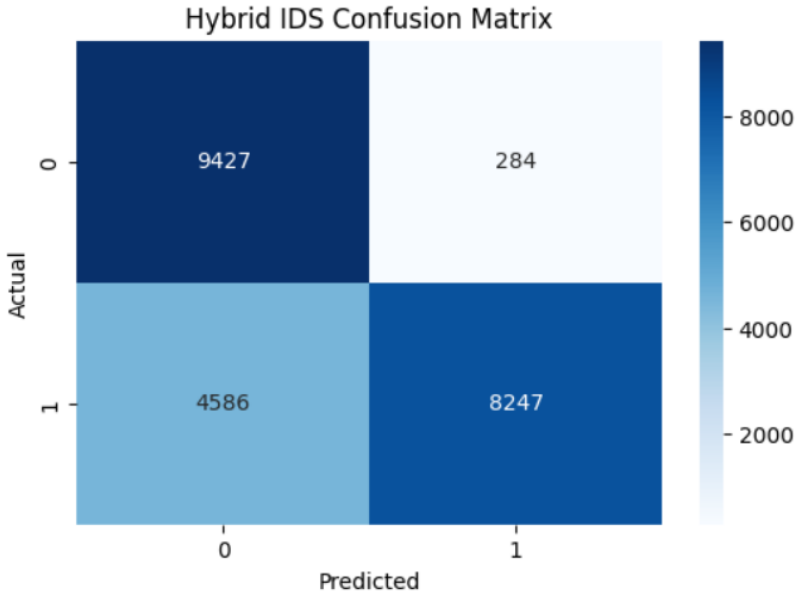
[41] Weiqi, X., Hooi, S. T. C., Sindiramutty, S. R. a. L., Asirvatham, D. a. L., Kumar, D., & Verma, S. (2024). Surface Anomaly Detection Using Machine Learning Technique. *2024 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC)*, 1–7. https://doi.org/10.1109/etncc63262.2024.10767562

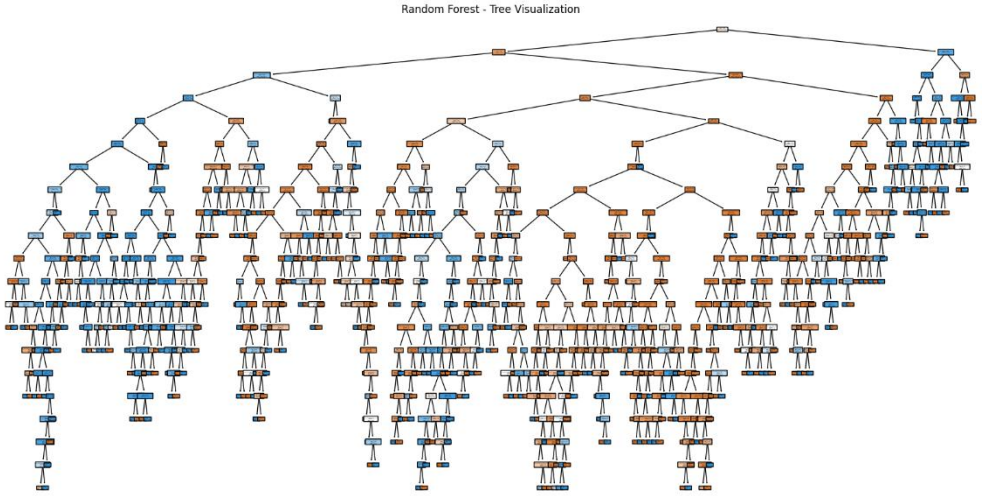
[42] What are the three types of IDS? (n.d.). Retrieved from Stamus Network: https://www.stamus-networks.com/blog/what-are-the-three-types-of-ids

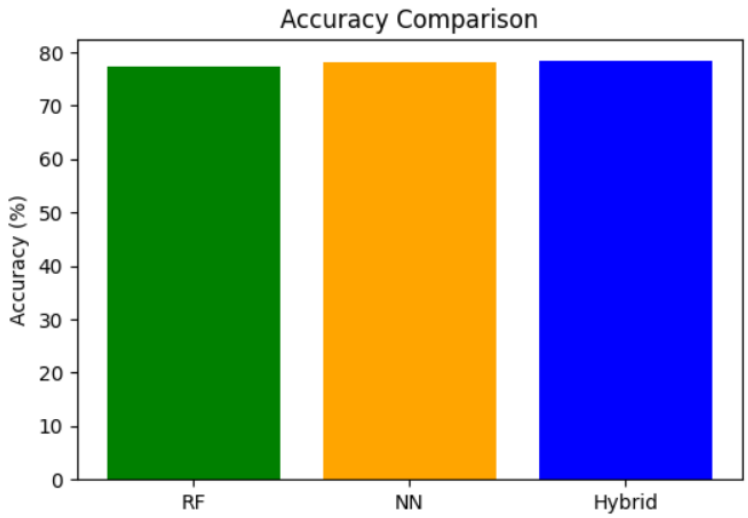
[43] WJARR. (2025). *Hybrid network intrusion detection system combining signature-based and anomaly-based detection. World Journal of Advanced Research and Reviews, 25*(2), 507–515. https://journalwjarr.com/sites/default/files/fulltext_pdf/WJARR-2025-0367.pdf

[44] Ying, X., Murugesan, R. K., Sindiramutty, S. R., Wei, G. W., Balakrishnan, S., Kumar, D., & Verma, S. (2024). Scene Text Recognition using Deep Learning Techniques. *2024 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC)*, 1–9. https://doi.org/10.1109/etncc63262.2024.10767484

11.0 Appendices






Accuracy Comparison
Accuracy (%)
0
10
20
30
40
50
60
70
80
RF
NN
Hybrid