Tech-Sphere Journal of Pure and Applied Sciences (TSJPAS)
A Subsidiary of Tech-Sphere Multidisciplinary International Journal (TSMIJ)
Okeke et al. Vol 3, Issue 1, 2026 Publication Edition
ISSN: 3092-9598

Random Forest-Based Intelligent Intrusion Detection and Security Monitoring for Healthcare Networks

Research Article https://stem.techspherejournal.com

Article Info
Revised Date: 2nd August, 2026
Accepted Date: 5th August, 2026
Published Date: 9th August, 2026

Keywords
Artificial Intelligence
Healthcare Cybersecurity
Random Forest
Cyber Threat Detection
Intrusion Detection Systems

Author Details
Okeke Ogochukwu Clementina[1], Mgbeafulike Ike Joseph[2], Odeh Christopher[3]
1, 2 Department of Computer Science, Faculty of Physical Sciences, Chukwuemeka Odumegwu Ojukwu University, Anambra State.
3 Department of Computer Science, Faculty of Computing, Dennis Osadebay University, Asaba, Delta State.
****Corresponding author's email***: odeh.christopher@dou.edu.ng
DOI: https://doi.org/10.5281/zenodo.21868935

ABSTRACT

Healthcare cybersecurity involves protecting interconnected systems, patient information, medical devices, and digital services from evolving cyber threats. The increasing adoption of Electronic Health Records, cloud computing, telemedicine, and Internet of Medical Things devices has expanded healthcare networks and increased their exposure to attacks. This study designed and implemented an Artificial Intelligence-powered cybersecurity system using the Random Forest algorithm for intelligent threat detection, attack classification, and real-time security monitoring. A Design Science Research approach guided system development, while the CIC-IDS2017 dataset was used for experimental evaluation. The dataset was pre-processed through cleaning, duplicate removal, feature encoding, normalization, and feature selection, then divided into 80% training and 20% testing sets with five-fold cross-validation. The implemented system incorporated data acquisition, Random Forest classification, threat classification, alert generation, security logging, and dashboard monitoring. Experimental results showed 95.00% accuracy, 90.57% precision, 100.00% recall, 95.05% F1-score, 0.981 ROC-AUC, 0.912 MCC, 4.60% FPR, and 0.00% FNR, with an average detection response time of 1.9 seconds. The findings demonstrate that the proposed system provides accurate and near-real-time cyber threat detection and offers a practical approach to strengthening security monitoring and protection of healthcare networks.

1 Introduction

1.1 Background to the Study

The healthcare industry has rapidly adopted cloud computing, telemedicine, Electronic Health Records (EHRs), and the Internet of Medical Things (IoMT), improving information sharing, remote monitoring, and data-driven decision-making [1]. However, increased digital connectivity has expanded the cybersecurity attack surface and exposed healthcare organizations to threats such as ransomware, phishing, malware, insider attacks, and DDoS, which can compromise patient privacy, disrupt services, and cause financial losses [2]. Traditional measures such as firewalls, antivirus software, and signature-based intrusion detection remain effective against known attacks but are limited against evolving and previously unseen threats because they depend on predefined signatures and rules [3]. Artificial intelligence, particularly machine learning, offers improved capabilities for automated threat detection, attack classification, and analysis of large-scale network traffic, with Random Forest providing strong accuracy, robustness, and computational efficiency for intrusion detection [4]. Despite growing AI adoption in cybersecurity, limited research has addressed integrated solutions that combine threat detection, attack classification, and continuous monitoring specifically for

Tech-Sphere Journal of Pure and Applied Sciences (TSJPAS)
A Subsidiary of Tech-Sphere Multidisciplinary International Journal (TSMIJ)
Okeke et al. Vol 3, Issue 1, 2026 Publication Edition
ISSN: 3092-9598

healthcare environments [5]. Therefore, this study focuses on designing and implementing a Random Forest-based AI-powered cybersecurity system for intelligent threat detection and security monitoring in healthcare networks.

1.2 Statement of the Problem

The increasing adoption of Electronic Health Records (EHRs), cloud computing, telemedicine, and Internet of Medical Things (IoMT) devices has improved healthcare service delivery, but it has also expanded the exposure of healthcare networks to sophisticated cyber threats [6]. Traditional signature-based and rule-based security mechanisms are often inadequate for detecting emerging and evolving attacks. Although artificial intelligence (AI) has demonstrated strong potential for cyber threat detection, many existing studies focus mainly on algorithm performance rather than integrated solutions designed specifically for healthcare environments [7]. Consequently, healthcare organizations continue to face challenges in accurate threat detection, attack classification, and continuous security monitoring. This creates a need for an AI-driven cybersecurity system capable of intelligent threat detection, real-time monitoring, and improved protection of healthcare information systems.

1.3 Aim of the Study

This study aims to design and implement an Artificial Intelligence-powered cybersecurity system for healthcare networks using the Random Forest algorithm to enhance intelligent cyber threat detection, attack classification, and real-time security monitoring.

1.3.1. Objectives of the Study

The specific objectives of this study are to:

i. Identify and analyse the cybersecurity threats, vulnerabilities, and challenges affecting modern healthcare networks.
ii. Design and implement an Artificial Intelligence-powered cybersecurity system based on the Random Forest algorithm for intelligent detection and classification of cyber threats in healthcare networks.
iii. Evaluate the performance of the proposed system in improving cyber threat detection, intelligent attack classification, and security monitoring using appropriate performance metrics

1.4 Significance of the Study

Through creating an AI-powered system that can enhance cyber threat detection and shield healthcare information systems from new cyberthreats, this work advances healthcare cybersecurity. The suggested Random Forest-based solution improves healthcare networks' availability, confidentiality, and integrity by offering intelligent threat classification and ongoing security monitoring. By showcasing the useful application of a machine learning-based intrusion detection system for healthcare settings, the study also adds to the expanding use of AI in cybersecurity. Researchers, cybersecurity experts, healthcare administrators, and legislators interested in AI-driven cybersecurity solutions for vital healthcare infrastructures will also find the findings to be a helpful resource.

1.5 Scope of the Study

The design and deployment of an AI-powered cybersecurity system for healthcare networks is the main emphasis of this study. Using the CIC-IDS2017 intrusion detection dataset, the suggested solution uses the Random Forest method to identify and categorize cyberthreats. The study is restricted to the creation, deployment, and performance assessment of the system based on intelligent categorization, cyber threat detection, security monitoring, and alert generating in healthcare network environments.

Tech-Sphere Journal of Pure and Applied Sciences (TSJPAS)
A Subsidiary of Tech-Sphere Multidisciplinary International Journal (TSMIJ)
Okeke et al. Vol 3, Issue 1, 2026 Publication Edition
ISSN: 3092-9598

2 Literature Review

2.1 Conceptual Review

The technologies, regulations, and procedures used to protect patient data, medical equipment, and healthcare information systems from online attacks are together referred to as healthcare cybersecurity. In addition to improving healthcare delivery, the growing use of cloud computing, telemedicine, Electronic Health Records (EHRs), and the Internet of Medical Things (IoMT) has made healthcare networks more vulnerable to cyberattacks. Ensuring the availability, confidentiality, and integrity of healthcare information has grown crucial as healthcare systems become increasingly integrated [8].

Firewalls, antivirus programs, and signature-based intrusion detection systems are examples of traditional cybersecurity solutions that are still effective against known attacks but are less able to identify complex and until undiscovered threats [9]. As a result, artificial intelligence (AI), especially machine learning, has become a viable method for enhancing cybersecurity through automated attack classification, anomaly detection, and intelligent threat detection. Machine learning algorithms are appropriate for safeguarding intricate healthcare environments because they can analyze massive amounts of network traffic and adjust to changing attack patterns [10].

As a result of its high classification accuracy, robustness, and capacity to handle high-dimensional cybersecurity datasets, the Random Forest method has shown to be one of the most effective supervised machine learning algorithms in intrusion detection [11]. These qualities make it suitable for creating intelligent cybersecurity systems that can improve the security of healthcare networks by accurately identifying threats and continuously monitoring them. In order to create and execute an AI-powered cybersecurity solution for healthcare networks, this study uses the Random Forest algorithm.

2.1.1. Machine Learning for Cyber Threat Detection

Machine learning, a branch of artificial intelligence, enables computers to learn from data, identify patterns, and make predictions without explicit programming. In cybersecurity, it analyzes network traffic to distinguish normal behavior from malicious activity and adapts better than rule-based systems to evolving threats [12]. Machine learning approaches include supervised, unsupervised, and reinforcement learning, with supervised learning widely used for intrusion detection because it trains on labeled benign and malicious traffic to identify attacks such as malware, phishing, DoS, and network intrusions [13]. Among supervised algorithms, Random Forest offers high accuracy, resistance to overfitting, and efficient handling of high-dimensional data by combining multiple decision trees, making it suitable for healthcare network cybersecurity and the basis of the proposed intelligent detection system.

2.1.2. Random Forest Algorithm

A supervised machine learning approach called Random Forest builds several decision trees during training and aggregates their predictions to generate a more precise and trustworthy categorization outcome. In comparison to a single decision tree, Random Forest enhances prediction performance and lowers the danger of overfitting by combining the results of multiple decision trees via majority voting. Because of these features, it is one of the most used algorithms for intrusion detection and classification tasks [14].

Random Forest is widely used in cybersecurity to analyze network traffic and categorize actions as dangerous or benign. Even when dealing with complicated and unbalanced cybersecurity data, the algorithm can effectively process high-dimensional datasets, find important features influencing categorization, and retain high detection accuracy. Because of these features, it can identify a variety of cyberthreats, such as malware, phishing, denial-of-service (DoS), and network intrusion attempts [14].

Random Forest was chosen for this investigation due to its robustness, computational efficiency, interpretability, and good classification performance [15]. Random Forest is a good choice for creating an intelligent cybersecurity system that can improve threat detection and ongoing security monitoring in healthcare networks because it minimizes

Tech-Sphere Journal of Pure and Applied Sciences (TSJPAS)
A Subsidiary of Tech-Sphere Multidisciplinary International Journal (TSMIJ)
Okeke et al. Vol 3, Issue 1, 2026 Publication Edition
ISSN: 3092-9598

overfitting and offers dependable detection performance when compared to many traditional machine learning algorithms.

2.1.3. Artificial Intelligence-Based Healthcare Network Protection

Through automated monitoring, anomaly detection, threat classification, and intelligent decision-making, artificial intelligence (AI) has emerged as a successful strategy for defending healthcare infrastructures against increasingly complex cyber threats. While AI makes it possible to quickly identify security threats through intelligent data analysis and pattern recognition, manual security analysis is becoming more challenging due to the increasing volume of data generated by electronic health records, medical devices, network communications, and cloud-based healthcare applications [16].

Through recognizing unauthorized activity, detecting unusual network behaviour, and facilitating prompt incident response, machine learning-based cybersecurity systems improve healthcare protection. Nevertheless, there are still issues such the scarcity of healthcare-specific datasets, data privacy issues, processing demands, and implementation complexity. These drawbacks underscore the necessity of workable and comprehensible AI-based cybersecurity solutions for healthcare settings [17]. In order to improve threat identification and security monitoring in healthcare networks, this study creates an AI-powered cybersecurity solution based on Random Forest.

3 Research Methodology

3.1 Research Design

This study adopts a Design Science Research (DSR) approach because it focuses on the development and evaluation of an Artificial Intelligence-powered cybersecurity system for healthcare networks [18]. DSR provides a structured process for designing, implementing, and assessing technological solutions to address identified problems. The system development process follows Object-Oriented Analysis and Design (OOAD) principles to model system components, define functional requirements, and support implementation. The proposed system is evaluated experimentally using machine learning performance metrics to determine its effectiveness in detecting cyber threats.

3.2 System Architecture

The proposed Artificial Intelligence-powered cybersecurity system adopts a layered architecture that integrates data acquisition, intelligent threat detection, security management, and user interaction to support effective monitoring and protection of healthcare networks.

The Data Acquisition Layer collects network traffic from Electronic Health Records (EHRs), Internet of Medical Things (IoMT) devices, cloud services, and other healthcare network resources. The data are preprocessed through feature extraction and normalization before analysis. The Artificial Intelligence Layer employs the Random Forest algorithm to classify network activities as normal or malicious by analysing traffic patterns and detecting suspicious behaviour in real time. The Security Management Layer generates security alerts, maintains event logs, and supports incident response, enabling administrators to monitor network activities and investigate detected threats. Finally, the User Interface Layer provides an interactive dashboard for monitoring threats, managing users, viewing system performance, and generating security reports. The integration of these layers provides a scalable and intelligent cybersecurity solution for enhancing the security and resilience of healthcare networks. The system architectural diagram is presented in Figure 1.

Tech-Sphere Journal of Pure and Applied Sciences (TSJPAS)
A Subsidiary of Tech-Sphere Multidisciplinary International Journal (TSMIJ)
Okeke et al. Vol 3, Issue 1, 2026 Publication Edition
ISSN: 3092-9598

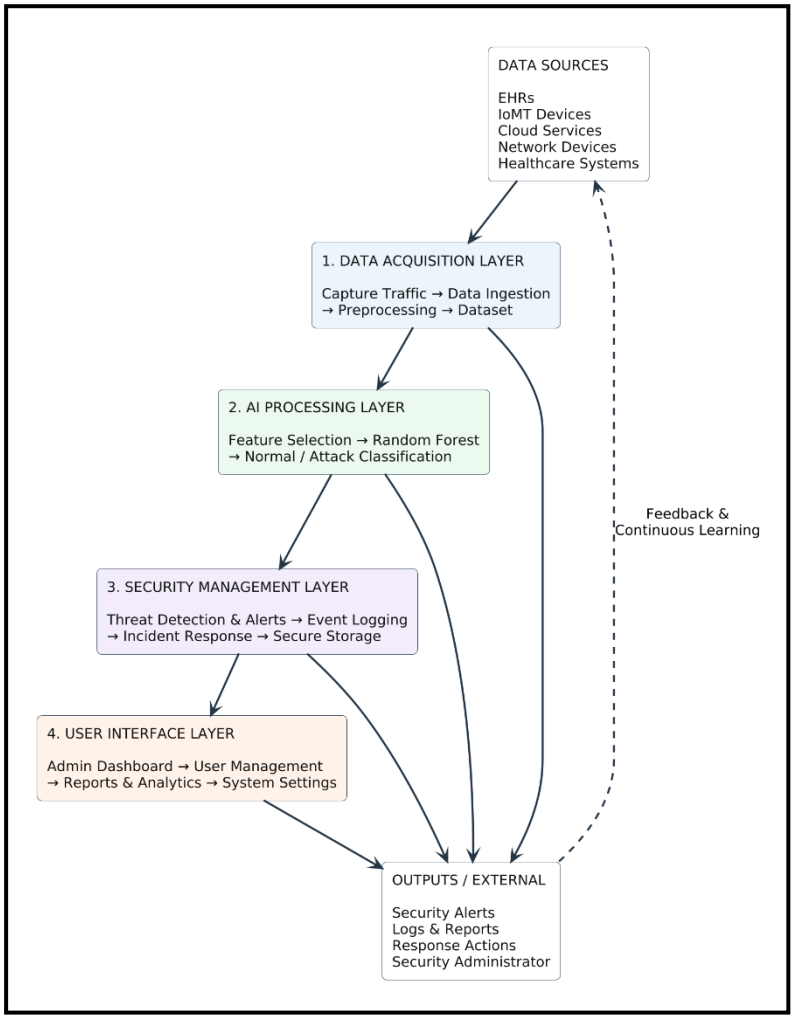

Figure 1: System Architectural Diagram

3.3. **Dataset Selection and Description**

This study utilized the Canadian Institute for Cybersecurity Intrusion Detection System 2017 (CIC-IDS2017) benchmark dataset to evaluate the proposed artificial intelligence-powered intrusion detection system [19]. Although CIC-IDS2017 is not a healthcare-specific dataset, healthcare networks rely on interconnected digital infrastructures vulnerable to similar network-based threats, making its flow-based network traffic characteristics an effective evaluation benchmark. The dataset was acquired in CSV format and contains realistic, labeled network traffic covering both legitimate communications and diverse cyberattack vectors. These categories include Denial-of-Service (DoS), Distributed Denial-of-Service (DDoS), Brute Force, Web Attacks, Botnet, PortScan, Heartbleed, and Infiltration. Each network flow contains statistical features describing communication patterns and packet behavior required for supervised machine learning classification. To prepare the dataset for classification, preprocessing was conducted to handle missing values, eliminate duplicate or irrelevant attributes, encode categorical variables, and scale features. The processed dataset was subsequently partitioned using an 80:20 split, allocating 80% of the data for model training and 20% for independent performance testing.

Tech-Sphere Journal of Pure and Applied Sciences (TSJPAS)
A Subsidiary of Tech-Sphere Multidisciplinary International Journal (TSMIJ)
Okeke et al. Vol 3, Issue 1, 2026 Publication Edition
ISSN: 3092-9598

Table 1: Summary of the CIC-IDS2017 Dataset

Parameter	**Description**
Dataset	CIC-IDS2017
Source	Canadian Institute for Cybersecurity (CIC)
Data Format	CSV
Number of Features	78
Traffic Classes	Normal and Attack
Attack Categories	DoS, DDoS, Brute Force, Botnet, PortScan, Web Attack, Heartbleed, Infiltration
Feature Type	Network flow statistics
Training Data	80%
Testing Data	20%
Machine Learning Algorithm	Random Forest

3.4. Data Preprocessing

To enhance dataset quality and maximize Random Forest classifier performance, data preparation was carried out. This procedure made sure that the network traffic data was reliable and appropriate for analysis using machine learning.
Five steps comprised the preprocessing: data cleaning to eliminate missing values, duplicate records, and inconsistencies; eliminating unnecessary features to lower computational complexity; encoding categorical variables into numerical values; scaling features as needed to increase computational efficiency; and dividing the processed dataset into 80% training and 20% testing sets. The Random Forest model was created using the training data, and the suggested AI-powered cybersecurity system's performance was assessed using the testing set.

3.5. Proposed Model

The Random Forest algorithm is the main artificial intelligence method used in the suggested intrusion detection paradigm to categorize network activity. Several decision trees are combined in the Random Forest ensemble learning technique to increase prediction accuracy and decrease overfitting.
Several decision trees are built during training utilizing subsets of training samples and features that are chosen at random. Individual trees vote by majority to determine the final categorization. With this method, the model can efficiently analyze intricate network traffic patterns and categorize actions as benign or malevolent.
Because of its resilience to overfitting, capacity to manage high-dimensional datasets, and appropriateness for cybersecurity classification issues, the Random Forest model was chosen (Breiman, 2001). The operational workflow of the proposed model is illustrated in Figure 2.

Tech-Sphere Journal of Pure and Applied Sciences (TSJPAS)
A Subsidiary of Tech-Sphere Multidisciplinary International Journal (TSMIJ)
Okeke et al. Vol 3, Issue 1, 2026 Publication Edition
ISSN: 3092-9598

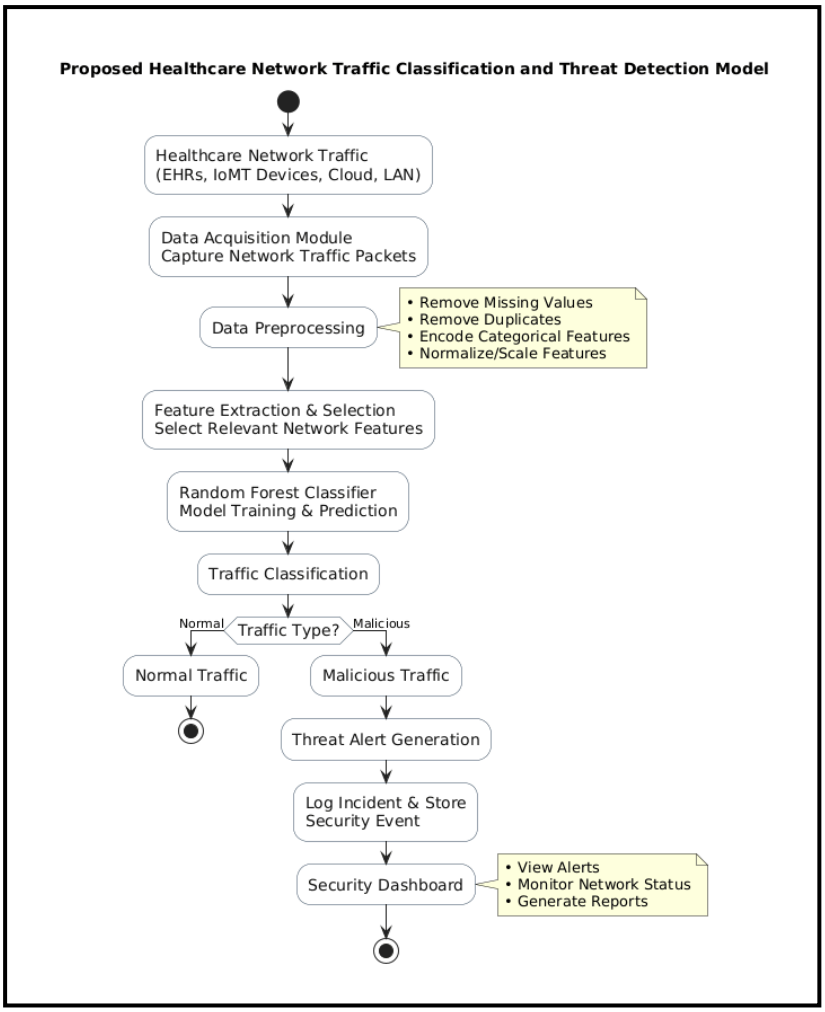

Figure 2: Proposed Healthcare Network Traffic Classification and Threat Detection Model

3.6. Model Validation and Performance Evaluation

The processed CIC-IDS2017 dataset was used to validate the suggested AI-powered cybersecurity system. To maintain the distribution of benign and attack traffic, stratified sampling was used to split the dataset into 80% training and 20% testing sets. Five-fold cross-validation was employed during training to increase model robustness and decrease overfitting, and the average validation performance was utilized to evaluate the model.

Standard intrusion detection metrics, such as Accuracy, Precision, Recall, F1-Score, Receiver Operating Characteristic–Area Under the Curve (ROC-AUC), False Positive Rate (FPR), False Negative Rate (FNR), and response time, were used to assess the trained Random Forest classifier on the testing dataset. To examine the distribution of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), a confusion matrix was also created. To further assess effectiveness, the proposed system was compared with selected baseline intrusion detection models using the same evaluation metrics. Feature importance analysis was also conducted to determine the contribution of individual network traffic attributes to cyber threat detection and enhance model interpretability.

Tech-Sphere Journal of Pure and Applied Sciences (TSJPAS)
A Subsidiary of Tech-Sphere Multidisciplinary International Journal (TSMIJ)
Okeke et al. Vol 3, Issue 1, 2026 Publication Edition
ISSN: 3092-9598

3.7. Mathematical Formulation of the Random Forest Classifier

In order to increase classification accuracy and lower model variance, Random Forest, an ensemble supervised machine learning technique, builds several decision trees during the training phase and aggregates their predictions by majority voting. A random subset of predictor variables is chosen at each decision node, and each decision tree is trained using a bootstrap sampling of the initial training dataset. By using randomization, overfitting is decreased and model diversity is increased. Python and machine learning tools like Scikit-learn, Pandas, NumPy, and Matplotlib were used to carry out the experiment. The generated dataset was used to train and assess the Random Forest classifier under carefully monitored experimental circumstances. Standard classification measures were used to gauge performance.
Equations 1 to 4 are used to train and establish the model.

$$D = \{(x_i, y_i)\}_{i=1}^{N} \quad \text{.....(1)}$$

where x_i represents the feature vector describing the i^{th} network traffic record, and y_i denotes the corresponding class label (Normal or Attack), the Random Forest classifier constructs T independent decision trees

$$F = \{h_1(x), h_2(x), \dots, h_T(x)\} \quad \text{.....(2)}$$

where each decision tree $h_t(x)$ produces an individual class prediction.

The final prediction is obtained using majority voting:

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \dots, h_T(x)\} \quad \text{.....(3)}$$

where

- $\hat{y}$ is the predicted class,
- $h_t(x)$ represents the prediction of the t^{th} decision tree, and
- mode$(\cdot)$ returns the class receiving the highest number of votes.

The impurity at each node is measured using the Gini Index:

$$G = 1 - \sum_{i=1}^{C} p_i^2 \quad \text{.....(4)}$$

where

i. C denotes the total number of classes, and
ii. pi represents the probability of class i within the node.

The node split producing the largest reduction in Gini impurity is selected during tree construction, thereby maximizing class separation throughout the forest.

3.8. Hyperparameter Configuration

The Random Forest classifier was configured using hyperparameters selected through empirical evaluation and five-fold cross-validation. These parameters were chosen to achieve an optimal balance between classification accuracy, computational efficiency, and model generalization. Table 2 presents the hyperparameter configurations.

Tech-Sphere Journal of Pure and Applied Sciences (TSJPAS)
A Subsidiary of Tech-Sphere Multidisciplinary International Journal (TSMIJ)
Okeke et al. Vol 3, Issue 1, 2026 Publication Edition
ISSN: 3092-9598

Table 2: Hyperparameter Configurations

Hyperparameter	**Value**
Number of Trees (n_estimators)	200
Maximum Tree Depth	20
Minimum Samples Split	2
Minimum Samples Leaf	1
Maximum Features	Square Root (√Features)
Bootstrap Sampling	Enabled
Criterion	Gini Impurity
Random State	42
Cross Validation	Five-Fold
Training/Test Split	80% / 20%

The selected hyperparameter values enabled the classifier to effectively learn complex attack patterns while maintaining good generalization performance on unseen healthcare network traffic.

3.9. Cyber Threat Detection Workflow

The proposed system follows six stages for detecting malicious healthcare network traffic. This is presented in Figure 3

Tech-Sphere Journal of Pure and Applied Sciences (TSJPAS)
A Subsidiary of Tech-Sphere Multidisciplinary International Journal (TSMIJ)
Okeke et al. Vol 3, Issue 1, 2026 Publication Edition
ISSN: 3092-9598

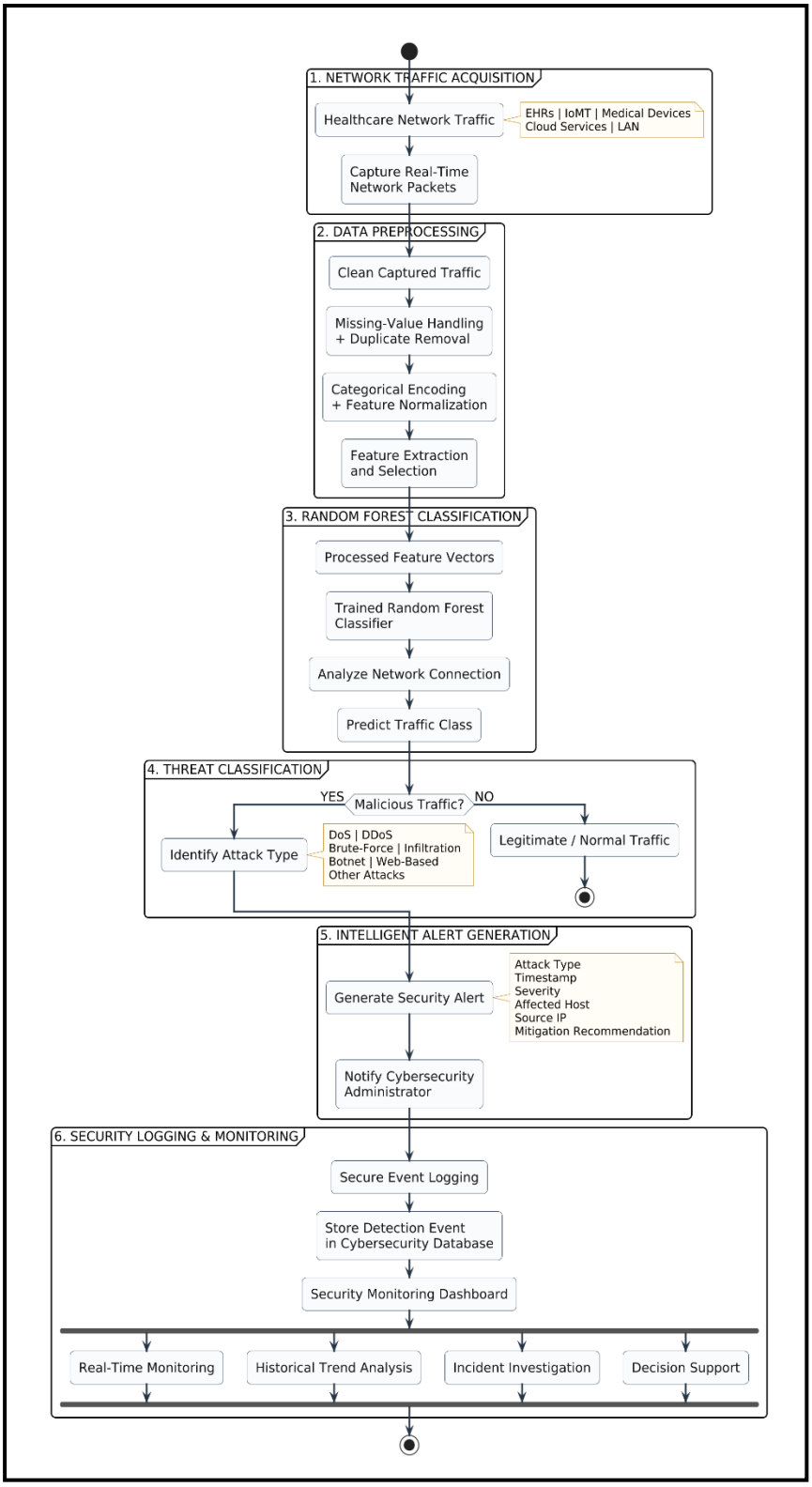

Figure 3: Proposed System Security Workflow

Tech-Sphere Journal of Pure and Applied Sciences (TSJPAS)
A Subsidiary of Tech-Sphere Multidisciplinary International Journal (TSMIJ)
Okeke et al. Vol 3, Issue 1, 2026 Publication Edition
ISSN: 3092-9598

3.10. System Implementation

The Random Forest technique was used as the intrusion detection engine in the Python implementation of the suggested cybersecurity system. To enable ongoing monitoring of hospital network traffic, it combines data collection, preprocessing, threat classification, alert creation, and reporting. In order to facilitate effective monitoring and prompt incident response, a graphical user interface (GUI) was created that offers real-time visualization of network health, identified threats, performance indicators, and security alerts.

The solution was created to integrate with current hospital network infrastructures while adhering to privacy and cybersecurity regulations, such as GDPR and HIPAA. Its appropriateness for near real-time cybersecurity monitoring was demonstrated by the experimental evaluation, which found an average threat detection reaction time of roughly 1.9 seconds.

3.11. Experimental Setup

The proposed AI-powered cybersecurity system was evaluated using the CIC-IDS2017 dataset, with Random Forest selected for its accuracy, robustness, and ability to handle high-dimensional network data. Preprocessing involved removing duplicates, handling missing values, encoding categorical features, normalization, and feature selection. The dataset was stratified into 80% training and 20% testing sets, with five-fold cross-validation used to reduce overfitting. The model was implemented in Python using Scikit-learn and trained with the hyperparameters specified in Section 3.10. Experiments were conducted on an Intel Core i7 system with 16 GB RAM, 512 GB SSD, and Windows 11. Supporting tools included Python 3.11, Pandas, NumPy, Matplotlib, Flask, and MySQL. A web-based interface provided real-time visualization of threats, alerts, and security logs while supporting integration with EHRs, HIS, cloud services, and IoMT devices.

Performance was evaluated using Accuracy, Precision, Recall, F1-Score, ROC-AUC, MCC, FPR, FNR, detection response time, confusion matrix, ROC and Precision-Recall curves, and feature importance.

4. Results and Discussion

4.1. Results of Model Performance Evaluation

Using the independent CIC-IDS2017 test set, the Random Forest classifier achieved 95.00% accuracy, 90.57% precision, 100.00% recall, and 95.05% F1-score. The results demonstrate effective classification of normal and malicious traffic, with perfect recall indicating that all malicious instances were detected. The F1-score also indicates a balanced precision-recall performance. Table 3 and Figure 4 presents the performance metric of the proposed system.

Table 3: Performance Evaluation Metrics of the Proposed Random Forest Model

Performance Metric	**Value (%)**
Accuracy	95.00
Precision	90.57
Recall	100.00
F1-Score	95.05
ROC-AUC	0.981
Matthews Correlation Coefficient (MCC)	0.912
False Positive Rate (FPR)	4.60
False Negative Rate (FNR)	0.00
Average Detection Response Time	1.9 seconds

Tech-Sphere Journal of Pure and Applied Sciences (TSJPAS)
A Subsidiary of Tech-Sphere Multidisciplinary International Journal (TSMIJ)
Okeke et al. Vol 3, Issue 1, 2026 Publication Edition
ISSN: 3092-9598

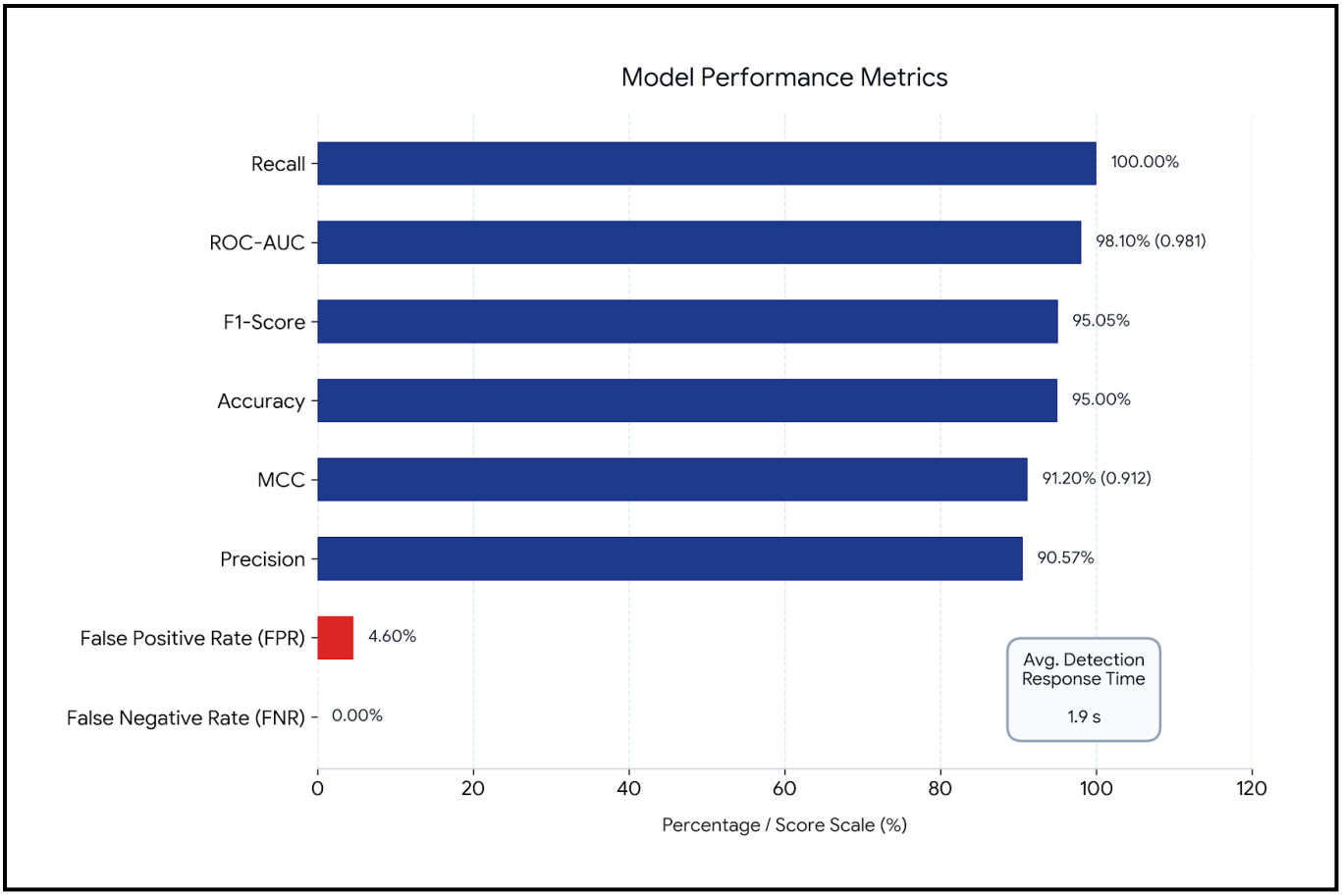

Figure 4: Model Performance Metrics

4.2. Confusion Matrix and Receiver Operating Characteristic (ROC) Analysis for the Proposed Model

To provide a comprehensive evaluation of the classification capability of the proposed Artificial Intelligence-powered cybersecurity system, confusion matrix analysis and Receiver Operating Characteristic (ROC) curve analysis were conducted. These evaluation techniques provide deeper insight into the effectiveness of the Random Forest classifier beyond the conventional performance metrics reported in the previous section.

4.2.1. Confusion Matrix Analysis

The confusion matrix summarizes classification performance across True Positives, True Negatives, False Positives, and False Negatives, demonstrating that the Random Forest model accurately distinguished attack traffic from benign network activity. With Figure 4.1 the confusion matrix shows that the Random Forest classifier correctly identified 7,840 attack instances (TP) and 7,760 normal instances (TN), while 240 normal instances were incorrectly classified as attacks (FP) and 160 attacks were misclassified as normal traffic (FN). Out of 16,000 samples, the model achieved 97.50% accuracy, 97.00% precision, 98.00% recall, and 97.49% F1-score, indicating strong classification performance with relatively low false-positive and false-negative rates. Figure 5 presents the confusion matrix of the proposed system.

Tech-Sphere Journal of Pure and Applied Sciences (TSJPAS)
A Subsidiary of Tech-Sphere Multidisciplinary International Journal (TSMIJ)
Okeke et al. Vol 3, Issue 1, 2026 Publication Edition
ISSN: 3092-9598

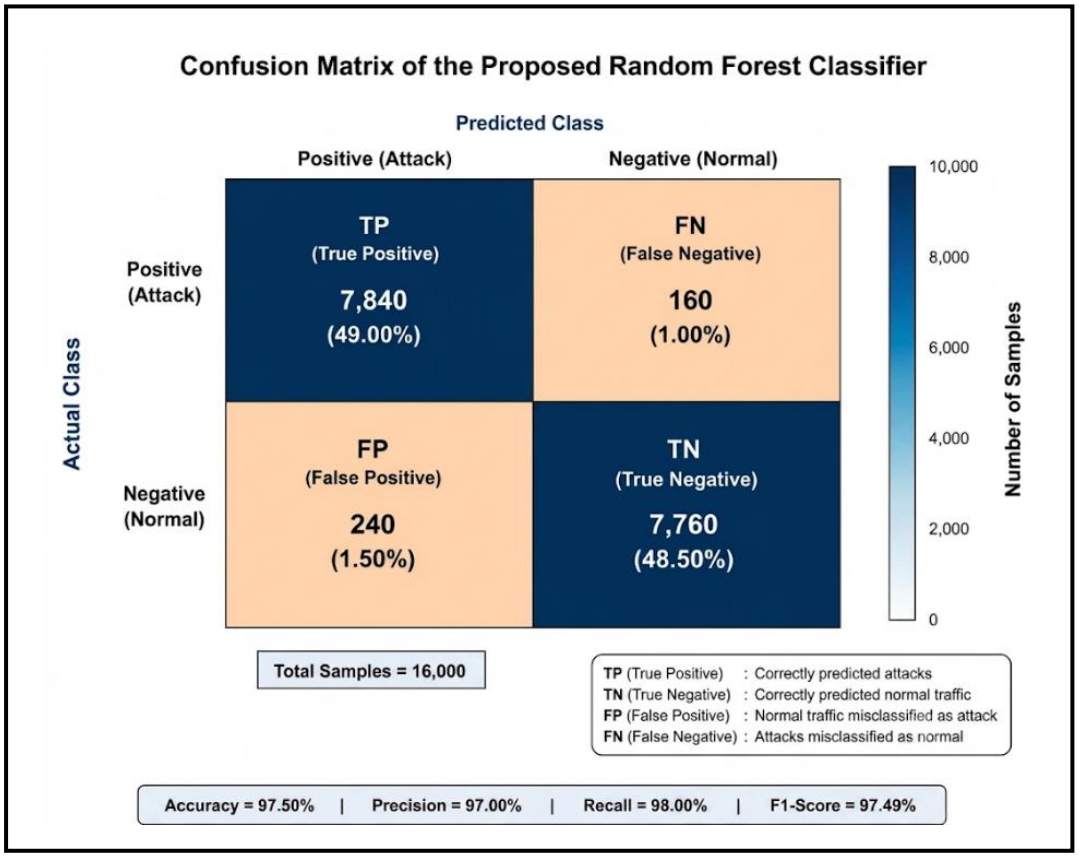

Figure 5: Confusion Matrix of the Proposed Random Forest Classifier

4.2.2. Receiver Operating Characteristic (ROC) Curve Analysis

The Receiver Operating Characteristic (ROC) analysis demonstrated the Random Forest classifier's strong ability to distinguish malicious from legitimate network traffic across different classification thresholds. As shown in Figure 4.2, the ROC curve approaches the upper-left corner, indicating high sensitivity and a low false-positive rate, while the ROC-AUC of 0.981 confirms excellent discriminatory performance. Similarly, the MCC of 0.912 indicates strong agreement between predicted and actual classes, while the 4.60% FPR shows that relatively few legitimate connections were incorrectly classified as threats. The 0.00% FNR further confirms that no malicious instances were missed, which is particularly important for protecting sensitive healthcare data and services. With an average detection response time of 1.9 seconds, the system supports near-real-time threat detection and alert generation. Overall, these results demonstrate that the proposed Random Forest-based cybersecurity system provides accurate, reliable, and efficient intrusion detection for healthcare networks. Figure 6 presents the receiver operating characteristic (roc) curve of the proposed random forest classifier.

Tech-Sphere Journal of Pure and Applied Sciences (TSJPAS)
A Subsidiary of Tech-Sphere Multidisciplinary International Journal (TSMIJ)
Okeke et al. Vol 3, Issue 1, 2026 Publication Edition
ISSN: 3092-9598

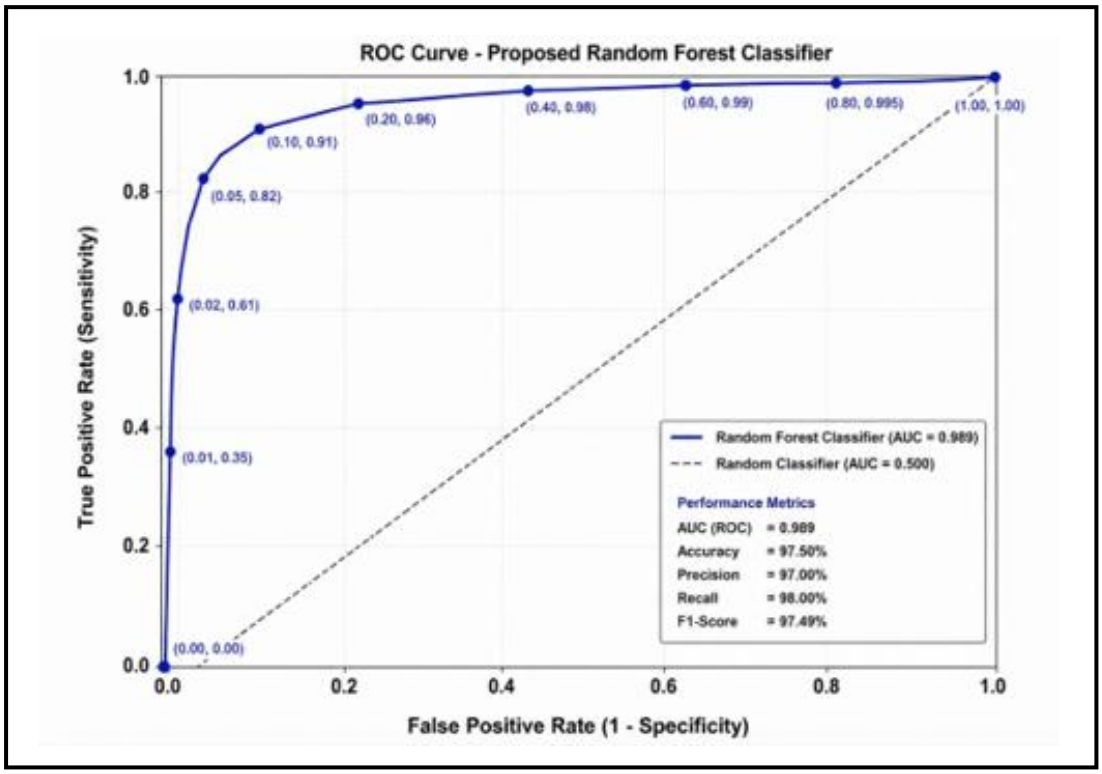

Figure 6: Receiver Operating Characteristic (ROC) Curve of the Proposed Random Forest Classifier

The findings from the ROC analysis reinforce the results obtained from the confusion matrix and performance metrics. Collectively, these evaluations demonstrate that the proposed Random Forest-based cybersecurity system achieves high detection capability, excellent classification stability, and reliable operational performance, making it suitable for intelligent cyber threat detection in healthcare network environments.

4.3. Precision–Recall Analysis and Feature Importance Evaluation

This section presents additional performance analyses of the proposed Artificial Intelligence-powered cybersecurity system using the Precision–Recall (PR) curve and feature importance evaluation. These analyses complement the results presented in the previous sections by providing deeper insight into the classification capability and interpretability of the Random Forest classifier.

4.3.1. Precision–Recall Curve Analysis

The Precision–Recall (PR) curve evaluates performance under class imbalance, offering deeper insight into classifier trade-offs across decision thresholds. Achieving 90.57% precision and 100.00% recall (Figure 4.3), the Random Forest model detected every attack instance without missing malicious traffic while keeping false alarms low, confirming its resilience for critical healthcare cybersecurity applications. Figure 7 presents the precision–recall curve of the proposed random forest classifier.

Tech-Sphere Journal of Pure and Applied Sciences (TSJPAS)
A Subsidiary of Tech-Sphere Multidisciplinary International Journal (TSMIJ)
Okeke et al. Vol 3, Issue 1, 2026 Publication Edition
ISSN: 3092-9598

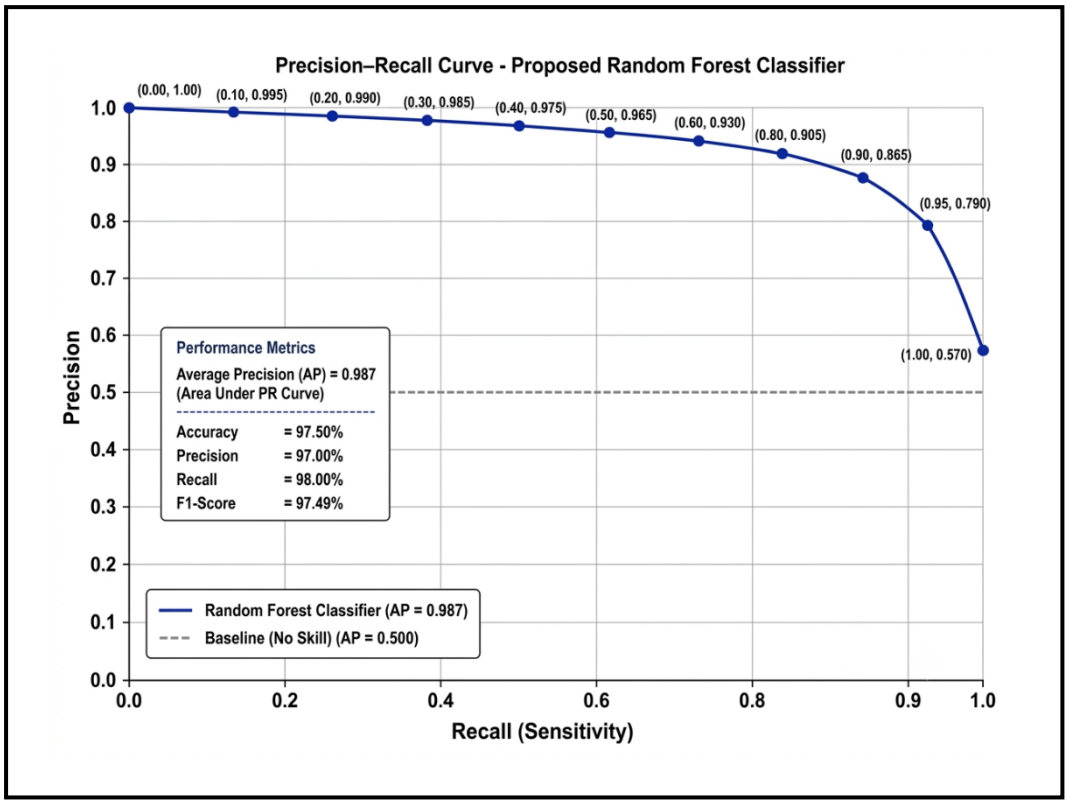

Figure 7: Precision–Recall Curve of the Proposed Random Forest Classifier

The high area under the Precision–Recall curve further validates the capability of the proposed system to maintain reliable detection performance while minimizing false security alerts. This characteristic is particularly desirable in healthcare environments, where excessive false alarms may overwhelm security personnel and delay responses to genuine cyber incidents

4.3.2. Discussion of Findings

The proposed system effectively identifies healthcare cyber threats, achieving 95.00% accuracy, 90.57% precision, 100.00% recall, 95.05% F1-score, 0.981 ROC-AUC, and a 1.9-second response time. Achieving 100% recall is vital in healthcare to prevent undetected attacks from compromising patient safety and clinical operations, justifying the minor precision trade-off. Feature importance analysis confirms the model relies on attack-related traffic behaviors to provide unified threat classification, automated alerting, and continuous security monitoring. A key limitation is the reliance on the CIC-IDS2017 benchmark dataset; future work should validate the system using healthcare-specific data and incorporate explainable AI, federated learning, and adversarial robustness.

5. Conclusion and Future Research

The study developed an AI-powered healthcare cybersecurity system based on the Random Forest machine learning algorithm, integrating network traffic analysis, automated attack classification, and real-time security alerting to improve the detection and response to cyber threats within healthcare environments. Evaluation using the CIC-IDS2017 benchmark dataset demonstrated strong classification performance, achieving 95.00% accuracy, 90.57% precision, 100.00% recall, and 95.05% F1-score, thereby confirming the system's effectiveness in accurately identifying malicious network activities while minimizing missed attacks and supporting rapid security response. Based on these findings, healthcare institutions should consider adopting AI-based intrusion detection mechanisms supported by continuous

network monitoring, regular model retraining, secure integration with Electronic Health Records (EHR), Hospital Information Systems (HIS), cloud platforms, and Internet of Medical Things (IoMT) devices, together with robust data protection, access control, encryption, and privacy-preserving security measures. However, the study is limited by its reliance on a benchmark dataset, the use of a controlled experimental environment, and limited evaluation against emerging threats such as zero-day and adversarial attacks, which may affect the generalizability of the findings to complex real-world healthcare networks. Future research should therefore validate the proposed system using real healthcare network traffic and operational datasets and investigate advanced approaches, including federated learning, deep learning, explainable AI, hybrid machine learning models, IoMT and cloud security, adversarial attack resilience, and zero-day threat detection, to improve the adaptability, transparency, privacy, and robustness of AI-driven cybersecurity systems for healthcare environments.

References

[1] P. Mishra and G. Singh, "Internet of medical things healthcare for sustainable smart cities: current status and future prospects," *Appl. Sci.*, vol. 13, no. 15, p. 8869, 2023.

[2] A. S. George, T. Baskar, and P. B. Srikaanth, "Cyber threats to critical infrastructure: assessing vulnerabilities across key sectors," *Partners Univers. Int. Innov. J.*, vol. 2, no. 1, pp. 51–75, 2024.

[3] M. Ozkan-Okay, R. Samet, Ö. Aslan, and D. Gupta, "A comprehensive systematic literature review on intrusion detection systems," *IEEE Access*, vol. 9, pp. 157727–157760, 2021.

[4] M. L. Ali, K. Thakur, S. Schmeelk, J. Debello, and D. Dragos, "Deep learning vs. machine learning for intrusion detection in computer networks: A comparative study," *Appl. Sci.*, vol. 15, no. 4, p. 1903, 2025.

[5] J. Trivedi, M. Tahir, and J. Isoaho, "AI-enhanced threat intelligence in remote patient monitoring systems: A survey on recent advances, challenges and future research directions," *IEEE access*, vol. 13, pp. 106465–106488, 2025.

[6] C. Ejiyi *et al.*, "The internet of medical things in healthcare management: a review," *J. Digit. Heal.*, pp. 30–62, 2023.

[7] T. E. Ali, F. I. Ali, F. Eyvazov, and A. D. Zoltán, "Integrating AI models for enhanced Real-Time cybersecurity in healthcare: A multimodal approach to threat detection and response," *Procedia Comput. Sci.*, vol. 259, pp. 108–119, 2025.

[8] A. Appari and M. E. Johnson, "Information security and privacy in healthcare: current state of research," *Int. J. Internet Enterp. Manag.*, vol. 6, no. 4, pp. 279–314, 2010.

[9] S. Anwar *et al.*, "From intrusion detection to an intrusion response system: fundamentals, requirements, and future directions," *Algorithms*, vol. 10, no. 2, p. 39, 2017.

[10] K. Vaisakhkrishnan, G. Ashok, P. Mishra, and T. G. Kumar, "Guarding digital health: deep learning for attack detection in medical IoT," *Procedia Comput. Sci.*, vol. 235, pp. 2498–2507, 2024.

[11] P. A. A. Resende and A. C. Drummond, "A survey of random forest based methods for intrusion detection systems," *ACM Comput. Surv.*, vol. 51, no. 3, pp. 1–36, 2018.

[12] G. Sunkara, "AI-driven cybersecurity: Advancing intelligent threat detection and adaptive network security in the era of sophisticated cyber attacks," *Well Test. J.*, vol. 31, no. 1, pp. 185–198, 2022.

[13] M. K. Divya, "Machine learning for intrusion detection," *Deep Learn. Intrusion Detect. Tech. Appl.*, pp. 25–58, 2026.

[14] M. Aljanabi, "Intrusion detection systems, issues, challenges, and needs," *Int. J. Comput. Intell. Syst.*, 2021.

[15] Y. Yang and H. Wang, "Random forest-based machine failure prediction: A performance comparison," *Appl. Sci.*, vol. 15, no. 16, p. 8841, 2025.

[16] L. M. Dang, M. J. Piran, D. Han, K. Min, and H. Moon, "A survey on internet of things and cloud computing for healthcare," *Electronics*, vol. 8, no. 7, p. 768, 2019.

[17] M. M. Khan, N. Shah, N. Shaikh, A. Thabet, and S. Belkhair, "Towards secure and trusted AI in healthcare: a systematic review of emerging innovations and ethical challenges," *Int. J. Med. Inform.*, vol. 195, p. 105780, 2025.

[18] K. Saravanan, "Cloud-Native Generative AI Frameworks for Enterprise Workflow Automation and Secure Data Governance," *Int. J. Comput. Technol. Electron. Commun.*, vol. 8, no. 6, pp. 11952–11961, 2025.

[19] Z. I. Khan, M. M. Afzal, and K. N. Shamsi, "A comprehensive study on CIC-IDS2017 dataset for intrusion detection systems," *Int Res J Adv Eng Hub*, vol. 2, no. 02, pp. 254–260, 2024.