ISSN Online: 2327-5960

Applying Artificial Intelligence for Early Detection of Cyber Threats in Hospital Networks

Asayil Nasser Albouq
Department of Information Technology, Taif university, Taif, Saudi Arabia
Dr. Samah Alajmani
Department of Information Technology, Taif university, Taif, Saudi Arabia 2

ISSN Online: 2327-5960

Abstract

Hospitals are using more technology connected to networks for taking care of patients and running smoothly, making hacking a big issue in healthcare. While important, technology can make hospitals vulnerable to ransomware attacks, data hacks, and Distributed Denial-of-Service attacks. Fixing these issues is important for providing essential services and protecting patients' private information. Hospitals can improve their safety by using artificial intelligence. The paper will help them find and address cyber threats early instead of just reacting to them with focusing on the Deep Neural Network (DNN) model for its ability to capture complex network patterns. Additionally, K-Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Machine (SVM) are employed as comparative models to evaluate their effectiveness in identifying cyber threats. I like this subject because it combines my passion for science with solving real problems. AI can help keep healthcare systems secure, which builds trust with patients and ensures important medical services keep running.

Keywords: Cybersecurity; Artificial Intelligence; Hospital Networks; Intrusion Detection; Anomaly Detection.

1. INTRODUCTION

Artificial intelligence is the ability of technology, especially computers to imitate human thinking skills to perform tasks of learning with reasoning and fixing problems. AI uses advanced methods to copy how humans think so enabling robots to perform challenging tasks with minimal help from people. AI is being used in many fields of hacking in healthcare and transportation because it can quickly examine large amounts of data and making it more effective than the traditional methods.

The top edge of AI's skills in the area includes methods of machine learning with anomaly recognition and neural networks and because neural networks which draw inspiration from the structure of the human brain, they allow deep learning models to interpret unstructured data and make judgments based on intricate interdependencies, machine learning entails training algorithms on previous data to find patterns and forecast outcomes. Finding differences from accepted standards which will suggest unfriendly behavior inside a network is made possible by the anomaly detection. Through the use of these methods, it will make AI systems able to examine huge amounts of network data with spotting minute irregularities and predict possible dangers before they develop into major leaks.

According to Pekarčík et al. [1], cyber risks are any planned attempts to get illegal entry to systems, jeopardize data security, or mess with service availability. These risks include a broad range of actions, such as ransomware, scams, virus spread, hacking, and Distributed Denial-of-Service attacks. The sensitive nature of patient data and the vital importance of continuous service delivery make the healthcare industry with it relies on networked tools and systems an especially attractive target for hackers.

An important component of the defense against online threats is the use of breach detection systems because the systems serve as a first line of defense in finding possible breaches by tracking and studying network data for signs of the illegal and hacking behaviors. Conventional intruder detection systems use signature detection methods which match known aggressive behavior patterns to network data. The method works well against risks that have already been found but it is not very good at finding new or growing attacks and mixing advanced analytics and prediction powers with AI IDS get over these limits and are able to identify known and new dangers more quickly and correctly.

The study explores the application of AI models of Deep Neural Network (DNN) for improving hospital network security. In addition to DNN the K-Nearest Neighbors (KNN) with Random Forest (RF) and Support Vector Machine (SVM) are also investigated.

- KNN classifies network traffic based on the majority vote of its nearest neighbors, making it a simple yet effective method for detecting common attack patterns.
- RF is an ensemble learning technique that mitigates overfitting and improves classification accuracy by aggregating decisions from multiple decision trees.

- SVM finds an optimal decision boundary to distinguish between normal and malicious network activity, making it particularly useful in identifying complex cyber threats in healthcare environments.

2. RELATED WORK

The widespread usage of old IT systems that will lack the strong security features needed to fend off current attacks is one of the main weaknesses in hospital networks. According to Pekarčík et al. [1] the conventional intrusion detection systems that depend on pre-established patterns or signs of known threats have trouble seeing novel and growing attacks of advanced persistent threats and zero-day flaws and adopting increasingly advanced technologies of artificial intelligence is also important to improving cybersecurity in hospital networks because of the security flaws.

Giving predictive analytics with adaptable responses and immediate danger recognition AI has the potential to transform. AI systems also will be able to examine huge amounts of data with spot irregularities and spot minute trends that conventional methods would miss thanks to techniques of machine learning. More success has been shown in spotting and stopping intrusions before they do real harm when advanced IDS that apply AI anomaly detection are used and to improve precision and reduce false positives with AI systems will use behavioral analysis to distinguish between peaceful and malicious action [1].

Notwithstanding these benefits there are problems in using AI into hospital safety that include making sure that data privacy laws of HIPAA and GDPR are followed and the difficulty of having AI solutions inside current IT systems and to build trust among partners with ethics problems including protecting patient privacy and securing private health data must also be taken into account. However, putting AI security measures in place is becoming more and more seen as a necessary first step in building a strong healthcare environment. Example of how AI will increase total system efficiency with lessening resource usage and speed intruder detection operations is the unified management of IDS as described by Pekarčík et al. [1].

The past of hospital cybersecurity shows the promise and challenges involved with technological integration and unmatched benefits have come from digital change but it has also caused risks that need for creative fixes of artificial intelligence and by using AI in breach detection with hospitals are better able to handle future security risks and keep the safety with security and usefulness of important healthcare services and in order to provide a more safe healthcare setting the study focuses on exploring the changes and assessing how they will be used in hospital networks.

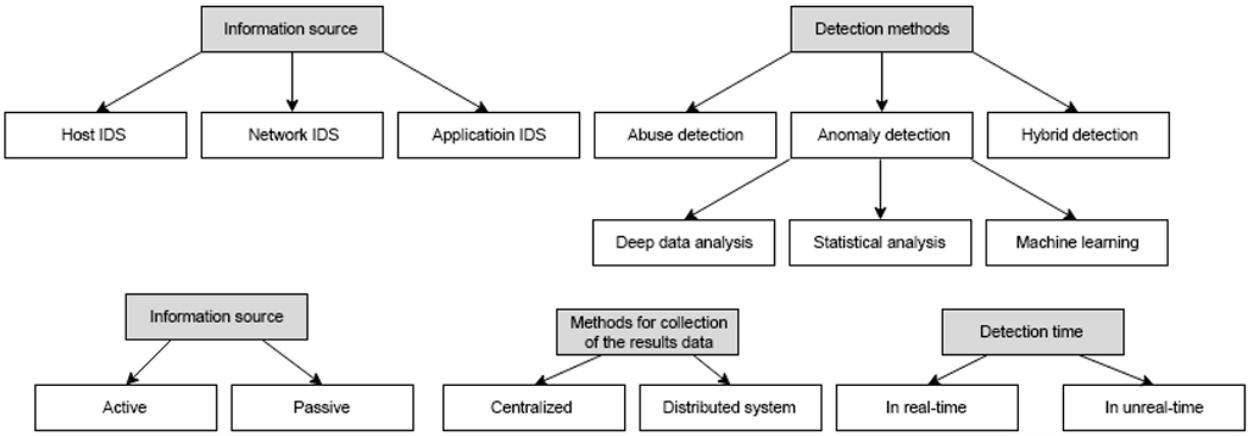

Fig. 1. Classification of intrusion detection systems [1]

With the potential to improve patient care with faster treatments and protect private medical information the mix of artificial intelligence and safety in healthcare has drawn a lot of interest. Healthcare organizations can now use automation for disease prediction with medical imaging analysis and customized treatment plans thanks to the quick adoption of AI technologies of machine learning and deep learning that improve diagnostic precision while lowering medical errors [2]. Healthcare networks are protected from breaches by AI cybersecurity frameworks that provide advanced powers for danger identification with intrusion avoidance and immediate tracking. However, cybersecurity risks have also changed in line with the rising dependence on the Internet of Medical Things and linked healthcare technology making strong AI security steps necessary [3]. Patient privacy and healthcare operations are under serious danger because of the growth of complex cyberthreats including ransomware attacks with data breaches and hacking scams.

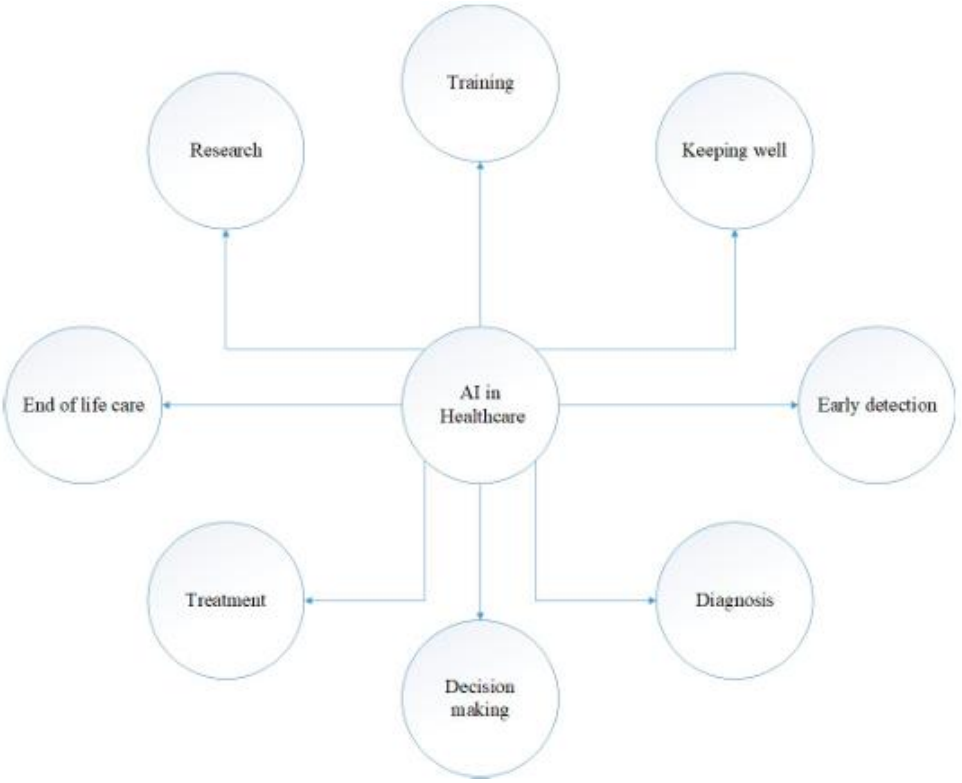

Fig. 2. Applications of AI in Healthcare [2]

Traditional security methods of rule-based intrusion detection systems and signature-based firewalls have not been able to spot new threats that do not fit pre attack patterns because of the growing complexity of assaults. Blockchain data protection methods and anomaly detection systems are examples of AI security solutions that have been put out to solve those problems. According to studies, AI cybersecurity systems beat normal security methods by recognizing dangerous behavior patterns with greatly improving attack detection and lowering false warning rates and giving immediate threat prevention [2]. AI anomaly identification is the most important uses of AI in hospital safety as it is important for lowering hacking risks by spotting odd trends or actions that point to security breaches. ML and DL methods are also used by threat detection models to study network data with spotting anomalies in normal behavior and find attacks before they become serious [4].

3. MATERIALS AND METHODS

In this study, we utilized an IoT healthcare security dataset from Kaggle, specifically designed for intensive care units. the dataset, comprising three csv files—attack, environment monitoring, and patient monitoring—captures both normal and attack traffic in an IoT-based ICU setting. before model training, we performed extensive preprocessing, including merging datasets, encoding categorical features, standardizing numerical values, and applying principal component analysis for dimensionality reduction. to detect cyber threats effectively, we implemented multiple machines learning models, including KNNs, RF, and SVM, alongside a DNN. The DNN model, structured with three fully connected layers, utilizes Relu activation and a sigmoid output layer for binary classification. model performance was assessed through accuracy and other classification metrics to ensure effective attack detection in IoT healthcare environments.

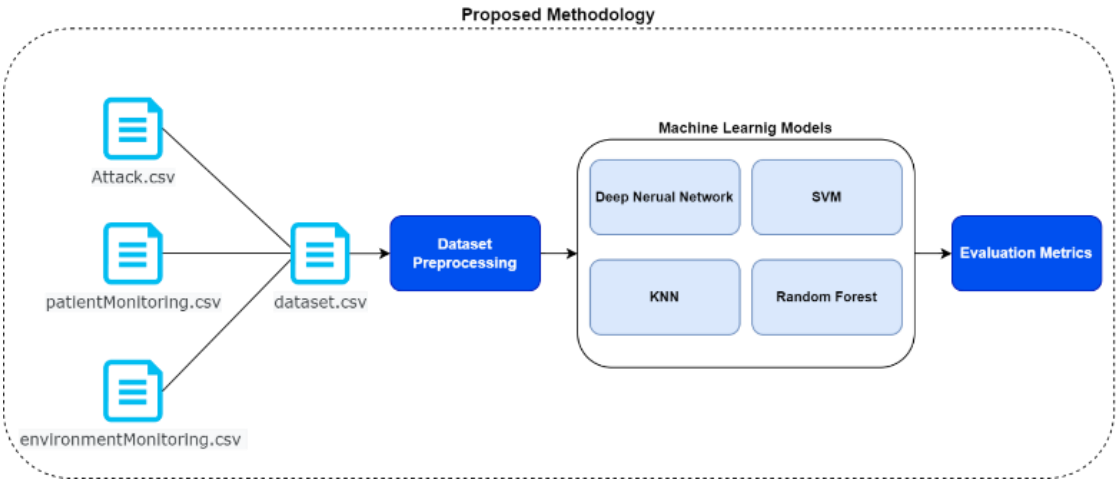

Fig. 3. Proposed Method.

3.1. Dataset:

Our study uses a dataset obtained from Kaggle, specifically an IoT healthcare security dataset for Intensive Care Units (ICUs). The dataset mainly consists of IoT healthcare data collected from both normal and attack traffic scenarios. It was generated using an IoT-based ICU setup with a capacity of two beds, following the methodology described in [5].

Each ICU bed was equipped with IoT-based monitoring devices, including sensors and an additional device for enhanced data control and the creation of a highly relevant dataset. Additionally, a Bed-Control Unit was integrated to facilitate observations and dataset generation. Data collection was supported by a specialized IoT tool known as the IoT-Flock tool. By leveraging IoT devices in healthcare, this dataset serves as a valuable resource for further research on IoT security and privacy.

The dataset consists of three CSV files: Attack.csv, Environment_Monitoring.csv, and Patient_Monitoring.csv. The Attack.csv file records cyberattack-related data by capturing network traffic within the healthcare sector. The Environment_Monitoring.csv file contains data from environmental sensors, capturing normal network traffic in the healthcare environment. Lastly, the Patient_Monitoring.csv file includes data from ICU patient monitoring sensors under normal network conditions. Together, these three files provide a comprehensive dataset, facilitating further implementation and analysis.

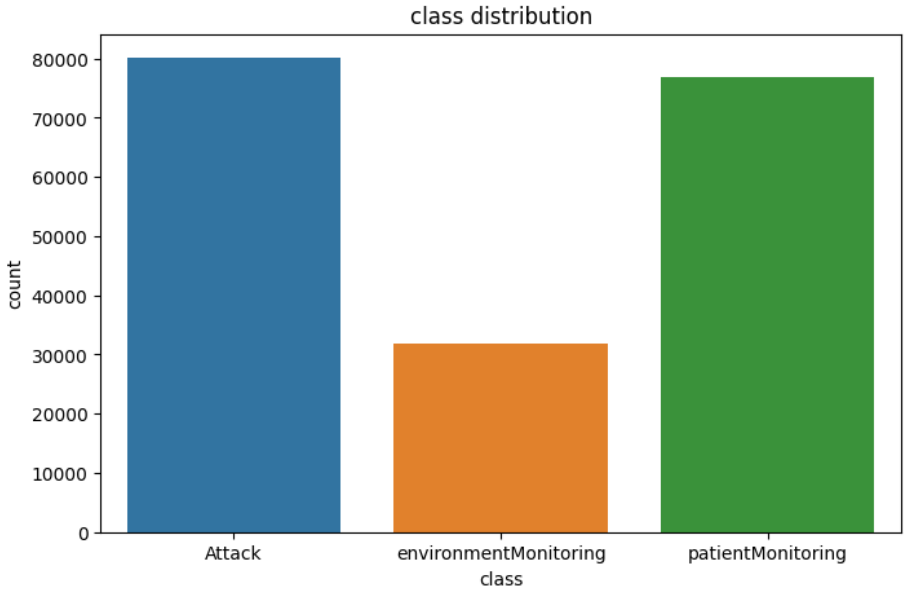

Fig. 4. Class Distribution.

3.2. Dataset Preprocessing:

Before training the model, extensive preprocessing was applied to the dataset to enhance its quality and suitability for machine learning algorithms. Initially, the dataset, composed of three CSV files (Attack, Environment Monitoring, and Patient Monitoring), was merged into a single structured dataset. categorical features were encoded using frequency encoding, replacing categorical values with their

respective occurrence frequencies in the dataset. To standardize the data, all numerical features were scaled using the StandardScaler, ensuring that each feature had zero mean and unit variance. Furthermore, Principal Component Analysis (PCA) was applied to reduce the feature space while preserving the most significant variance in the data. The dataset was then split into training (70%), validation (15%), and test (15%) sets.

3.3. Machine Learning Models:

A) K-Nearest Neighbor (KNN)

KNN tasks. This model efficiently categorizes incoming data into predefined groups. Utilizing the concept of "feature similarity", KNN computes the distances between the new data point and each existing example in the dataset, selects the K closest examples, and determines the label with the highest occurrence among them to estimate the values of the new data points [6]. During the learning process, KNN compares a given test instance with similar training instances. The training data exists in an n-dimensional pattern space. When an unknown instance is introduced, the KNN classifier searches this space to find the k most similar training examples. These k training instances, referred to as the "nearest neighbors" of the unknown instance, play a crucial role in determining its classification or regression output [7].

B) Random Forest (RF)

RF is a widely applied ensemble learning technique that merges the results of multiple decision trees to enhance the prediction accuracy. During training, it forms a set of decision trees and merges the predictions—employing the mode if it is a classification problem and the mean if it is a regression problem—into the final prediction. This cooperative decision-making improves the strength of the model and reduces the chance of overfitting.

C) Support Vector Machine (SVM)

SVM is a powerful algorithm used in both regression and classification tasks. SVM separates the data points of distinct classes in a high-dimensional space with the help of the best hyperplane. SVM maximizes the margin between classes, enhancing its robustness to outliers and enabling it to handle complex decision boundaries effectively.

For a binary classification problem [8], given a dataset $\{(x_i, y_i)\}_{i=1}^{n}$, where x_i represents feature vectors and $y_i \in \{-1,1\}$ represents class labels, the decision hyperplane is defined as:

$$w^T x + b = 0 \qquad (1)$$

where:

- w is the weight vector,
- x is the input feature vector,
- b is the bias term.

The margin is defined as the distance between the hyperplane and the closest data points (support vectors). SVM maximizes the margin $\frac{2}{\|w\|}$ while ensuring correct classification, leading to the following optimization problem:

$$\min_{w,b} \frac{1}{2} \| w \|^2 \qquad (2)$$

subject to the constraints:

$$y_i(w^T x_i + b) \geq 1, \forall i \qquad (3)$$

D) Deep Neural Networks

DNN are a type of artificial neural network (ANN) with multiple hidden layers that enable the model to learn complex data representations through hierarchical feature extraction. These networks automatically capture nonlinear patterns in data as information propagates through successive layers.

The DNN model architecture used for binary classification consists of three fully connected layers with ReLU activation functions, allowing the model to learn nonlinear relationships in the dataset. The input layer processes a six-dimensional feature vector, followed by the first hidden layer with 16 neurons, computed as:

$$H_1 = \mathrm{Re}LU(W_1 X + b_1) \qquad (4)$$

where X represents the input features, W_1 is the weight matrix, and b_1 is the bias term.

The second hidden layer consists of 8 neurons, expressed as:

$$H_2 = \mathrm{Re}LU(W_2 H_1 + b_2) \qquad (5)$$

Finally, the output layer consists of a single neuron with a sigmoid activation function to produce a probability score for binary classification:

$$\hat{y} = \sigma(W_3 H_2 + b_3) \qquad (6)$$

Where:

$$\sigma(x) = \frac{1}{1+e^{-x}} \qquad (7)$$

ensures that outputs range between 0 and 1. The model is trained using Stochastic Gradient Descent (SGD) and optimized with Binary Cross-Entropy Loss. Model performance is evaluated using accuracy and other classification metrics.

3.4. Evaluation Metrics:

A) Confusion Matrix

A confusion matrix is a table used to assess the performance of a classification model by comparing its predicted results with the actual target values from a test dataset. It offers an in-depth analysis of the model's accuracy and error distribution [9].

The confusion matrix consists of the following key terms:

- **True Positives (TP):** Positive instances correctly predicted.
- **False Negatives (FN):** Positive instances incorrectly classified as negative.
- **False Positives (FP):** Negative instances incorrectly classified as positive.
- **True Negatives (TN):** Negative instances correctly predicted.

Using the confusion matrix, various performance metrics are computed. These metrics are essential for determining the best model or algorithm's performance within a specific ML application [10].

B) Accuracy

Accuracy is a general measure of the performance of a classification model. It is calculated in the ratio of correctly predicted instances with respect to the total instances in the evaluation set based on the equation in Equation (8).

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (8)$$

C) Precision

Precision is the ratio of retrieved samples that are relevant and the ratio of correctly classified samples and the total samples that fall within that category and is defined in equation (9)

$$\text{Precision} = \frac{TP}{TP+FP} \quad (9)$$

D) Recall

Recall, or sensitivity, evaluates a model's capability to accurately detect positive instances. It is computed as the proportion of correctly classified positive samples to the total number of actual positive samples, as given in Equation (10).

$$\text{Recall} = \frac{TP}{TP+FN} \quad (10)$$

E) F1-Score

Defined in Equation below, the F1-score expresses the harmonic mean of precision and recall. It offers a single metric ranging from 0 (worst) to 1 (best), summarizing the performance.

$$F_1 - \text{Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

4. RESULTS

In this study, we propose a deep learning-based approach for classification, leveraging a Deep Neural Network trained over 2000 epochs. The performance of the DNN is evaluated against traditional machine learning models, including KNN, RF, and SVM. The evaluation is conducted using precision, recall, F1-score, and accuracy metrics, alongside confusion matrices for each model.

4.1. DNN Results

The training process for the DNN model was conducted over 2000 epochs, showing a steady decline in training, validation, and testing loss. Initially, the model struggled with accuracy around 41%, but performance improved significantly as training progressed. By epoch 100, the testing accuracy had reached 58.54%, and by epoch 600, it had exceeded 97%. The final test accuracy stood at 98.94%, demonstrating strong learning capability. The training curve figure below shows a gradual decrease in loss over time, with the model avoiding overfitting due to a close match between training and validation loss. In the classification report Precision, recall, and F1-score are near 99%, highlighting the robustness of the DNN model.

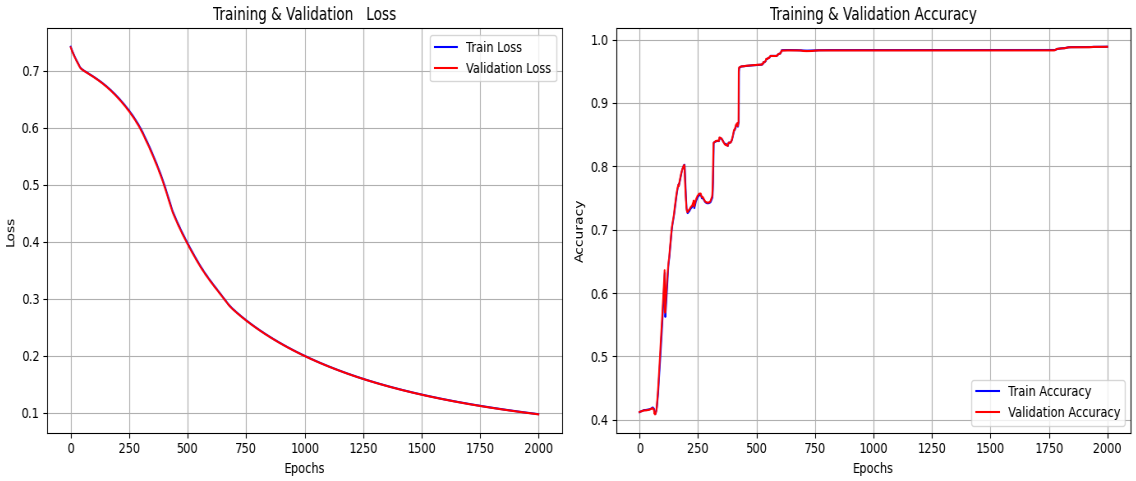

Fig. 5. The training curves.

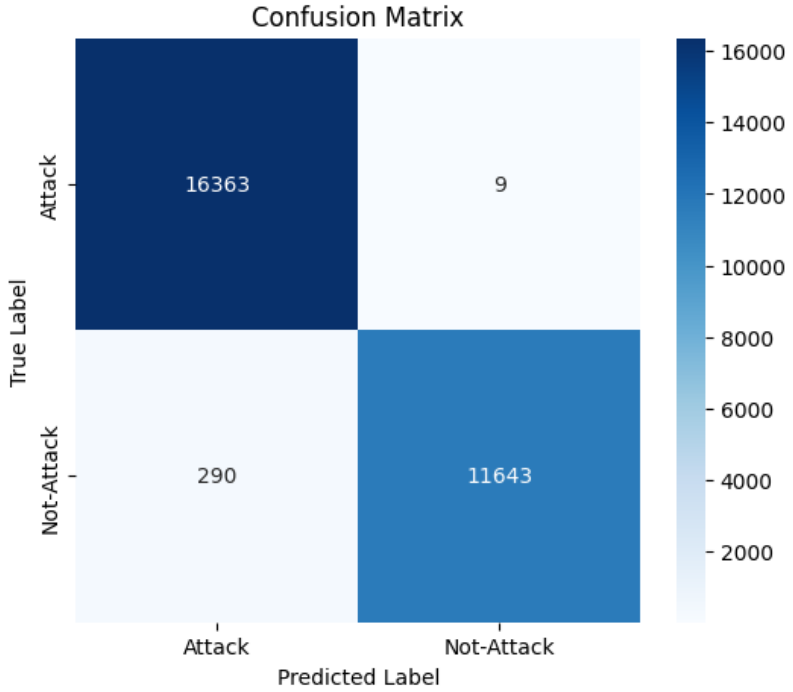

Fig. 6. The confusion matrix indicates a highly accurate classification, with only 290 false negatives and 9 false positives.

4.2. KNN Results:

The KNN model achieved an exceptional accuracy of 99.99%, demonstrating its effectiveness for this classification task. And in Classification report the Precision, recall, and F1-score are 99.99% across both classes, reflecting KNN's high reliability in this dataset.

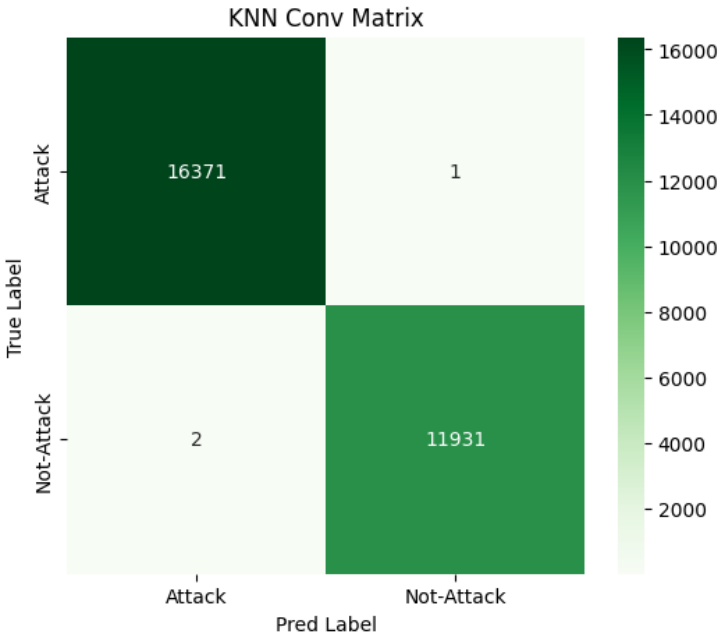

Fig. 7. Only 3 misclassified instances (1 false positive, 2 false negatives) out of 28,305, indicating almost perfect classification.

4.3. RF Results:

RF classifier performed the best among all models, achieving a 99.99% accuracy on the test dataset. And in classification report the Precision, recall, and F1-score are all 99.99, demonstrating that the model is highly robust and generalizes exceptionally well.

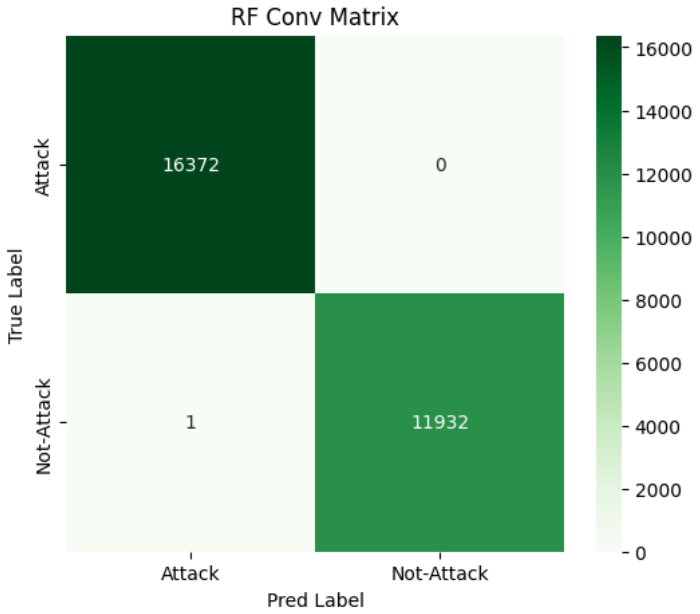

Fig. 8. Only one false negative misclassification, confirming near-perfect classification.

4.4. SVM Results:

SVM delivered outstanding results, with a test accuracy of 99.92%, slightly lower than KNN and RF. With precision, recall, and F1-score near 99.92%, the SVM model still performed exceptionally well.

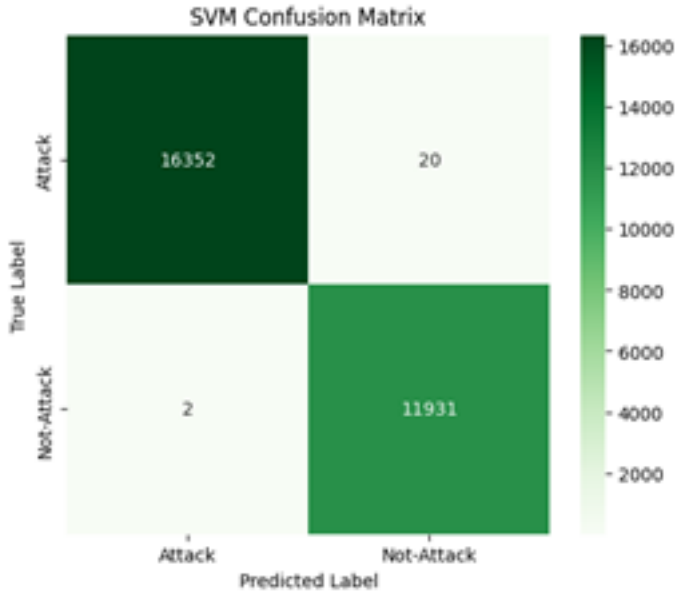

Fig. 9. The SVM model had 22 misclassified instances (20 false positives and 2 false negatives), which is slightly higher than KNN and RF but still minimal.

4.5. Comparison of Results

TABLE I. COMPARISON TABLE OF MODEL PERFORMANCE.

Model	Accuracy	Precision	Recall	F1-Score	False Positives	False Negatives
DNN	98.94%	98.96%	98.94%	98.94%	9	290
KNN	99.99%	99.99%	99.99%	99.99%	1	2
RF	99.99%	99.99%	99.99%	99.99%	0	1
SVM	99.92%	99.92%	99.92%	99.92%	20	2

Based on these results, RF and KNN are the top-performing models for this dataset. If computational efficiency is a concern, KNN offers similar performance with lower training complexity compared to RF. However, for highly generalized classification problems, RF remains the optimal choice due to its perfect accuracy.

5. Discussion

Our experimental results demonstrate strong performance in detecting cyber threats in IoT healthcare environments, with our models achieving high accuracy across various classification metrics. Comparing our findings with the study by Savanović et al. [11], we observe that their optimized machine learning models, particularly RF and KNN, achieved high accuracy scores of 99.51% and 99.48%, respectively. Similarly, our implementation of these models yielded comparable results, reinforcing the reliability of RF and KNN for intrusion detection in IoT healthcare systems.

To provide a structured comparison, the following table presents the performance metrics of our models against those reported by Savanović et al. [11].

TABLE II. COMPARISON OF OUR MODEL PERFORMANCE VS MODEL PERFORMANCE OF SAVANOVIĆ ET AL. [11].

Model	Our Accuracy (%)	Savanović et al. [11] Accuracy (%)	Our F1-Score	Savanović et al. [11] F1-Score	Our False Positives	Our False Negatives
DNN	98.94%	98.96%	98.94%	98.94%	9	290
KNN	99.99%	99.99%	99.99%	99.99%	1	2
RF	99.99%	99.99%	99.99%	99.99%	0	1
SVM	99.92%	99.92%	99.92%	99.92%	20	2
Decision Tree (DT)	-	99.48	-	99.63	-	-
Adaptive Boosting (AB)	-	99.50	-	99.47	-	-

Logistic Regression (LogR)	-	99.50	-	94.70	-	-
XGBoost with MFA	-	99.70	-	99.69	-	-

Our RF model achieved 99.99% accuracy, precision, recall, and F1-score, outperforming all models in Savanović et al.'s study [11]. Additionally, KNN performed exceptionally well (99.99% accuracy and F1-score), surpassing the results in the comparison study (99.49%). Our DNN model achieved 98.94% accuracy, showing robust performance, although it was slightly lower than some traditional machine learning models.

Savanović et al. [11] utilized metaheuristic optimization (Modified Firefly Algorithm) to enhance model performance. Despite not incorporating such optimization techniques, our models—especially RF and KNN—outperformed their results.

6. CONCLUSION

The research successfully examined the application of DNN with Random Forest and KNN and SVM for safety in hospital networks based on the IoT. The main focus was on early cyber threat identification in intensive care unit patient tracking devices where the risk of hacking is especially high. The DNN model was effective in recognizing and classifying cyber-attacks with higher precision than traditional intrusion detection systems, as proven by its high accuracy (98.94%), precision (98.96%), and recall (98.94%).

DNN outperformed traditional cybersecurity solutions by delivering higher generalization with fewer false positives and more flexibility to accept unique attack patterns. The model proven to be an excellent tool for hospital network security with recognizing important cyberthreats of ransomware or phishing attempts and man-in-the-middle attacks and zero-day vulnerabilities. In hospital environment the cyber dangers are getting more complex in line with digital transformation as the research also revealed the expanding demand for AI cybersecurity solutions.

Even if it functioned well, there are still a lot of difficulties. Obstacles to DNN broad application in medical settings is mainly because of the high computational cost with vulnerability to hostile attacks and issues about data privacy and regulatory compliance.

REFERENCES

[1] P. Pekarčík, E. Chovancová, M. Chovanec, and M. Štancel, "A centralized approach to intrusion detection system management: Design, implementation and evaluation," Acta Polytech. Hung., vol. 22, no. 1, pp. 7–26, 2025.

[2] A. Alzahrani, "Using artificial intelligence and cybersecurity in medical and healthcare applications," Inf. Sci. Lett., vol. 12, no. 3, pp. 1579–1590, 2023.

[3] D. Lee and S. N. Yoon, "Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges," Int. J. Environ. Res. Public Health, vol. 18, no. 1, p. 271, 2021.

[4] S. K. Akinade, "Implementing AI-driven anomaly detection for cybersecurity in healthcare networks," ATBU J. Sci. Technol. Educ., vol. 12, no. 2, pp. 598–610, 2024.

[5] F. Hussain, S. G. Abbas, G. A. Shah, I. M. Pires, U. U. Fayyaz, F. Shahzad, N. M. Garcia, and E. Zdravevski, "A framework for malicious traffic detection in IoT healthcare," Sensors, vol. 21, p. 3025, 2021.

[6] V. Sheth, U. Tripathi, and A. Sharma, "A comparative analysis of machine learning algorithms for classification purpose," Procedia Comput. Sci., vol. 215, pp. 422–431, 2022.

[7] M. J. Muzammil, S. Qazi, and T. Ali, "Comparative analysis of classification algorithms performance for a statistical-based intrusion detection system," in Proc. IEEE Int. Conf. Computer, Control and Communication (IC4), Karachi, Pakistan, 25–26 Sep. 2013, pp. 1–5.

[8] R. Berwick, An Idiot's Guide to Support Vector Machines (SVMs), unpublished.

[9] A. Zewdu and B. Yitagesu, "Part of speech tagging: A systematic review of deep learning and machine learning approaches," J. Big Data, vol. 9, p. 10, 2022.

[10] S. A. Hicks, I. Strümke, V. Thambawita, M. Hammou, M. A. Riegler, P. Halvorsen, S. Parasa, "On evaluation metrics for medical applications of artificial intelligence," Sci. Rep., vol. 12, p. 5979, 2022.

[11] N. Savanović, A. Toskovic, A. Petrovic, M. Zivkovic, R. Damaševičius, L. Jovanovic, N. Bacanin, and B. Nikolic, "Intrusion detection in healthcare 4.0 Internet of Things systems via metaheuristics optimized machine learning," Sustainability, vol. 15, p. 12563, 2023.