IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 19, April 2026
Impact Factor: 8.2

A Proactive Cyber Defense Framework For Healthcare Infrastructure Using Federated Learning, Honeypots, And Hybrid AI

Bryson Lopes[1], Brett Lopes[2], Sherwin Dmello[3], Jeff Dsouza[4]
Student, Department of Artificial Intelligence and Data Science
Fr. Conceicao Rodrigues College of Engineering, Mumbai, Maharashtra, India[1–4]
10067@crce.edu.in[1], 10066@crce.edu.in[2],
10444@crce.edu.in[3], 10048@crce.edu.in[4]

Abstract: *Cyber threats in healthcare have increased rapidly due to the growing use of connected medical devices and digital systems. While these technologies improve efficiency and patient care, they also create new security risks such as ransomware attacks and zero-day exploits. Traditional intrusion detection systems often struggle to identify complex or evolving threats, and they may not meet strict data privacy requirements in healthcare environments.In this paper, we present HealthSentinel, a cybersecurity framework designed specifically for healthcare systems. Instead of relying only on centralized data processing, our approach uses Federated Learning so that multiple healthcare institutions can train detection models together without sharing sensitive data.The system combines OSQuery with a Long Short-Term Memory (LSTM) Autoencoder to analyze system-level logs and detect unusual patterns over time. It also uses an Isolation Forest to identify unknown or zero-day attacks. In addition, honeypots are deployed to act as decoys and capture attacker behavior at an early stage.To ensure the system is practical and easy to use, the framework is implemented using a FastAPI backend and visualized through a secure web dashboard protected by Multi-Factor Authentication. We also include explainable AI techniques such as SHAP and integrate a language model to generate clear, human-readable threat reports.Experimental results show that the proposed hybrid approach achieves high detection accuracy with a low false-positive rate, making it suitable for protecting sensitive healthcare infrastructure while maintaining data privacy.*

Keywords: Healthcare Cybersecurity, Federated Learning, Intrusion Detection System, LSTM, Anomaly Detection, Honeypot, Explainable AI.

I. INTRODUCTION

Modern healthcare systems are becoming increasingly dependent on connected technologies such as Smart Healthcare Systems (SHS) and digital medical infrastructure. These advancements help improve patient care and hospital efficiency, but they also introduce serious cybersecurity risks. Attackers can exploit these systems to disrupt operations, manipulate medical data, or even interfere with critical devices, which can have severe consequences.

Recent ransomware attacks on hospitals have shown that traditional security systems are no longer sufficient. Most existing intrusion detection systems rely on signature-based methods, which are not effective against advanced or unknown threats. Although machine learning-based solutions have been proposed, many of them require centralized data collection. This creates privacy concerns and may violate regulations such as HIPAA. In addition, many current systems are reactive, meaning they detect attacks only after damage has already begun.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 19, April 2026 Impact Factor: 8.2

To address these challenges, we propose HealthSentinel, a proactive and privacy-focused cybersecurity framework. The system is designed to shift from reactive detection to intelligent and early threat prevention. It combines deep learning techniques, decentralized training, and active defense strategies to improve overall security.

The key contributions of this work are as follows:

First, we introduce a hybrid anomaly detection approach that combines OSQuery with an LSTM Autoencoder and Isolation Forest. This allows the system to detect both known attack patterns and previously unseen threats by analyzing system-level logs over time.

Second, we implement a Federated Learning architecture that enables multiple healthcare institutions to collaboratively train models without sharing raw data, thereby preserving privacy.

Third, we incorporate honeypots into the network to act as decoy systems. These help in detecting attackers early and collecting valuable threat intelligence.

Fourth, we use explainable AI techniques such as SHAP along with language model integration to make the system's decisions easier to understand for security analysts.

Finally, the entire system is deployed using a FastAPI backend with a secure web dashboard protected by Multi-Factor Authentication, enabling real-time monitoring and practical usability in healthcare environments.

II. RELATED WORK

The intersection of artificial intelligence and healthcare cybersecurity has developed quickly over the last decade. As Smart Healthcare Systems (SHS) and the Internet of Medical Things (IoMT) grow, researchers have turned to machine learning to address increasing cyber threats.

A. Machine Learning in Healthcare Cybersecurity

Traditional intrusion detection systems (IDS) rely heavily on signature-based methods. These methods are limited when dealing with polymorphic malware or zero-day exploits. To tackle this challenge, researchers have suggested behavioral anomaly detection models. A key framework in this area is HealthGuard [1], created by Newaz et al. (2019). HealthGuard demonstrated that machine learning algorithms, such as Artificial Neural Networks and Decision Trees, can monitor physiological vital signs and spot unusual activities that indicate a cyberattack on medical devices. While effective for specific device attacks, it mainly focused on physiological data rather than system-level network logs.

Later research built on deep learning techniques for broader intrusion detection. Shamsolmoali et al. (2024) [2] reviewed deep learning technologies for IoT intrusion detection. They pointed out that sequence-based models, like Long Short-Term Memory (LSTM) networks, perform better in capturing temporal dependencies in network traffic compared to static models. Similarly, Bensaoud et al. (2024) [3] showed that hybrid deep learning architectures greatly enhance detection rates for evasive malware.

B. The Threat Landscape: Ransomware and Data Breaches

The need for better IDS in healthcare is highlighted by the changing threat landscape. Thamer and Alubady (2021) [4] conducted an in-depth survey of ransomware attacks in healthcare. They noted that this sector is particularly vulnerable due to its critical need for constant data access, making hospitals prime targets for extortion. Additionally, Reddy et al. (2023) [5] examined the main causes of healthcare data breaches. They found that while external hacking (malware/ransomware) is common, insider threats and poor access controls (internal breaches) are also significant vulnerabilities.

C. Identified Research Gaps

Despite the progress in ML-based IDS, our literature review reveals three key research gaps that current frameworks do not sufficiently address:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 19, April 2026
Impact Factor: 8.2

- Privacy Bottlenecks: Most high-performing deep learning models require centralized data pooling for training. Various regulatory studies have shown that pooling raw patient data and hospital network logs violates strict privacy laws like HIPAA and GDPR.
- Lack of Proactive Deception: Current models, including HealthGuard, are essentially reactive. They identify anomalies during or after a breach. There is a noticeable absence of proactive defense measures, such as honeypots, that can trap attackers early in the reconnaissance phase.
- The "Black Box" Problem: Deep learning models, especially LSTMs, struggle with interpretability. Security Operations Center (SOC) analysts often cannot trust an AI's alert if it cannot explain why specific network traffic was flagged as unusual.

HealthSentinel is specifically designed to address these gaps by integrating deep sequence learning (LSTM) on OS-level logs with Federated Learning for privacy, honeypots for proactive deception, and SHAP-driven explainability to clarify the AI's decision-making process..

III. PROPOSED METHODOLOGY

Architectural Overview

The HealthSentinel framework is a multi-tiered architecture for privacy-preserving cybersecurity in Smart Healthcare Systems. The methodology evolves from proactive network deception to deep host-based telemetry analysis, culminating in a hybrid deep learning detection engine trained via Federated Learning. System Architecture Overview The architecture consists of five distinct operational phases: Proactive Deception: Deployment of simulated medical assets (Honeypots) to trap lateral movement. Telemetry Extraction: Deep system-level log harvesting using OSQuery. Sequential Feature Engineering: Transformation of unstructured logs into time-series tensors. Hybrid AI Engine: A Dual-Core anomaly detection system utilizing a Long Short-Term Memory (LSTM) Autoencoder and an Isolation Forest. Decentralized Training & Secure Deployment: Federated Learning across hospital nodes, served via a FastAPI backend to a Multi-Factor Authentication (MFA) secured dashboard.

Proactive Deception via Honeypots

A basic limitation of traditional Intrusion Detection Systems (IDS) is that they only detect an attacker after a legitimate asset has been targeted. HealthSentinel uses high-interaction honeypots deployed in the Smart Healthcare System (SHS) subnet to avoid this asymmetric disadvantage.

- Decoy Asset Generation: These honeypots are purposely configured to simulate vulnerable Internet of Medical Things (IoMT) devices, unpatched Electronic Health Record (EHR) databases, and administrative SSH terminals that are left open.
- High-Fidelity Threat Intelligence: Since real hospital staff and automated clinical processes would have no reason to interact with these decoy assets, any connection attempt, port scan, or payload execution against a honeypot will produce a zero-false-positive alert.
- Pre-emptive Signature Extraction: The moment an adversary tries lateral movement into the honeypot, it logs the attacker's IP, execution behavior, and payload signatures for immediate use in dynamically updating host-based detection models across the legitimate network.

Deep Host-Based Telemetry Extraction

Network-level packet inspection, for example, PCAP analysis, often does not work against modern encrypted payloads, fileless malware, and insider threat escalation. That is why HealthSentinel uses deep host-based telemetry via OSQuery. OSQuery exposes the operating system as a high-performance relational database; thus, our framework can run SQL-driven queries to extract low-level system states. It continuously queries critical system tables like: processes to monitor process trees parent-child relationships execution paths. listening_ports and process_open_sockets to detect unauthorized outbound connections or reverse shells. logged_in_users and last for tracking anomalous authentication patterns. This

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 19, April 2026 Impact Factor: 8.2

results in an ordered stream of operating system events that delineate exactly how the behavioral rhythm looks within a hospital's digital infrastructure.

Sequence Formatting and Feature Engineering

The telemetry data that OSQuery retrieves is raw and structured in categories; therefore, it is not suitable for direct ingestion by a neural network. HealthSentinel has a detailed preprocessing pipeline: Categorical Embedding: Features like process names, execution paths, and user roles are converted into dense vector forms through embedding layers which maintain the semantic relationship between analogous system processes. Temporal Windowing: Cyberattacks are never instantaneous but rather sequences of events (for example, unauthorized login followed by privilege escalation and registry modification leading finally to data exfiltration). To represent such chronological dependencies, the tokenized log data gets divided into sliding temporal windows. Let X=$\{x_1,x_2 ,\ldots,x_T\}$ denote a sequence of system events over some time window T. Each $x_t \in R^n$ is an n-dimensional feature vector describing the state of the system at time step t.

Privacy Preserving Federated Learning

Healthcare organizations must combine extremely sensitive network and user data into a single repository in order to use traditional centralized machine learning models, which creates a huge single point of failure and violates laws like HIPAA. HealthSentinel uses a Federated Learning architecture to address this issue.The local hospital nodes (clients) in our framework receive the LSTM and iForest models. The model is trained by each hospital using its proprietary, localized OSQuery logs. Each hospital only sends the calculated mathematical gradients to a central aggregator server rather than raw data.

API Architecture and Secure Deployment

The pipeline is designed for practical implementation so that hospital IT personnel can actively use the mathematical models.

FastAPI Integration: A FastAPI framework encloses the Dual-Core AI engine. FastAPI serves as a fast bridge between the OSQuery agents and the detection engine, enabling HealthSentinel to process thousands of concurrent log evaluations with sub-millisecond latency by utilizing Python's asynchronous capabilities (asyncio).

Secure Web Interface: A Streamlit web application is used to visualize the threat telemetry. Since this dashboard serves as the hospital's Security Operations Center (SOC), Multi-Factor Authentication (MFA), which combines standard credentials with time-based one-time passwords (TOTP) to prevent unauthorized administrative access, strictly gates access.

Explainable AI for Threat Intelligence

The "black box" nature of deep neural networks is a significant obstacle to AI adoption in healthcare cybersecurity. Without knowing where an alert came from, security analysts cannot simply believe it. The model's output is interpreted by HealthSentinel using SHapley Additive exPlanations (SHAP).In order to assign a marginal contribution score to each feature (such as a particular API call or IP address) toward the final anomaly prediction, SHAP draws on cooperative game theory. For feature i, the SHAP value Øi is computed as follows:

$$\phi_i(v) = \sum_i \frac{|S|!\,(|N| - |S| - 1)!}{|N|!}\big(v(S \cup \{i\}) - v(S)\big)$$

where N is the set of all features, S is a subset of features, and v(S) is the model output for subset S.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 19, April 2026
Impact Factor: 8.2

IV. EXPERIMENTAL SETUP AND IMPLEMENTATION DETAILS

To rigorously assess the performance of the HealthSentinel framework, a comprehensive experimental testbed was created to simulate a distributed network of Smart Healthcare Systems. This section describes the hardware specifications, dataset generation, model hyperparameters, and software stack used for end-to-end implementation

Experimental Testbed and Network Topology

The Federated Learning architecture necessitates deployment in a distributed system. We simulated a multi-hospital consortium using separate virtual environments. The topology included: Client Nodes (Hospital Endpoints): Three different virtual machines representing separate hospital networks (Hospital A, B, and C). Each node had an OSQuery daemon running to collect local host telemetry. Central Aggregator Server: A central server managing the Federated Averaging process which does not receive any raw OSQuery logs but only encrypted gradient weight updates from client nodes. Hardware Specifications Central Aggregator and Client Nodes were simulated on a high-performance workstation with an AMD Ryzen 9 processor, 64 GB RAM, and NVIDIA RTX 3080 Ti (12GB VRAM) to accelerate LSTM Autoencoder matrix multiplications using CUDA.

Dataset Generation and Threat Simulation

As real-world hospital OS logs with live zero-day attacks are highly classified and not available due to laws protecting patient privacy, we created a synthetic high-fidelity telemetry dataset in our testbed. Benign Telemetry Generation: We simulated normal hospital activities for 72 hours. This included automated scripts that acted like doctors using EHR APIs, system updates running in the background, network backups happening regularly, and IoT devices sending heartbeats as they normally would. OSQuery captured this baseline behavior which amounted to around 500,000 benign sequential log events. Malicious Threat Injection: In order to train and test the models on active threats, we conducted controlled cyberattacks against the testbed and our deployed Honeypots. The simulated attack vectors included: Ransomware Execution (e.g., simulated Ryuk/Conti behavior): Fast file traversal, unauthorized encryption API calls, and shadow copy deletion. Insider Threat Privilege Escalation: Unauthorized execution of PowerShell scripts and modifications to registry keys. IoT Botnet Lateral Movement: Sequential port scanning and SSH brute-force attempts on the simulated medical devices.

Neural Network and Model Hyperparameters

Dual-Core AI was built with Python 3.10 and used TensorFlow/Keras for the LSTM Autoencoder and Scikit-Learn for Isolation Forest.

1) LSTM Autoencoder Architecture: input telemetry was reshaped to 3d tensors (samples, time_steps, features) Using a sliding window of size time_steps = 10, we captured the sequence number (last 10 OSQuery events).

Encoder: Two LSTM layers with 64 units and 32 units, using ReLU activation functions to avoid vanishing gradients.

Latent Space: RepeatVector layer connecting encoder with the decoder.

Decoder: Two LSTM layers with 32 and 64 units respectively, followed by a TimeDistributed Dense layer with Linear activation function to reconstruct the original input shape.

Optimization: The model was compiled with the Adam optimizer. We used early stopping with a patience of 5 epochs and monitored validation loss to prevent overfitting. We trained for 50 epochs with a batch size of 128.

2) Isolation Forest Settlement: The Isolation Forest was optimized to distinguish point-anomalies in the telemetry stream. Set hyperparameters to n_estimators = 200 (number of isolation trees) and max_samples = 'auto' The contamination hyperparameter (the expected proportion of outliers in the dataset), for the method adapting to such changes was dynamically determined at a value of 0.05 (5%) based on historical frequency of anomalies previously captured in honeypot logs.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 19, April 2026
Impact Factor: 8.2

Federated Learning Framework

Training was decentralized, performed using the Flower (flwr) framework. The training protocol was synchronous and followed:

The global LSTM model weights on the Central Server are initialized.

During every communication round, the server sends the weights to Hospital Nodes A, B and C.

local_epochs = 3: Each node trains on its local OSQuery dataset.The nodes send the updated gradients back to the server.

Application Programming Interface and Deployment

To turn the mathematical models into a usable cybersecurity product, we built a strong backend and frontend.

1) FastAPI Backend: We wrapped the inference engine using FastAPI. The OSQuery daemon on the host machines was set up to send HTTP POST requests with JSON-formatted system logs to the FastAPI /predict endpoint. Using asynchronous routing, the API processes the telemetry, runs it through the loaded .h5 LSTM model and Isolation Forest, and returns a binary classification (Benign/Malicious) along with the calculated anomaly score in less than 45 milliseconds.

2) Secure SOC Dashboard (Streamlit): We developed the Security Operations Center (SOC) interface using Streamlit. To protect against the risk of compromised administrative credentials, the dashboard includes Multi-Factor Authentication (MFA). We used a Time-Based One-Time Password (TOTP) system with the pyotp library, making it necessary for administrators to verify their identity through an authenticator app (like Google Authenticator) before accessing the threat intelligence platform.

3) Explainability Integration: We added the shap Python library to the FastAPI backend. When the system detects an anomaly, the backend calculates the SHAP values for the specific input sequence. These values are sent to the frontend, creating a SHAP force plot that visually highlights the specific OS processes (such as an unauthorized vssadmin.exe execution) that caused the anomaly score to exceed the threshold. At the same time, an integrated Large Language Model (LLM) API processes this output to produce a clear, plain-text incident response report for the SOC analysts.

V. RESULTS AND EVALUATION

The performance of the HealthSentinel framework was thoroughly assessed using the synthesized hospital telemetry dataset. This evaluation looked at how well the Dual-Core AI engine detects threats, the stability of the Federated Learning network, and the improvements from the Explainable AI (XAI) integration.

Evaluation Metrics

To measure the performance of the intrusion detection models, we used standard classification metrics based on the confusion matrix: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Since reducing false alarms is crucial for minimizing alert fatigue in a hospital Security Operations Center (SOC), we focused on Precision and the False Positive Rate (FPR). The metrics are mathematically defined as follows:

Accuracy: The overall proportion of correctly classified telemetry sequences.

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

Precision*:* The ratio of correctly predicted cyberattacks to all predicted cyberattacks.

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

Recall (Sensitivity): The ratio of correctly predicted cyberattacks to all actual cyberattacks.

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

F1-Score*:* The harmonic mean of Precision and Recall, providing a balanced metric for uneven class distributions.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 19, April 2026
Impact Factor: 8.2

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

False Positive Rate (FPR): The probability of a benign system event being falsely flagged as an attack.

$$\text{FPR} = \frac{FP}{FP + TN}$$

Threat Detection Performance

The goal for the AI engine was over 95% accuracy on detection and a false-positive rate<3%. The Dual-Core architecture (LSTM Autoencoder + Isolation Forest) was benchmarked against the baseline models, such as a standard Random Forest classifier, and the standalone CNN-LSTM architecture proposed in the HealthGuard study (with an accuracy of 91%).The testing dataset contained 100k sequence windows (80k benign, 20k malicious). The results show that the hybrid method outperforms.

1) Model Comparison:

Standard Random Forest: Had issues with sequential data (88.4% accuracy, 7.2 % FPR — too high).

Baseline LSTM (HealthGuard Equivalence): 91.5% accuracy with an FPR of 4.1%. Although good at detecting previously known patterns, it failed to miss multiple new simulated zero-day behaviors.

HealthSentinel Dual-Core (Proposed)- Combining LSTM Autoencoder with Isolation Forest results in 96.8% accuracy, with Precision of 97.2% and an FPR of only 1.8%.

TABLE I: PERFORMANCE COMPARISON OF INTRUSION DETECTION MODELS

Model	Accuracy	Precision	Recall	F1-Score	FPR
Random Forest	88.4%	85.1%	83.9%	84.5%	7.2%
LSTM	91.5%	90.2%	89.8%	90.0%	4.1%
HealthSentinel	96.8%	97.2%	96.1%	96.6%	1.8%

The project objectives were successfully achieved by the Dual-Core model. Reconstruction error alone allowed for the LSTM to disregard 94% of sequences that contained simulated ransomware and insider threat, while an Isolation Forest was able to capture the remaining lateral movement anomalies that were still slipping by under temporal threshold limits set on individual data streams.

Federated Learning Convergence Analysis

The central question of this study was whether decentralized training through Federated Learning would lead to a degradation in model performance relative to a centralized model that violates privacy constraints. The global model validation loss was tracked over 20 communication rounds. Centralized Baseline: A model trained on all the data aggregated at one server converged after 12 epochs to a minimum loss of $\{L\}$ = 0.015. Federated Implementation: The HealthSentinel FL model showed minor fluctuations in early communication rounds due to the non-IID nature of different hospital logs; however, by round 15, the FedAvg algorithm had effectively smoothed out gradient updates. The FL model converged at global loss {L} = 0.018 in round 18, which indicates that Federated Learning incurs only a very small performance penalty for providing an absolute mathematical guarantee of data privacy across the healthcare consortium.

Explainability and Incident Response

Quantitative accuracy is insufficient if security analysts cannot interpret the alerts; unlike quantitative analysis, this approach to evaluating the SHAP and LLM integration on the SOC dashboard can be purely qualitative and avoid specific numbers altogether.

During simulated ransomware executions, the core principle of the analysis is the use of statistical models, such as the SHAP explainer; these models and algorithms are used to process large amounts of data to get a grasp on security

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 19, April 2026 Impact Factor: 8.2

performance and risks, and the resulting visualizations accurately isolated the exact API calls driving the anomaly score.For example, during a simulated "Ryuk" attack, the sudden, unauthorized execution of vssadmin.exe delete shadows generated a massive positive SHAP contribution, triggering the isolation threshold, while normal processes like svchost. exe had a near-zero contribution.Furthermore, the LLM successfully translated these SHAP vectors into readable reports, providing specific quantifiable results that are easier to communicate to security analysts and senior-level management; instead of providing the analyst with a raw JSON payload of API calls, the dashboard outputted a high severity alert, dramatically reducing the cognitive load on non-technical staff and drastically accelerating the Mean Time to Respond (MTTR).

VI. CONCLUSION AND FUTURE WORK

The groundbreaking digitization through the adoption of Smart Healthcare Systems (SHS) and interlinked medical devices has revolutionized patient access; however, it has also rendered critical infrastructure vulnerable to cataclysmic cyber threats. Traditional, reactive intrusion detection systems struggle when faced with sophisticated zero-day exploits and typically rely on centralized data architectures that outstrip stringent healthcare privacy regulations. In this paper, we have presented and evaluated HealthSentinel — a holistic, proactive and privacy-preserving cybersecurity framework for the health-care ecosystems. Using deep host-based telemetry to pivot the defensive model from network perimeters, our framework leverages OSQuery to log highly granular, time stamped system state data. A Dual-Core AI engine further extends the state of the art for high quality sequential profiling using an LSTM Autoencoder followed by point-anomaly detection through Isolation Forest achieving a 96.8% accuracy with extremely low False Positive rate at 1.8%.Most importantly, HealthSentinel solves the long-standing healthcare dilemma of “privacy versus security” with a Federated Learning architecture. Our experimental results show that collaborative, decentralized model training can achieve near-centralized accuracy while mathematically ensuring patient and system logs never leave their local hospital networks. In addition, the inclusion of high-interaction Honeypots facilitates active deception, and with MFA for authentication/authorization at the FastAPI backend as well as SHAP based Explainable AI (XAI) dashboard functionality is also ensured that it is not just a dry theory but a highly deployable, transparent, and user-friendly Security Operations Center (SOC) solution.

Future Work: Although HealthSentinel serves as a good defense mechanism, the future work would extend its functionalities. We also wish to incorporate Reinforcement Learning (RL) agents that can autonomously mitigate threats—for example, by dynamically quarantining compromised subnets without human intervention. Second, we aim to scale the Federated Learning consortium by including heterogeneous inter-hospital datasets (e.g., rural clinics vs. metropolitan hospitals) to generalise threat detection parameters of AI even further. Third, to improve our reactive deception layer we will introduce a dynamic honeypot in which can morph their vulnerabilities depending on real-time reconnaissance traffic as they learn more about the modifications made by adversaries.

REFERENCES

[1]. K. M. I. Newaz, A. K. Sikder, M. A. Rahman, and A. S. Uluagac, "HealthGuard: A Machine Learning-Based Security Framework for Smart Healthcare Systems," *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 2019, pp. 389-396.

[2]. H. Shamsolmoali, M. Hosseinzadeh, and R. Zhang, "Deep Learning Technologies for Intrusion Detection in IoT: A Comprehensive Survey," *IEEE Internet of Things Journal*, vol. 11, no. 3, pp. 2264-2289, 2024.

[3]. A Bensaoud, J. Kalita, and M. Bensaoud, "A Survey of Malware Detection Using Deep Learning," *Machine Learning with Applications*, vol. 16, p. 100546, 2024.

[4]. N. Thamer and R. Alubady, "A Survey of Ransomware Attacks for Healthcare Systems: Risks, Challenges, Solutions and Opportunity of Research," *2021 1st Babylon International Conference on Information Technology and Science (BICITS)*, 2021, pp. 210-216.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 19, April 2026
Impact Factor: 8.2

[5]. J. Reddy, N. Elsayed, Z. ElSayed, and M. Ozer, "A Review on Data Breaches in Healthcare Security Systems," *International Journal of Computer Applications*, vol. 184, no. 45, pp. 1-7, Feb. 2023.

[6]. S. E. Schmeelk, "Where is the Risk? Government Reported Patient Medical Data Breaches," in *IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 2018, pp. 636-641.

[7]. E. A. Al-Qarni, "Cybersecurity in Healthcare: Attacks and Mitigation Strategies," *Journal of Healthcare Engineering*, vol. 2023, Article ID 5532456, 2023.

[8]. R. Gupta, P. Kumar, and S. Singh, "Security Attacks in Electronic Healthcare Systems: A Review," in *Proceedings of IEEE CTCEEC*, 2017, pp. 204-210.

[9]. M. Rahman, A. Hosseinzadeh, and T. U. Ahmed, "Securing Healthcare with Deep Learning: A CNN-Based Model for Medical IoT Threat Detection," *arXiv preprint arXiv:2410.23306*, 2024.

[10]. A Kaur and R. Kaur, "A Review on Data Breaches in Healthcare Security Systems," *International Journal of Computer Applications (IJCA)*, vol. 184, no. 45, pp. 12-18, 2023.

[11]. Franco, A. Malvoni, A. L. Dos Santos, and J. M. Silva, "A Survey of Honeypots and Honeynets for Internet of Things, Industrial Internet of Things, and Cyber-Physical Systems," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 4, pp. 2351-2383, Fourthquarter 2021.

[12]. J. P. A. Silva, F. Silva, and R. Maciel, "Explainable AI for Intrusion Detection Systems: LIME and SHAP Applicability on Multi-Layer Perceptron," *IEEE Access*, vol. 12, pp. 31540-31555, 2024.

[13]. M. A. Rahman, et al., "Intrusion Detection System for IoHT Devices using Federated Learning," *2023 IEEE International Conference on Communications (ICC)*, 2023, pp. 1-6.

[14]. O. A. Osanaiye, et al., "Machine Learning Explainability for Intrusion Detection in the Industrial Internet of Things," *IEEE Internet of Things Magazine*, vol. 7, no. 3, pp. 34-40, May 2024.

[15]. A Imteaj, U. Thakker, S. Wang, J. Li, and M. Amini, "A Survey on Federated Learning for Resource-Constrained IoT Devices," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 1-24, Jan. 2022.

[16]. Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai, "Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection," *Network and Distributed System Security Symposium (NDSS)*, San Diego, CA, USA, 2018, pp. 1-15.

[17]. F. T. Liu, K. M. Ting, and Z. Zhou, "Isolation Forest," *2008 Eighth IEEE International Conference on Data Mining*, Pisa, Italy, 2008, pp. 413-422.

[18]. I. C. C. M. de Souza, M. A. A. da Cruz, and R. de Oliveira, "Host-Based Intrusion Detection Systems Using Deep Learning: A Comprehensive Survey," *IEEE Access*, vol. 11, pp. 25302-25325, 2023.

[19]. A Alghamdi, A. Lasebae, and M. Aiash, "Microservices-Based Architecture for Edge-Cloud Healthcare Applications," *2021 IEEE International Conference on Communications (ICC)*, Montreal, QC, Canada, 2021, pp. 1-6.

[20]. M. S. Ali, M. Tariq, and P. K. Sharma, "Zero Trust-based Security Framework for Smart Healthcare Systems," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 4, pp. 2481-2495, July-Aug. 2022