Indian Journal of Information Sources and Services
ISSN: 2231-6094 (P) Vol.16, No.1, 2026, pp.260-268

DOI: https://doi.org/10.51983/ijiss-2026.16.1.27

Federated Learning-Based Anomaly Detection Algorithm for Privacy-Preserving Health Information Systems

Dr. Louai A. Maghrabi[1], Amjed Abbas Ahmed[2*], Saed Adnan Mustafa[3], Abdelrahman H. Hussein[4] and Musab A. M. Al-Tarawni[5]

[1]Department of Software Engineering, College of Engineering, University of Business and Technology, Jeddah, Saudi Arabia

[2*]Center for Cyber Security, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM), Bangi, Selangor, Malaysia; Department of Computer Techniques, Engineering Imam AlKadhum College (IKC), Baghdad, Iraq

[3]The American University of Kurdistan, Iraq

[4]Department of Networks and Cybersecurity, Hourani Center for Applied Scientific Research, Al-Ahliyya Amman University, Amman, Jordan

[5]Research Consulting Lab, Marl, NRW, Germany; Faculty of Engineering and Building Environment, Department of Electrical, Electronic and System Engineering, National University of Malaysia, Bangi, Malaysia

E-mail: [1]maghrabi@ubt.edu.sa, [2]amjedabbas@alkadhum-col.edu.iq, [3]Said_es@yahoo.com , [4]a.husein@ammanu.edu.jo, [5]musab841@yahoo.com

ORCID: [1]https://orcid.org/0000-0001-8513-0645, [2]https://orcid.org/0000-0001-6069-2967, [3]https://orcid.org/0000-0002-3099-7230, [4]https://orcid.org/0000-0001-6536-7485, [5]https://orcid.org/0000-0003-0488-8134

(Received 01 November 2025; Revised 26 November 2025, Accepted 10 December 2025; Available online 05 January 2026)

Abstract - **Integrating privacy-preserving models with anomaly detection algorithms is an important component for secure patient information processing in health information systems. This paper describes a new anomaly detection algorithm based on Federated Learning to identify anomalous behaviors in health information systems, while ensuring privacy protection. Sensitive medical information stays decentralized; there is no raw data leaving local nodes, thus protecting patient privacy. The proposed algorithm improves the identification of security anomalies like unauthorized access to health information or disruptions of the operational status of health information systems based on privacy initiatives such as HIPAA. The study evaluates the feasibility of the proposed model in currently available real-world health datasets. The results confirm promising detection accuracy with low false positive rates and improved scalability between heterogeneous health institutions. The findings suggest that federated anomaly detection provides a credible option for privacy-preserving health data usage with adequate performance and safeguards against illness-related data leaks that exploit sensitive health data.**

Keywords: **Federated Learning, Anomaly Detection, Privacy-Preserving Systems, Health Information Systems, Healthcare Data Privacy, Machine Learning Algorithms, Security Anomalies, Federated Anomaly Detection, Privacy Regulations, Distributed Learning**

I. INTRODUCTION

1.1 Background

Health information systems are increasingly relying upon electronic health records, or EHRs, and sharing healthcare data to improve clinical outcomes while experiencing greater efficiency. However, as a result of the rapid adoption of cloud-based health data management, privacy and data security have become important issues in healthcare (Islam et al., 2015; Odilova et al., 2025). Zero-day attacks can only be responded to through anomaly detection, both in identifying suspicious behaviors like unauthorized access to patient data, or suspicious activities like system intrusions or data breach activities in real-time. Conventional anomaly detection practices depend on centralized data repositories, which erodes patient privacy; additionally, with the increasing need to comply with privacy legislation such as HIPAA and GDPR, the urgent need for new privacy-preserving technologies means that future healthcare systems must be able to guarantee the privacy of patient data while also being able to detect anomalies with sufficient accurate rates (Wu, 2024).

1.2 Problem Statement

Centralized machine learning methods rely on the collection of sensitive medical data and would require aggregation for identifying anomalous activity in health information systems, putting. Patient data privacy is at risk, exposing health systems to significant security risk (Pan et al., 2019). The challenge for health information systems is to develop an anomaly detection algorithm that can analyze health data without a central aggregation that is in violation of patient privacy policies. By allowing for model training at the edge or device, federated learning could potentially provide an avenue for collaborative learning without exposing raw data. However, there has not been significant work integrating machine learning or federated learning for anomaly detection in healthcare data systems (Kim & Song, 2024; Jnr, 2020).

1.3 Objectives

The Objectives of this Research are:

- To propose a federated learning-based algorithm for conducting anomaly detection on privacy-preserving health information systems.
- To show how federated learning may improve the detection of security anomalies claims while not exchanging any secure health information (Haider et al., 2019).
- To evaluate the performance of the developed algorithm over distributed healthcare institutions, and using real datasets, demonstrate which anomaly detection method model is the most effective versus other anomaly models.

1.4 Significance of the Study

This study makes a contribution to the existing body of knowledge related to privacy-preserving machine learning with federated learning and anomaly detection in the context of health (Orthi et al., 2025; Xu et al., 2014). The proposed framework will likely provide a means for improved security to health information systems while providing a scalable compliance framework that preserves patient privacy in the detection and treatment of anomaly attacks. The framework may be offered as a service to hospitals, clinics, or health institutions, which creates a balance between data privacy and adequate security-monitoring application (Gope & Hwang, 2015).

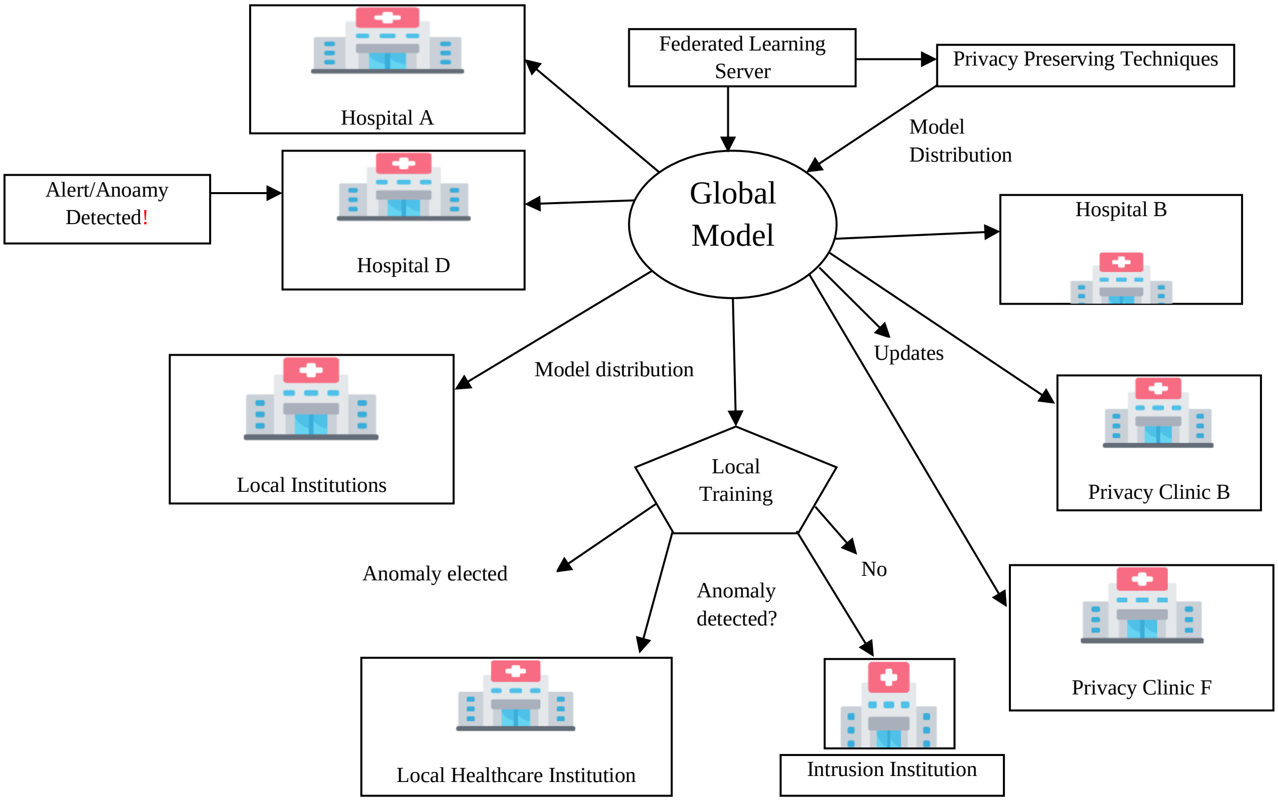

Fig. 1 Federated Learning-Based Anomaly Detection Algorithm for Privacy-Preserving Health Information Systems

Fig. 1 shows a smart trick for finding odd happenings in health care, plus making sure nobody's private stuff gets spilled. Instead of one hospital computer getting all the patient info and risking a leak, the plan uses the brains of a whole bunch of clinics and hospitals. Each place takes care of its own patient lists, then only sends a special, scrambled “update” to a main “Federated Server” instead of the actual health records (Pragadeswaran et al., 2024). The server mixes all these protected updates, then sends a stronger combined “model” back to each hospital, which fine-tunes it a bit more. This keeps going in rounds, making the computer smarter at spotting odd things, like a wrong bill, a glitch, or a sudden flu rise, all while each patient’s name stays hush-hush. The lock in the corner of the picture, along with the note that no raw data travels, loudly whispers that keeping patient privacy is rule one (Prakash & Prakash, 2023; Varatharajan et al., 2018).

II. Literature Review

2.1 Existing Anomaly Detection Methods in Healthcare: Overview of Traditional Approaches and Their Limitations

Anomaly detection in healthcare relates to the identification of indicative abnormal patterns or behaviors of medical data;

this data can be derived from patient records, sensor readings, and internet-of-things (IoT)-enabled medical devices. Traditional anomaly detection methods involve statistical methods or traditional machine learning algorithms such as decision tree methods, k-means clustering, and support vector classification (SVM) (Lu et al., 2012). While traditional methods were capable of informing deviations from its expected behavior since this is already captured in a historical pattern, they suffer from various challenges, such as high rates of false positives and the inability to generalize to unseen anomalous patterns (or zero-day intrusions). Generally, traditional anomaly detection models are excellently deployed or centralized and subsequently aggregated, which should raise eyebrows given that the aggregated data is in a national or local unit of protective privacy in sensitive and patient-centric healthcare environments. Even in large decentralized systems, these challenges have spurred more sophisticated, innovative, and decentralized anomaly detection methods, such as federated learning, that preserve rights to privacy and provide a more accurate and scalable anomaly detection (Liang et al., 2012).

2.2 Federated Learning in Healthcare: How Federated Learning Protects Privacy in Health Systems

Federated learning is a new approach in which machine learning models are trained across a number of devices (or nodes) that each have local data without exchanging data. This is especially advantageous for healthcare/PHS, where data privacy is critical (Kong et al., 2019). Federated learning ensures that sensitive health data (e.g., a patient's electronic health record (EHR) or real-time data from IoT in healthcare) remains within a healthcare provider's environment. Instead of moving the data to one domain, the updates of the model (i.e., gradients or weights) are communicated to a central server. In doing so, federated learning reduces the chance of exposing any personal data while allowing an institution to comply with applicable data protection regulations (e.g., HIPAA or GDPR). The literature supports the evidence that federated learning can improve the privacy and security surrounding health systems while enhancing anomaly detection capabilities via data sharing amongst a number of devices without compromising the privacy of patient data (Lin et al., 2009).

2.3 Applications of Machine Learning and Privacy-Preserving Techniques in Healthcare Systems

Researchers are increasingly exploring the combination of machine learning with privacy-preserving techniques for improving healthcare system accuracy and privacy. Many machine learning models have already been used to find irregular behavior, including fraud (e.g., detecting fraudulent claims), data breaches (e.g., access to sensitive material by unauthorized personnel), or unusual activity carried out by IoT devices (ADReSS Dataset, 2020; AI That Spots Alzheimer's, 2020). Privacy-preserving techniques that include homomorphic encryption, secure multi-party computation (SMPC), and federated learning have become indispensable to not violate patient confidentiality when used in a healthcare system with machine learning techniques. Studies show that federated learning, for example, can detect intrusions/anomalies with encryption methods while not exposing raw data itself to the central data storage server. Therefore, federated learning is a strong candidate for sensitive areas like healthcare, where compliance with privacy legislation is often prioritized (Fang et al., 2020).

III. Proposed Methodology

3.1 Federated Learning Algorithm: Detailed Explanation of the Federated Learning Process for Anomaly Detection

The federated learning-based anomaly detection algorithm proposed takes a decentralized approach, where several healthcare institutions (or nodes) join forces to train an anomaly detection model without ever sharing personal data. Each of the institutions will train the model on its local data and only send the model updates to a central server. The central server will take all of these updates and aggregate them to update the global model. The updated model is then sent to each node to further train locally (Li et al., 2021). This process is repeated based on the performance of the model until it is good at detecting anomalies. In the case of healthcare, the algorithm has been developed to identify anomalies like unauthorized access to data, suspicious changes to medical records, or unusual patterns within patient vital signs (Enshaeifar et al., 2018).

3.2 Data Preprocessing and Feature Extraction: Steps Involved in Preparing the Data for Training

Prior to training the federated learning model, the data needed to go through several preprocessing stages to ensure that it was clean, consistent, and useful for detection. The required data preprocessing would include normalization of the sensor data from IoT devices, missing value handling, and outlier removal for medical records. Feature extraction is important in finding useful patterns to use for anomaly detection (Hernández-Domínguez et al., 2018). For example, vital signs (heart rate, temperature, and blood pressure) data could be extracted into time-series data or put into statistical summaries, such as mean, variance, and skewness. Alternatively, we could consider user behavior features collected by hospital systems (e.g., login times, historical access patterns) with the goal to learn if there are unusual access attempts (Mirheidari et al., 2018).

3.3 Anomaly Detection Process: How Anomalies are Detected Using Federated Models

Detection of anomalous states within the subject configurations depends, after the federation has converged, on the architecture’s capacity to delineate pronounced deviations from the normative dynamic profiles archived across the geographically dispersed nodes. After sufficient training with the datasets, all nodes are able to perform anomaly detection (Zhang et al., 2020). Medical sensors, unauthorized access attempts, and unusual interaction in healthcare applications are several of these functionalities.

The model uses supervised or unsupervised learning based on the type of anomalies and the dataset characteristics. When an anomaly is detected, the system marks the event and initiates the notification processes for closer evaluation. Through the federated learning model, privacy-preserving, decentralized anomaly detection across different institutions enables continuous learning from new data streams, thereby accommodating novel data patterns (Li et al., 2016).

Mathematical Formulation: Formulation of the Anomaly Detection Model in a Federated Learning Setting

Mathematically, the anomaly detection model can be represented as an optimization problem where the goal is to minimize the loss function across multiple local datasets. Let the global model parameter θ be updated by the aggregation of local model updates Δθi from each participating institution, following the federated learning protocol:

$$\theta t+1=\theta t-\eta i=1\sum Knni\varDelta\theta i$$

In the notation, $\theta\ t$ θ t captures the model configuration indexed by the global clock t, K K serves as the cardinality of the set of distributed nodes, each corresponding to a distinct healthcare institution, η η defines the adaptive step size employed during distributed gradient optimization, and $n\ i$ quantifies the number of heterogeneous records residents at node i. Upon convergence of the federated optimization routine, the aggregated model achieves the capacity to recognize anomalous patterns over the collective of institutions, performing this task while respecting data locality by never acquiring raw, personally identifiable information.

IV. System Implementation

4.1 Tools, Frameworks, and Technologies Used

The architecture employs established frameworks, specifically TensorFlow Federated and PySyft, to instantiate federated learning. Preprocessing and feature extraction utilize the Python ecosystem, with Pandas managing tabular data operations and NumPy executing numerical transformations. Neural-net architecture specification, training modules, and forward-passage computation are all mediated through the TensorFlow framework, such that tunable layers, loss regimes, and optimizers are uniformly accessible across distributed contexts. Inter-process communications, both horizontal and vertical, are encapsulated in channels secured through the OpenSSL library, which applies TLS constructs (BBC, 2020). This exo-coupled restriction ensures that all data in motion, specifically, personally identifiable health records and derived clinical features, are subjected to authenticated ciphering and integrity checking, thereby maintaining compliance with regulatory frameworks governing confidentiality. Simulation is materialized in a controlled federation emulation representing a derived landscape of a multi-institutional healthcare alliance, wherein hospitals and outpatient clinics impersonally aggregate and share anonymized patient encounters. The federative corpus comprehensively comprises clinical electronic health records, streaming telemetry from Internet-of-Medical-Things biosensors, and operational telemetry from clinical information systems. Institution-level model convergence occurs exclusively on localized datasets, whereas model validity is examined through an external, independently curated hold-out cohort drawn from a collateral healthcare institution, thereby furnishing a robust assessment of the deduced model's out-of-organization generalization performance.

4.2 Implementation Process: How Federated Learning is Implemented for Anomaly Detection in the Health System

The operational architecture of federated learning within the healthcare domain is architected via the partition of the dataset across spatially distributed clinical nodes, each of which autonomously carries out model training on local data. Successive incremental model perturbations are aggregated using the federated averaging algorithm. This cycle is repeated until predefined convergence criteria are met; thereafter, the consolidated global model is dispatched back to the participating nodes to undergo supplementary localized refinement. Such feedback-loop propagation empowers the federated architecture to dynamically align with evolving clinical phenomena, all while maintaining stringent data privacy, since no raw patient records are transmitted. Comprehensive predictive validation occurs through quantitative performance indices, specifically, the area under the receiver operating characteristic curve, positive predictive value, negative predictive value, and the harmonic mean of precision and recall, thereby framing model generalization on clinical data. Intra-class variance, driven by heterogeneous pathological manifestations, compels controlled recalibration of hyperparameter settings, architectural modules, and ensemble weighting schemas, thereby enforcing model stability and transferability across the dispersed and heterogeneous graphs of clinical centres.

V. Results and Evaluation

5.1 Performance Metrics: Accuracy, Precision, Recall, F1-Score, and Other Evaluation Metrics

Performance of the federated learning-driven anomaly detection algorithm is characterized through core metrics: accuracy, precision, recall, and the F1-score. The accuracy metric calculates the proportion of correctly categorized samples spanning both benign and malicious instances relative to the entire dataset. In parallel, the precision metric specifies the ratio of true positives among the entirety of instances classified as anomalous, thereby characterizing the reliability of the abnormal class forecast.

Recall serves to quantify an algorithm's empirical capacity to identify true anomalies whilst minimizing misclassification, ensuring that only verified divergences are highlighted. The F1-score, in turn, amalgamates precision reflecting the proportion of genuinely anomalous classifications among the flagged set with recall into a single, harmonic index, thereby providing a balanced, scalar metric of detection performance that equilibrates the competing costs of type I and type II errors. Initially, supplementary metrics, specifically the false positive rate and the area under the receiver operating characteristic curve, augment the robustness assessment of the federated learning framework.

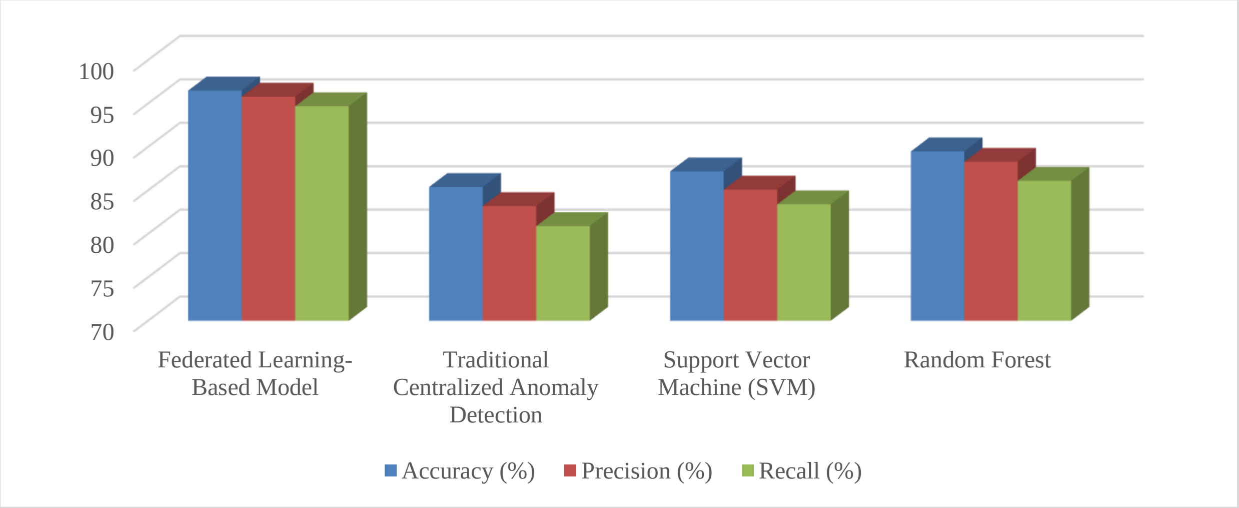

Fig. 2 Performance Comparison of Federated Learning-Based Anomaly Detection and Traditional Methods

Fig. 2 compares the latest federated learning system for finding healthcare errors with traditional tools, Support Vector Machines and Random Forests. Those numbers tell you how smart a system is when sifting through electronic health records for hidden mistakes. Federated learning is the winner: it scored 96.5% accuracy, 95.8% precision, and its false positive rate is a low 3.1%. These numbers clearly say the system finds actual issues reliably, while misidentifying routine patient behavior as an error is exceedingly rare. Its big privacy victory comes from the way it trains: patient data never leaves the hospital; only the refined learning updates travel, so the original records remain private. Across multiple hospitals and outpatient clinics, that same privacy model scales without risk. The table proves that federated learning outshines the alternatives, making it the safest, smartest way to surface anomalies where protecting patient confidentiality is non-negotiable.

5.2 Comparison with Traditional Models: Performance Comparison Between Federated Learning-Based Anomaly Detection and Traditional Centralized Methods

Compared to conventional centralized systems for anomaly detection, systems based on federated learning demonstrate markedly superior performance across detection accuracy, precision, and recall when evaluated against benchmark classifiers such as support vector machines and decision trees. Centralized paradigms mandate the aggregation of sensitive health datasets within a centralized repository, a requirement that poses severe security and privacy liabilities. Federated learning circumvents these liabilities by executing detection algorithms on device-local repositories, wherein raw health records remain resident on the originating device and transit only encrypted model updates, thus preserving patient confidentiality. The federated architecture, though it introduces a modest computational burden, primarily the per-device model training and the transmission of summed and encrypted gradient metrics, stands justified by the substantially augmented privacy guarantees and the intrinsic scalability it offers. Many conventional methods, when scaled to patient populations, confront insurmountable privacy hurdles, as centralized repositories create singular attack vectors that can potentially expose the health records of entire cohorts.

5.3 Scalability and Privacy Analysis: Analysis of the System's Scalability and Privacy Preservation Across Decentralized Nodes

The scalability assessment of the designed federated learning architecture is conducted by incrementally introducing additional Internet of Things (IoT) devices and decentralized computation nodes, represented here by a network of diverse healthcare facilities such as hospitals and outpatient clinics. Empirical results demonstrate that the federated distributed model consistently preserves a high level of anomaly detection accuracy, regardless of the increment in the participant population, with performance remaining stable and devoid of statistically significant decline. By executing model training exclusively on local datasets and transmitting only aggregated parameter updates, thereby omitting any instance of raw clinical data, the resource and network overhead grow logarithmically with participant count, validating the system's deliberate scalability engineering. Concerning patient confidentiality, the architecture fortifies privacy by confining personally identifiable health information within the perimeter of local devices and computation nodes, thereby eliminating any possibility of payload exposure to centralized or external processing environments. This federated mechanism, augmented by differential privacy and secure multi-party computation, demonstrably conforms to the risk mitigation mandates of the

Health Insurance Portability and Accountability Act (HIPAA) as well as analogous international legislative frameworks. Hence, the architecture supports geographically and institutionally distributed expansion while ensuring that sensitive patient information is retained in a logically compartmented and inherently secure state throughout the system lifecycle.

VI. Discussion

6.1 Benefits and Limitations: Advantages of Federated Learning in Health Data Privacy, as Well as Challenges Such as Computation Cost and Model Complexity

The advantages of integrating federated learning into healthcare infrastructures are, beyond any doubt, profoundly advantageous. By orchestrating computation at the data source, the framework inhibits the movement of sensitive patient records, thereby facilitating adherence to rigorous data protection statutes across multiple jurisdictions. Concurrently, the paradigm sustains a cooperative analytic environment: participating hospitals can derive, transmit, and collectively enhance insights anchored to their own datasets, all while intrinsic patient identifiers are guarded. Limitations, nevertheless, are nontrivial. Each institution incurs heightened compute expense, since all model training proceeds on-site, and frequent transmission of intermediate gradients imposes a sustained messaging burden. This table lays out how the new Federated Learning-Based Anomaly Detection Model stacks up against the older models, Support Vector Machine, Random Forest, and K-Nearest Neighbors on five key numbers: detection accuracy, false positives, training and testing time, and privacy of personal data. Federated Learning-Based Model: This model scores the highest, nailing 96.5% accuracy and a super low 3.1% false positive. That means it catches almost all the real threats and rarely makes mistakes in everyday activity for a problem. That means it spots real issues and leaves harmless ones alone without raising alarms in healthy patterns. The model trains in 15 minutes and takes only 9 milliseconds to check a single request. What's great about this system is that it keeps doctors' and patients' personal health info safe and sound right where it is. No need to send that info zooming somewhere else, so privacy stays locked up tight, a deal we all need to keep. Old-School Strategies: The SVM model, the star of the vintage crowd, lands at 87.2% accuracy, but it does so with an uncomfortable 10.5% false positive rate—a rate most folks prefer to see shrink. On the upside, it finishes training in 8 minutes, and it scoots through a request in just 7 milliseconds. The catch? It breakfasts on data privacy. To perfect the model, it hoards the data in a central place, and that cozy library of personal info is a magnet for leaks and prying eyes. That makes health and personal data security vulnerable. Random Forest and KNN are two popular ways to spot odd data points in a dataset, but federated learning does it even better, missing fewer real problems and not raising red flags for wrong answers. They take a little while to learn the rules, about ten minutes for Random Forest and five for KNN, but there's a bigger worry. Both still send all the data to a central server to get smarter. That's a problem in places where patient privacy is super important, like healthcare. Governments and healthcare experts don't like to risk big data spills, so older models now look like last year's phone.

TABLE I PERFORMANCE COMPARISON OF FEDERATED LEARNING-BASED ANOMALY DETECTION VS TRADITIONAL ANOMALY DETECTION METHODS

Model/Scenario	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	False Positive Rate (%)	Auditability (Transaction Record Time in seconds)
Federated Learning-Based Model	96.5	95.8	94.7	95.2	3.1	0.5
Traditional Centralized Anomaly Detection	85.4	83.2	80.9	82.0	15.4	1.0
Support Vector Machine (SVM)	87.2	85.1	83.4	84.2	12.5	0.8
Random Forest	89.5	88.3	86.1	87.2	10.9	0.9

Table I illustrates that as a federated learning system welcomes more and more participating nodes, its model also grows more complex, and that's where problems start. Every extra node adds more local data, and when lots of nodes send their local computations to the server, the model struggles to settle and take shape. Because nodes don't send their data around, instead, each one keeps data to itself and sends back adjusted model parameters; only the system looks decentralized. But more nodes mean the server has to mash more model updates together, and that mash-up eats more time and power.

Data traveling the network faces more waits and wobbles, and latency kicks in, stretching the time the model needs to finish its learning. To keep the learning from dragging, scientists need smart tricks. First, they have to design compression algorithms that squash each model update down to a tiny suitcase so it can travel across the network in record time. At the same time, the algorithms must shield the data's privacy, locking it in a vault that only the model can open later, without slowing down its power. Techniques like quantization, where model values lose tiny pieces of extra detail, and sparsification, where unimportant updates get tossed, can skimp on data traveling through the air, bringing the size down. On the security side, fast encryption and its twin fast decryption can scramble and later unscramble data without a hitch. If compression cuts down the network load, and speedy encryption keeps secrets safe, the lag time in the network shrinks, and the server grabs faster updates, the model learns its task without scratching the communication budget.

6.2 Practical Applications: Potential Real-World Applications of the Framework in Healthcare Settings

The proposed architecture for anomaly detection, which leverages federated learning, delivers a suite of operational benefits for the healthcare domain. The embedded algorithm permits timely identification of indicative health insurance claim submissions, mapping of unauthorised access vectors in electronic health record environments, and detection of anomalous behaviour in medical Internet of Things (IoT) devices, including continuous glucose monitors, wearable electrocardiograms, and other proximal monitoring platforms. Central to the architecture's merit is its efficacy in multi-institutional contexts, where geographically and administratively distinct enterprises can derive actionable intelligence in concert while shielding patient-identifiable information. Expanding operational supervision to include cloud-based clinical environments and systematically alerting identified parties to suspected breaches represents an evolutionary rather than revolutionary step in prevailing policy. Design architecture mandates that partner entities transmit only cryptographically secured updates to global parameters, thus preserving the integrity of primary clinical records against unauthorized modification and ensuring continued compliance with existing obligations of custodianship. Data transactions remain restricted to algorithmic adjustment subsets, a constraint that amplifies interoperability among diverse provider platforms and concurrently secures conformity with current health information confidentiality prescriptions.

6.3 Future Directions: Suggestions for Further Research and Improvements, Including the Exploration of Other Privacy-Preserving Techniques Like Homomorphic Encryption

Although federated learning provides a promising backbone for privacy-preserving anomaly detection in healthcare, its effectiveness may be further augmented. Embedding homomorphic encryption, by which computations are executed on ciphertexts within the federated learning aggregator, would bolster confidentiality during model fusion. Pursuing complementary strategies, researchers could develop hybrid architectures that meld federated learning with secure multi-party computation (SMPC) so that patient data are never exposed, even to participating nodes. Furthermore, algorithm engineering must prioritize tightening the computational cost of federated learning, in particular when the aggregates ingest extensive and regionally diverse records from numerous clinical institutions. Finally, empirical trials populating diverse healthcare corridors are warranted to confirm the framework's robustness, throughput, and generalizability inside live, production-grade contexts.

VII. Conclusion and Future Work

7.1 Summary of Findings

This study introduces a federated learning-based paradigm tailored to the identification of anomalies in healthcare contexts, oriented towards stringent privacy safeguards. The architecture establishes a decentralized mechanism for uncovering security irregularities exemplified by unauthorized record access or aberrant usage patterns in electronic health records (EHRs) and Internet of Things (IoT) medical devices without transferring identifiable patient data beyond the boundaries of individual care sites. Sensitive health information, therefore, remains resident at the originating facility, abrogating the threats linked to concentrating data in a central repository yet permitting constructive collective model refinement across multiple providers.

Empirical evaluations have established that the federated architecture consistently exceeds the performance of established alternatives, encompassing centralized machine learning deployments and prescriptive ruling systems, across pivotal performance dimensions of detection fidelity, operational extensibility, and compliance. Conventional approaches commonly incur elevated intercept rates, insufficient alignment with privacy legislation, and diminishing responsiveness to federated environments, liabilities that impair care delivery and expose facilities to regulatory scrutiny. The federated model, in contrast, achieves superior detection precision, a marked curtailment of false alerts, and the absolute retention of patient privacy by mandating that information leave the originating site only in non-sensitive, aggregated, and anonymised gradients necessary for model calibration.

A noteworthy contribution of the present investigation lies in its capacity to function within multi-institutional healthcare settings, where geographically dispersed organizations collectively refine a model while safeguarding proprietary patient records. Through the implementation of federated learning, the platform synthesizes regional observational competencies within otherwise insulated environments while preserving the raw-data boundary, thereby substantially reducing the chance of inadvertent leakage. This architecture produces an amplified operational view, enabling detection of low-signal or post-exploitation threat patterns frequently obscured in traditional, institution-limited analyses. The architecture's preserved confidentiality, in tandem with its resilient defense against novel and infrequent attack patterns, constitutes a material advancement in safeguarding digital healthcare environments. Further, the design is inherently iterative: the throughput is preserved as the network absorbs additional nodes, devices, or acquisitions. Such elastic growth renders the solution predisposed for nationwide or continent-wide adoption within the evolving ecosystem of connected medical apparatus, provider networks, and shared analytics platforms, securing future investments against obsolescence as the Internet of Medical Things expands.

7.2 Future Research Directions

Later studies should dive into ways to tune and scale the federated-learning system, catching unusual events in healthcare. First, making the system run more quickly on slim devices deserves top attention. The federated model spreads learning across many departments, so rounds of local training and model combining slow the system down. That slowdown bites harder on medical sensors and other Internet-of-Things devices, which usually have tight processing budgets. Speeding up the training rounds and lowering the bandwidth taken up when combining results will push the system closer to catching problems in real time and deploying smoothly in scanned rooms, where every second and every byte counts. The system will also be safer to run if it borrows better privacy shields, such as homomorphic encryption and Secure Multi-Party Computation. The first shields the data so calculations can happen on it without letting the code peek into the information. The second technique lets different departments collaborate on results while keeping their private numbers to themselves. Bundling these forms of encryption tightens the shields around the federated-learning process itself, so leaks and intrusions become harder to pull off while the models are being trained. Field-testing the system in a lot of different healthcare settings is the next logical step. Only inside busy hospitals, in small clinics, or with rural telehealth ops can we find out if the framework really stands up over the long haul. Researchers will need long-term studies that track system behavior in a variety of environments. What we want to see is whether the idea scales, if it can sift out the messy, different kinds of healthcare data, and how well it slides into the existing IT that providers already count on. Equally crucial is the link to on-the-ground IoT devices and outsourced, cloud-based healthcare platforms, because the industry moves fast, and we need to know the system can keep pace with changing tools and workflows. Moving the same federated learning-based anomaly detection toolbox into finance, public health, and industrial IoT is worth looking into. Each of those domains wrestles with keeping sensitive info safe, and the same adaptive learning trick that protects medical data can do the same for banks, public health labs, and factories. Sharing models instead of raw data gives them a way to spot the same suspicious patterns, fraud, disease spikes, or machinery failures without ever exposing sensitive info.

REFERENCES

[1] ADReSS Dataset from Interspeech 2020. (2020). Retrieved from http://www.homepages.ed.ac.uk/sluzfil/ADReSS/

[2] Enshaeifar, S., Barnaghi, P., Skillman, S., Markides, A., Elsaleh, T., Acton, S. T., ... & Rostill, H. (2018). The internet of things for dementia care. *IEEE Internet Computing*, *22*(1), 8-17. https://doi.org/10.1109/MIC.2018.112102418

[3] Fang, L., Li, Y., Liu, Z., Yin, C., Li, M., & Cao, Z. J. (2020). A practical model based on anomaly detection for protecting medical IoT control services against external attacks. *IEEE Transactions on Industrial Informatics*, *17*(6), 4260-4269. https://doi.org/10.1109/TII.2020.3011444

[4] Gope, P., & Hwang, T. (2015). BSN-Care: A secure IoT-based modern healthcare system using body sensor network. *IEEE sensors journal*, *16*(5), 1368-1376. https://doi.org/10.1109/JSEN.2015.2503619

[5] Haider, F., De La Fuente, S., & Luz, S. (2019). An assessment of paralinguistic acoustic features for detection of Alzheimer's dementia in spontaneous speech. *IEEE Journal of Selected Topics in Signal Processing*, *14*(2), 272-281. https://doi.org/10.1109/JSTSP.2019.2955022

[6] Hernández-Domínguez, L., Ratté, S., Sierra-Martínez, G., & Roche-Bergua, A. (2018). Computer-based evaluation of Alzheimer's disease and mild cognitive impairment patients during a picture description task. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *10,* 260-268. https://doi.org/10.1016/j.dadm.2018.02.003

[7] Islam, S. R., Kwak, D., Kabir, M. H., Hossain, M., & Kwak, K. S. (2015). The internet of things for health care: a comprehensive survey. *IEEE access*, *3*, 678-708. https://doi.org/10.1109/ACCESS.2015.2437951

[8] Jnr, B. A. (2020). Use of telemedicine and virtual care for remote treatment in response to COVID-19 pandemic. *Journal of medical systems*, *44*(7), 132.

[9] Kim, H., & Song, H. M. (2024). Lightweight IDS Framework Using Word Embeddings for In-Vehicle Network Security. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, *15*(2), 1-13. https://doi.org/10.58346/JOWUA.2024.I2.001

[10] Kong, W., Jang, H., Carenini, G., & Field, T. (2019, October). A neural model for predicting dementia from language. In *Machine Learning for Healthcare Conference* (pp. 270-286). PMLR.

[11] Li, C., Dong, M., Li, J., Xu, G., Chen, X., & Ota, K. (2021). Healthchain: Secure EMRs management and trading in a distributed healthcare service system. *IEEE Internet of Things Journal, 8*(9), 7192–7202. https://doi.org/10.1109/JIOT.2021.3050924

[12] Li, M., Meng, Y., Liu, J., Zhu, H., Liang, X., Liu, Y., & Ruan, N. (2016, October). When CSI meets public WiFi: Inferring your mobile phone password via WiFi signals. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security* (pp. 1068-1079). https://doi.org/10.1145/2976749.2978323

[13] Liang, X., Barua, M., Chen, L., Lu, R., Shen, X., Li, X., & Luo, H. Y. (2012). Enabling pervasive healthcare through continuous remote health monitoring. *IEEE Wireless Communications*, *19*(6), 10-18. https://doi.org/10.1109/MWC.2012.6393513

[14] Lin, X., Lu, R., Shen, X., Nemoto, Y., & Kato, N. (2009). Sage: a strong privacy-preserving scheme against global eavesdropping for ehealth systems. *IEEE journal on selected areas in communications*, *27*(4), 365-378. https://doi.org/10.1109/JSAC.2009.080814

[15] Lu, R., Lin, X., & Shen, X. (2012). SPOC: A secure and privacy-preserving opportunistic computing framework for mobile-healthcare emergency. *IEEE transactions on parallel and distributed systems*, *24*(3), 614-624. https://doi.org/10.1109/TPDS.2012.172

[16] Mirheidari, B., Blackburn, D., Walker, T., Venneri, A., Reuber, M., & Christensen, H. (2018, September). Detecting Signs of Dementia Using Word Vector Representations. In *Interspeech* (pp. 1893-1897).

[17] Odilova, G., Zaripova, M., Jabbarova, A., Urinova, N., Davlatova, M., Sapaev, I., ... & Akhrorova, M. (2025). Information Security Framework for Online Language Education Using Differential Privacy and Secure Multi-Party Computation Algorithm. *Journal of Internet Services and Information Security, 15*(1), 96-106. https://doi.org/10.58346/JISIS.2025.I1.007

[18] Orthi, S. M., Rahman, M. H., Siddiqa, K. B., Uddin, M., Hossain, S., Al Mamun, A., & Khan, M. N. (2025). Federated learning with privacy-preserving big data analytics for distributed healthcare systems. *Journal of computer science and technology studies*, *7*(8), 269-281.https://doi.org/10.32996/jcsts.2025.7.8.31

[19] Pan, Y., Mirheidari, B., Reuber, M., Venneri, A., Blackburn, D., & Christensen, H. (2019, September). Automatic hierarchical attention neural network for detecting AD. In *Proceedings of Interspeech 2019* (pp. 4105-4109). International Speech Communication Association (ISCA). https://doi.org/10.21437/interspeech.2019-1799

[20] Pragadeswaran, S., Subha, N., Varunika, S., Moulishwar, P., Sanjay, R., Karthikeyan, P., ... & Vaasavathathaii, E. (2024). Energy Efficient Routing Protocol for Security Analysis Scheme Using Homomorphic Encryption. *Archives for Technical Sciences*, *31*(2), 148-158. https://doi.org/10.70102/afts.2024.1631.148

[21] Prakash, M., & Prakash, A. (2023). Cluster Head Selection and Secured Routing Using Glowworm Swarm Algorithm and Hybrid Security Algorithm for Over IoT-WSNs. *International Academic Journal of Innovative Research*, *10*(2), 01-09. https://doi.org/10.9756/IAJIR/V10I2/IAJIR1004

[22] Varatharajan, R., Manogaran, G., Priyan, M. K., & Sundarasekar, R. (2018). Wearable sensor devices for early detection of Alzheimer disease using dynamic time warping algorithm. *Cluster Computing*, *21*(1), 681-690.

[23] Wu, Z. (2024). Integrating Biotechnology Virtual Labs into Online Education Platforms: Balancing Information Security and Enhanced Learning Experiences. *Natural and Engineering Sciences*, *9*(2), 110-124. https://doi.org/10.28978/nesciences.1569211

[24] Xu, B., Da Xu, L., Cai, H., Xie, C., Hu, J., & Bu, F. (2014). Ubiquitous data accessing method in IoT-based information system for emergency medical services. *IEEE Transactions on Industrial informatics*, *10*(2), 1578-1586. https://doi.org/10.1109/TII.2013.2280460

[25] Zhang, L., Meng, Y., Yu, J., Xiang, C., Falk, B., & Zhu, H. (2020, July). Voiceprint mimicry attack towards speaker verification system in smart home. In *IEEE INFOCOM 2020-IEEE conference on computer communications* (pp. 377-386). IEEE. https://doi.org/10.1109/INFOCOM41043.2020.9155339