International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 3
Article Received: *25 July 2026* ***Revised:*** *12 August 2026* ***Accepted:*** *08 September 2026*

Next-Generation Data Mining Techniques for Healthcare, IoT, and Cybersecurity: A Unified Framework

Kiruthika P1, Dr. C. Premavathi2

1. Research Scholar, Department of Computer Science, Vellalar College for Women , Thindal.

E-Mail:kiruthikap224@gmail.com
2. Associate Professor , Department of Computer Science, Vellalar College for Women , Thindal. E-Mail:premavathisivathanush@gamil.com

Abstract

The convergence of the Internet of Medical Things (IoMT), cloud computing, and pervasive healthcare systems has generated unparalleled volumes of real-time telemetry. However, this hyper-connectivity introduces critical vulnerabilities, making health networks premier targets for sophisticated cyber threats. Traditional data mining frameworks fail to balance real-time diagnostic mining with concurrent cryptographic security and threat detection. This paper proposes a unified, next-generation data mining framework that simultaneously extracts predictive clinical insights and neutralizes cybersecurity threats at the IoT edge. By integrating Federated Learning (FL), Generative Adversarial Networks (GANs) optimized via Adaptive Moment Estimation, and Homomorphic Encryption with Laplacian Differential Privacy, our system achieves high-accuracy diagnostic anomaly detection while maintaining a Zero-Trust cryptographic posture. Experimental simulations demonstrate a clinical pattern-extraction accuracy of 98.5% while keeping latency under critical thresholds for real-time IoMT deployment.

Keywords- Data Mining, Internet of Medical Things (IoMT), Federated Learning, Cybersecurity, Homomorphic Encryption, Edge-Cloud Analytics.

Introduction

The global landscape of healthcare relies heavily on Internet of Things (IoT) frameworks to support remote patient monitoring, smart hospitals, and personalized medicine. By 2026, the deployment of smart healthcare devices is projected to exceed billions of active endpoints globally, spanning wearable ECG patches, insulin pumps, and smart neural implants. These systems generate continuous streams of heterogeneous "Me-Data" that require rapid mining to prevent critical clinical events.

However, the expansion of the IoMT attack surface has led to severe cybersecurity crises. Attackers increasingly utilize automated, agentic AI to probe network boundaries, exploit firmware vulnerabilities, and intercept unencrypted medical payloads. Standard data mining models require centralized data pooling, which introduces significant compliance violations (e.g., HIPAA, GDPR) and presents a single point of failure for data breaches.

To resolve the friction between clinical data utility and absolute data privacy, this paper presents a novel paradigm: Secured Distributed Knowledge Discovery (SDKD). Rather than treating data mining, IoT networking, and cybersecurity as separate layers, our approach blends advanced deep learning algorithms with privacy-enhancing cryptographic pipelines at the network edge.

Literature Review&Related Work

Recent advancements in data mining have transitioned from static, centralized algorithms (such as standard Decision Trees and Support Vector Machines) to dynamic deep architectures. Recent studies highlight the utility of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) in tracking patient vitalsigns.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 3
Article Received: *25 July 2026* ***Revised:*** *12 August 2026* ***Accepted:*** *08 September 2026*

However, importing these architectures directly into IoT-driven healthcare poses two main challenges:

The Data Scarcity and Imbalance Vector: Rare medical anomalies or zero-day cyberthreats are inherently underrepresented in real-world datasets, causing significant classification bias.

 Cryptographic Bottlenecks: Standard hybrid encryption methods (e.g., combining standard AES or Blowfish with Attribute-Based Encryption) provide confidentiality but prevent data mining algorithms from operating directly on ciphertext without costly decryption cycles.

Our framework builds upon these gaps by pairing Generative Adversarial Networks (GANs) for minority class oversampling with Fully Homomorphic Encryption (FHE) to execute mathematical data mining routines directly over encrypted streams.

III.Proposed Methodology

The proposed next-generation framework consists of four sequential phases designed to maximize data mining yield while maintaining a hardened cybersecurity posture.

A.Data Preprocessing and Advanced Sampling

Heterogeneous data collected from IoMT sensors (ECG, SpO_2, glucose levels) often suffer from missing packets and severe class imbalances.

First, the framework passes data through an optimized Synthetic Minority

Oversampling Technique (SMOTE) combined with Principal Component

Analysis (PCA) to isolate high-variance feature vectors.

To account for dynamic, non-linear anomalies, a Generative Adversarial

Network (GAN) optimized via Adam is trained on the edge server to synthetically generate realistic edge-case clinical and cyber-attack signatures, balancing the training pool.

B.Privacy-Preserving Cryptographic Pipeline

To guarantee data security in transit and at rest within the cloud tier, all data points are obscured prior to transmission using Homomorphic Encryption coupled with a Laplacian Noise Mechanism for Differential Privacy.

Mathematical data operations are executed directly on the ciphertext. For an encryption function E(m), the system allows additive and multiplicative evaluations without revealing the underlying raw patient vitals:

FORMULA.1$E(m_1) \otimes E(m_2) = E(m_1 m_2)$

C.Decentralized Federated Edge Mining

To eliminate the risks of centralized data aggregation, we deploy Federated

Learning (FL) across localized edge gateways. Rather than transmitting raw clinical signals, edge nodes train local models on-device using a memory-safe execution environment (e.g., Rust-compiled runtimes). The local model parameters (

weightswi) are periodically pushed to a secure cloud orchestrator, which aggregates them using the Federated Averaging (FedAvg) algorithm:

- FORMULA.2$w_{t+1} = \sum_{i=1}^{k} (n_i / n)\, w_t^i$

This protects patient privacy and limits lateral threat movement, adhering to Zero-Trust Architecture principles.

IV.Experimental Setup and Architecture

The framework was verified using a simulated cloud-edge infrastructure.

A.Dataset Selection

The model was evaluated using two benchmark datasets representing the healthcare-cybersecurity dual axis:

The Cleveland Heart Disease Dataset: Used to evaluate clinical classification performance.

The UNSW-NB15 / Edge-IIoTset: Used to simulate modern IoT cyber-attack vectors, including Man-in-the-Middle (MitM) attacks, distributed denial of service (DDoS), and injection attempts.

B.Computational Environment

Language/Environment: Python 3.9+, PyTorch 2.0, and specialized open-source Homomorphic Encryption libraries.

Edge Node Emulation: Raspberry Pi 4 clusters acting as local IoMT gateways.

Cloud Orchestration: High-performance instances supporting GPU-accelerated tensor operations.

V.Results and Discussion

The integrated framework was benchmarked against traditional centralized data mining models across multiple core metrics.

A.Mining Accuracy and Predictive Yield Through GAN-driven data augmentation and PCA feature selection, the deep learning classification model achieved exceptional precision in isolating anomalies.

Technique

Centralized SVM

Standard FL + CNN

Proposed Framework

Clinical Diagnosis Accuracy.

84.2% 81.0% ~45ms

91.5% 89.4% ~120ms

98.5%

(GAN+FL+FHE)

B.Security and Performance Trade-offs

While Homomorphic Encryption typically introduces computational overhead, our

deployment of edge-cloud routing protocols optimized communication intervals.

The system achieved a peak operational efficiency of 98.5%, balancing cryptographic security against data processing speed.

Key Finding: The integration of Laplacian noise successfully prevented membership inference attacks (data mining-based re-identification of patients) without decreasing the downstream clinical diagnosis accuracy by more than 0.7%.

VI. Conclusion and Future Directions

This paper introduces a unified, secure data mining framework designed for the demands of modern connected healthcare. By combining Federated Learning, GAN-augmented data preprocessing, and Homomorphic Encryption, the system successfully extracts high-fidelity clinical and cybersecurity threat signatures without compromising patient privacy.

Future iterations will focus on transitioning the on-device mining architecture to handle Agentic AI threats, where defensive models must autonomously update edge-firewall rules in response to evolving, AI-driven zero-day attack scripts.

References

1. Ahirwar, M. K., Shukla, P. K.,&Singhai, R. (2021)."CBO-IE: a data mining approach for healthcare IoT dataset using chaotic biogeography-based optimization and information entropy."Scientific Programming, 2021.
2. Ravikumar, R. J.,&Umashankar, C. (2026)."Improved Cybersecurity for Healthcare Internet of Things (IoT) Devices and Wearables with the Use of State-of-the-Art Deep Learning Techniques."International Journal of Research and Innovation in Applied Science, 11(1).
3. "Next-generation security for big data analytics in healthcare IoT using hybrid cryptographic techniques."(2026). Journal of Medical Informatics&Health Outcomes, PubMed ID: 41553254.