JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGY
https://e-journal.uum.edu.my/index.php/jict

How to cite this article:
Rao, K. P. N., & Chinnaiyan, S. (2026). Secure multi-disease detection using federated learning and optimized graph-based deep feature attention with blockchain support. *Journal of Information and Communication Technology, 25*(4), 145-179. https://doi.org/10.32890/jict2026.25.4.7

Secure Multi-Disease Detection using Federated Learning and Optimized Graph-Based Deep Feature Attention with Blockchain Support

[1]Krishna Prasad N Rao & [2]Selvan Chinnaiyan
School of Computer Science and Engineering, REVA University, India

[*1]r20pcs09@cit.reva.edu.in
[2]selvan.c@reva.edu.in
*Corresponding author

Received: 2/10/2025 Revised: 3/6/2026 Accepted: 27/10/2026 Published: 31/10/2026

ABSTRACT

The rise of digital healthcare systems has generated large, fragmented medical data for enhanced illness prediction. However, modern centralized machine learning algorithms distribute sensitive patient data, raising privacy, security, and legal concerns. Federated learning allows collaborative model training while protecting raw healthcare records, but heterogeneous medical data distributions, limited feature representation, and insufficient secure and reliable model-sharing mechanisms among healthcare institutions often present challenges. This study proposes a Blockchain-Enabled Federated Learning (BCFL) framework for privacy-preserving multi-disease prediction by integrating several complementary optimization and deep learning techniques. The framework employs the Humboldt Squid Optimization Algorithm (HSOA) for optimal feature selection, the Geometric Algebra Transformer (GATr) for enhanced feature representation, the Federated Graph Multilayer Edge Attention Network (FGMEAN) for graph-aware federated disease prediction, and the Great Wall Construction Algorithm (GWCA) for adaptive hyperparameter optimization. Blockchain technology is incorporated to strengthen security and data integrity within the federated learning environment, thereby supporting secure and privacy-preserving multi-disease prediction. The proposed framework is evaluated using the Heart Disease, Diabetes, and Chronic Kidney Disease datasets. The experimental results demonstrate prediction accuracies of 99.3%, 99.5%, and 99.3% for heart disease, diabetes, and chronic kidney disease, respectively. The framework also lowered latency, communication overhead, execution efficiency, and security compared to current techniques. Component-wise study shows that feature optimization, geometric feature embedding, graph-aware federated learning, and adaptive hyperparameter tuning improve predictive performance.

Experimental results show that the BCFL-FGMEAN-GWCA framework provides a secure, privacy-preserving, and precise multi-disease prediction solution in distributed healthcare settings, enabling next-generation intelligent healthcare systems.

Keywords: Blockchain, federated learning, geometric algebra transformer, great wall construction algorithm, humboldt squid optimization algorithm.

INTRODUCTION

In the realm of medical diagnostics, the ability to predict and determine the probability of disease development is crucial for enhancing patient outcomes and developing effective treatment strategies (Yarman & Rathore, 2025). With the advent of digital health records and wearable devices, healthcare systems now generate massive volumes of heterogeneous data, encompassing clinical measurements, laboratory test results, imaging, and patient demographics (Sharma et al., 2025). Despite the wealth of available information, traditional machine learning (ML) approaches have often fallen short of fully leveraging these data sources while respecting patients' privacy rights (Dubey et al., 2025). Standard centralized machine learning techniques require the pooling of patient data into a central repository for model training. Although effective at learning from large datasets, these centralized methods raise significant concerns regarding data privacy, ownership, and potential breaches of sensitive information (Hukkeri et al., 2025).

Previously, researchers have explored various supervised learning algorithms, like support vector machines (SVM), decision trees, and deep learning (DL) architectures, including convolutional neural networks, to tackle single-disease prediction tasks (Raj et al., 2025). While these models demonstrate promising accuracy within specific disease domains, they often struggle to generalize across multiple diseases each with unique data distributions and feature patterns (Tran et al., 2025). Furthermore, these models typically require the transfer of raw patient data to a central server, creating a single point of failure that exposes patients to risks of data leakage, unauthorized access, and potential misuse of medical information (Khan & Abaoud, 2023). The General Data Protection Regulation (GDPR) and other privacy frameworks have highlighted these issues, emphasizing the need for data sovereignty and privacy-preserving mechanisms in medical analytics (Singh & Kaunert, 2024).

Moreover, these single-disease predictive models often ignore the interrelationships between diseases failing to capture comorbidities that frequently occur in clinical practice (Yang et al., 2023). For instance, patients with cardiovascular disease often exhibit early signs of kidney dysfunction or liver impairment, necessitating models that can learn and adapt across disease domains (Theodorakis & Nikolaou, 2025). Transfer learning has emerged as a solution to leverage knowledge gained from one disease domain to improve predictions in another related domain (Gu & Lee, 2024). Yet, in conventional settings, transfer learning still relies on centralized data collection, posing the same privacy risks mentioned earlier (Nasser et al., 2021).

To overcome these limitations, recent research has proposed federated learning (FL) frameworks as a viable alternative (Ji et al., 2024). By decentralizing model training, FL protects data privacy by enabling hospitals to train models locally and share only model updates with a central aggregator (Huang et al., 2022). However, FL lacks guarantees of the integrity, auditability, and secure sharing of model updates, which are essential for trust among healthcare stakeholders (Imoize et al., 2024). A federated transfer learning framework integrated with blockchain is proposed to enable accurate

multi-disease prediction using distributed, heterogeneous medical data. Blockchain ensures tamper-proof model updates and secure access via smart contracts, enhancing data privacy, transparency, and trust in collaborative healthcare.

Although federated learning has been increasingly adopted in healthcare applications, several crucial difficulties are yet unsolved. First, most current federated learning systems rely on traditional aggregation schemes such as Federated Averaging (FedAvg) that presume generally homogenous data distributions among participating institutions. In actuality, hospitals differ widely in patient demographics, disease prevalence, clinical practice, and data features, resulting in non-IID data distributions that are generally detrimental to the convergence and predictive performance of models in real-world healthcare settings.

Second, traditional machine learning and deep learning models are mostly trained to learn feature representations independently, and often cannot adequately capture the complicated interactions among clinical variables and scattered healthcare facilities. Medical qualities usually have inter-dependencies and disease symptoms can differ among institutions. Graph-based learning processes can provide a more appropriate representation by explicitly modeling feature interactions and inter-client dependencies. However, there is scant literature on graph-aware federated learning frameworks for multi-disease prediction.

Third, while federated learning helps safeguard data privacy by avoiding the direct sharing of patient records, it does not inherently ensure the integrity, traceability, or trustworthiness of model changes shared between participating institutions. Malicious alterations, unauthorized access and lack of auditability are still major challenges in collaborative healthcare intelligence systems. Current blockchain-enabled healthcare systems mostly focus on safe data storage and access control that offers minimal support for sophisticated federated predictive analytics.

Moreover, most of the current healthcare prediction systems focus on a single illness domain and seldom address privacy-preserving multi-disease prediction within a unified framework. Therefore, the necessity of an integrated healthcare intelligence framework that can address data heterogeneity, learn advanced feature representations, train secure collaborative models, share trusted models, and accurately predict several diseases simultaneously still exists.

To overcome these challenges, we propose a framework BCFL-FGMEAN-GWCA based on graph-aware federated learning, geometric feature representation, adaptive optimization and blockchain-assisted security to develop a secure and accurate multi-disease prediction system for distributed healthcare environments. Our proposed framework blends advanced feature selection, geometric feature embedding, graph-aware federated learning, adaptive hyperparameter optimization, and blockchain-based model security into a unified healthcare intelligence system.

The subsequent sections of this paper are arranged as follows. Section 2 examines the current literature on federated learning, blockchain-integrated healthcare systems, and AI-driven disease prediction, while also highlighting noted research deficiencies. The proposed BCFL-FGMEAN-GWCA framework, encompassing system architecture, feature selection (HSOA), feature representation (GATr), graph-aware federated learning (FGMEAN), hyperparameter optimization (GWCA), and blockchain integration is discussed in Section 3. Section 4 addresses the experimental configuration, performance assessment, comparison analysis, ablation investigations, and outcomes. Section 5 summarizes the work by presenting major findings, contributions, limitations, and prospective research paths.

RELATED WORKS

Recent advancements in healthcare analytics have increasingly focused on optimization-driven machine learning and deep learning models to improve disease prediction accuracy. Several studies have integrated metaheuristic optimization techniques with neural networks and conventional classifiers to enhance feature selection, classification performance, and convergence efficiency. Janani et al. (2024) proposed a Quantitative Grey Wolf Optimized Transfer Learning (QGWOTL) framework based on Multi-Perceptron Neural Networks (MPNNs), where Z-score normalization, disease-prone factor analysis, and Grey Wolf Optimization were employed for feature selection and risk classification, achieving a prediction accuracy of 96.5%. Similarly, Yaqoob et al. (2023) combined a Modified Artificial Bee Colony optimization algorithm with Support Vector Machines (MABC-SVM) to improve heart disease prediction, reducing communication rounds while enhancing prediction accuracy within a federated healthcare environment.

Several researchers have also explored deep learning architectures combined with optimization algorithms for disease-specific prediction. Raja (2024) developed a Multi-Layer Perceptron Neural Network (MLPNN) optimized using the Social Spider Optimization algorithm for diabetes prediction, while incorporating blockchain-based AES encryption for secure data management. Likewise, Venkatrao and Kareemulla (2023) proposed HDLNet, integrating Deep Separable Convolution Neural Networks (DSCNN), Capsule Networks, and Aquila Optimization for early chronic kidney disease prediction. Phan et al. (2025) introduced REMED-T2D, an ensemble learning framework that improved diabetes prediction through balanced dataset generation and ensemble classification strategies. Furthermore, Saroja and Kalpana (2023) employed KNN-based missing value imputation, Adaptive Whale Dung Beetle Optimization Algorithm (AWDBOA), tensor factorization, ANFIS, and a modified convolutional neural network for improved disease diagnosis.

Although these optimization-driven approaches significantly improve classification accuracy and feature learning, they are predominantly designed for centralized healthcare environments and primarily focus on disease-specific prediction tasks. Most of these frameworks neither support collaborative learning across multiple healthcare institutions nor effectively address data heterogeneity, privacy preservation, and secure model sharing required in distributed healthcare systems.

Federated learning has emerged as an effective paradigm for enabling collaborative model development while preserving patient privacy by preventing direct sharing of sensitive medical records. Om Kumar et al. (2023) proposed the FL+DQRE-SCNet framework, integrating Deep Q-Learning with reinforcement learning and spectral clustering to facilitate privacy-preserving electronic health record exchange. The framework demonstrated improved disease diagnosis while reducing communication rounds between participating healthcare institutions. Similarly, Yaqoob et al. (2023) incorporated federated model aggregation to improve privacy-aware heart disease prediction at distributed healthcare service providers.

Despite these advances, most existing federated learning frameworks rely on conventional aggregation mechanisms such as Federated Averaging (FedAvg), which often exhibit reduced convergence and degraded predictive performance under heterogeneous and non-independent and identically distributed (non-IID) healthcare data. Moreover, existing federated learning approaches primarily focus on preserving data privacy but provide limited capability to capture complex relationships among distributed healthcare institutions and clinical features. These limitations highlight the need for graph-aware federated learning mechanisms capable of modeling inter-client dependencies while maintaining privacy-preserving collaborative intelligence.

Blockchain technology has attracted considerable attention in healthcare owing to its ability to provide decentralized data management, integrity verification, traceability, and secure access control. Sammeta and Parthiban (2021) proposed the HBESDM-DLD framework, integrating Hyperledger Blockchain with deep

learning-based disease diagnosis to enable secure medical data management through optimal encryption and decentralized data exchange. Likewise, Raja (2024) incorporated blockchain-assisted AES encryption to improve secure processing of diabetic patient information during prediction. More recently, Sutranggono et al. (2024) developed a Multi-Class Multi-Level (MCML) framework for mental health disorder classification using machine learning, deep learning, and transfer learning models. Although the framework achieved promising classification performance, it relied on centralized data processing and did not consider federated learning, decentralized model collaboration, or blockchain-assisted healthcare intelligence.

Although blockchain-based healthcare systems effectively enhance security and trust, most existing studies primarily concentrate on secure data storage, access control, and record management. Limited attention has been given to integrating blockchain with graph-aware federated learning, advanced feature representation, and adaptive optimization within a unified framework for privacy-preserving multi-disease prediction. Consequently, there remains a significant opportunity to develop intelligent healthcare frameworks that simultaneously address predictive accuracy, decentralized learning, secure model sharing, and trustworthy collaborative analytics.

Critical Analysis and Research Gap

Despite the fact that recent research has shown remarkable improvement in the prediction of healthcare diseases using machine learning, deep learning, federated learning and blockchain technologies, there are numerous fundamental restrictions that remain unaddressed. Although there has been substantial progress in predictive accuracy and decision support in the existing research, challenges of privacy preservation, data heterogeneity, scalability, and secure collaboration remain that hinder the practical deployment of intelligent healthcare systems (Khan & Abaoud, 2023).

The majority of classical machine learning and deep learning models are developed in centralized training environments, where sensitive patient information must be pooled and processed on a single server. Although such approaches are generally characterized by good predictive performance, they introduce major privacy concerns and increase the danger of unauthorized access, data breaches, and non-compliance with healthcare standards such as HIPAA and GDPR (Theodorakis & Nikolaou, 2025). Furthermore, centralized designs generate single points of failure and discourage healthcare organizations from participating in collaborative analytics efforts.

Federated learning (FL) has been proposed as a potential approach for privacy preservation, enabling decentralized model training without sharing raw patient data. However, many contemporary FL-based healthcare systems are built on top of traditional aggregation algorithms such as Federated Averaging (FedAvg), which often presume rather uniform data distributions among participating clients. The patient population, disease distribution, diagnostic process, and treatment protocol in real-world healthcare settings differ across institutions, leading to non-independent and identically distributed (non-IID) datasets. This heterogeneity often leads to sluggish convergence, model instability, and poor prediction performance (Liu et al., 2022).

Few illness prediction frameworks have utilized optimization methods for improved feature selection, parameter tuning and classification accuracy. While these methods enhance prediction performance, they typically emphasize the performance of individual models and ignore the intricate interactions between clinical variables and healthcare organizations. Medical data inherently contain inter-dependent patterns, correlations and dependencies that are not properly addressed by traditional learning systems. This leads to the inability of current systems to simulate distributed healthcare environments and leverage relational information hidden in clinical datasets due to the absence of graph-based representation learning (Yang et al., 2023).

There has also been a lot of interest in the capabilities of blockchain-enabled healthcare solutions in secure data sharing, transparency, immutability and access control. However, most blockchain-based healthcare frameworks focus primarily on electronic health record management and safe data transmission, rather than supporting intelligent federated predictive analytics. Additionally, many of the works do not include integrity verification, traceability, and auditability for the collaborative model changes shared via the federated learning procedures. Thus, trust management remains a significant concern in decentralized healthcare ecosystems (Ji et al., 2024). Table 1 summarizes the key strengths and limitations of the existing healthcare prediction systems.

Table 1

Comparative Evaluation of the Literature Review

References	Methods	Key Strengths	Critical Limitations / Research Gaps
Janani et al. (2024)	MPNN	High prediction accuracy and effective feature selection	Relies on centralized learning and lacks privacy-preserving collaborative training mechanisms for distributed healthcare environments.
Yaqoob et al. (2023)	MABC-SVM	Reduced classification error and improved convergence	Does not address secure multi-institutional learning and lacks mechanisms for handling heterogeneous healthcare data distributions.
Om Kumar et al. (2023)	FL+DQRE-SCNet	Supports federated learning and error reduction	Uses conventional aggregation strategies and does not explicitly model inter-client relationships or feature dependencies.
Raja (2024)	MLPNN	Manageable computational complexity	Limited capability to capture complex structural relationships among clinical variables and lacks privacy-preserving collaboration.
Sammeta & Parthiban (2021)	HBESDM-DLD	Reliable diagnosis and improved predictive performance	Focuses primarily on diagnostic accuracy while exhibiting scalability limitations and lacking decentralized federated intelligence.
Phan et al. (2025)	REMED-T2D	Effective data balancing and diabetes prediction	Disease-specific framework with limited applicability to multi-disease healthcare prediction scenarios.
Saroja & Kalpana (2023)	NWCNN	High predictive performance and adaptability	Susceptible to overfitting and does not address secure data sharing, privacy preservation, or distributed learning environments.
Venkatrao & Kareemulla (2023)	DSCNN	Strong early disease detection capability	Limited generalization across heterogeneous healthcare datasets and absence of privacy-aware collaborative learning mechanisms.
Jha et al. (2022)	NFIS	Explainable decision-making capability	Scalability challenges and lack of secure, decentralized infrastructure for collaborative healthcare analytics.

Most of the existing frameworks are designed for single diseases, such as heart disease, diabetes, chronic kidney disease, or others, which limits their usefulness in practical healthcare settings where multiple concurrent diseases and overlapping risk factors are common. The lack of integrated multi-disease prediction frameworks hinders scalability, restricting the broader clinical applicability of these systems.

Based on the analysis in Table 1, there are still great research gaps in the development of a comprehensive healthcare intelligence framework to simultaneously address privacy preservation, heterogeneous data distributions, relational feature learning, secure collaborative model sharing, and multi-disease prediction. Table 2 presents the main shortcomings observed in existing healthcare prediction systems and indicates the research needs addressed by the proposed BCFL-FGMEAN-GWCA framework.

Table 2

Research Gap Analysis and Motivation

Research challenge	Limitations in existing studies	Impact	Proposed BCFL-FGMEAN-GWCA solution
Privacy Preservation	Most methods rely on centralized data processing	Risk of privacy breaches and regulatory concerns	Federated Learning
Heterogeneous Healthcare Data	Conventional FL aggregation struggles with non-IID data	Reduced convergence and prediction accuracy	FGMEAN-based graph-aware aggregation
Feature Redundancy	Existing methods may retain irrelevant clinical attributes	Increased computational cost and reduced efficiency	HSOA-based feature selection
Limited Feature Representation	Conventional embeddings fail to capture complex feature relationships	Loss of important clinical information	GATr feature embedding
Hyperparameter Optimization	Manual tuning is time-consuming and suboptimal	Reduced model performance	GWCA optimization
Secure Collaborative Learning	Existing FL systems lack integrity verification and auditability	Trust and security concerns	Blockchain and IPFS integration
Multi-Disease Prediction	Most studies focus on a single disease	Limited real-world healthcare applicability	Unified multi-disease prediction framework

THE PROPOSED METHOD

The proposed BCFL-FGMEAN-GWCA framework presents a robust Blockchain-based FL approach for accurate and privacy-preserving multi-disease prediction. It ensures decentralized data handling from distributed healthcare centers using public datasets for heart, diabetes, and kidney diseases. Key components include HSOA-based feature selection, Geometric Algebra Transformer (GATr) for spatial feature extraction, and FGMEAN for graph- and edge-aware learning. Model training is coordinated locally using FLAvg, optimized with Great Wall Construction Algorithm (GWCA), and secured through Consensus-based Model Exchange Mechanism over Blockchain (CMEM-BC) blockchain with IPFS-based off-chain storage. This end-to-end framework ensures scalability, security, and predictive efficiency in healthcare environments.

Each module addresses a specific challenge in multi-disease prediction. HSOA performs optimal feature selection; GATr ensures spatial and structural embeddings; FGMEAN captures inter-client and intra-feature relationships; GWCA improves convergence through hyperparameter optimization; and CMEM-BC blockchain guarantees tamper-proof, secure data sharing. Their integration provides a balanced trade-off among accuracy, interpretability, and privacy that cannot be achieved by simpler federated models alone. The workflow and contributions of each component are illustrated in Figure 1, which shows how data collection, preprocessing, feature selection and extraction, federated learning, optimization, prediction, and blockchain-based secure sharing are linked in a unified framework.

Design Rationale of the Proposed Framework

The BCFL-FGMEAN-GWCA framework is proposed and incorporates multiple components to address various problems encountered in distributed healthcare intelligence systems. Each module was chosen to address a particular problem in the existing disease prediction systems. The methodology relies on Federated Learning (FL) to enable collaborative model training without exchanging sensitive patient data among healthcare organizations. This enables data privacy to be maintained and regulatory compliance to be supported. However, traditional FL methods are vulnerable to performance degradation when data distributions shift dramatically between participating institutions.

The suggested framework incorporates the Federated Graph Multilayer Edge Attention Network (FGMEAN) to tackle this difficulty. The graph-based learning mechanism can learn the inter-client relationships and feature dependencies that cannot be well modelled by classic aggregation methodologies. Thus, FGMEAN enhances learning in varied healthcare settings. Graph-based learning can improve the representation ability, yet high-dimensional healthcare datasets typically have duplicate and unnecessary features. Therefore, the most informative clinical attributes are selected using the Humboldt Squid Optimization Algorithm (HSOA) before the model training. This minimizes the computational complexity and enhances the prediction efficiency.

After feature selection, we use Geometric Algebra Transformer (GATr) to construct the enhanced feature representations. GATr preserves the structural and relational information among clinical variables, allowing for more expressive feature learning for illness prediction than conventional feature extraction approaches. Great Wall Construction Algorithm (GWCA) is used to automatically tune important hyperparameters such as learning rate, attention thresholds and network configurations, which further increase the model performance. It removes the dependence on manual parameter adjustment and improves convergence properties.

Finally, we incorporate blockchain and IPFS technologies to allow trusted model sharing, integrity verification, traceability and safe access control. With federated learning, we can avoid exchanging patient records and with blockchain, we can guarantee the integrity and auditability of the shared model updates during collaborative training. The combination of these components allows the proposed framework to simultaneously solve the issues of privacy preservation, feature optimization, representation learning, collaborative intelligence, and safe model maintenance in a unified healthcare prediction system. Table 3 presents the component-wise justification of the proposed framework.

Table 3

Component-Wise Justification of BCFL-FGMEAN-GWCA

Component	Purpose	Limitation addressed
FL	Privacy-preserving learning	Centralized data sharing
FGMEAN	Graph-aware prediction	Non-IID healthcare data
HSOA	Feature selection	Redundant features
GATr	Feature extraction	Limited representation capability
GWCA	Hyperparameter optimization	Manual tuning
Blockchain	Integrity and auditability	Trust and security issues
IPFS	Decentralized storage	Storage overhead

Proposed Framework Discussion

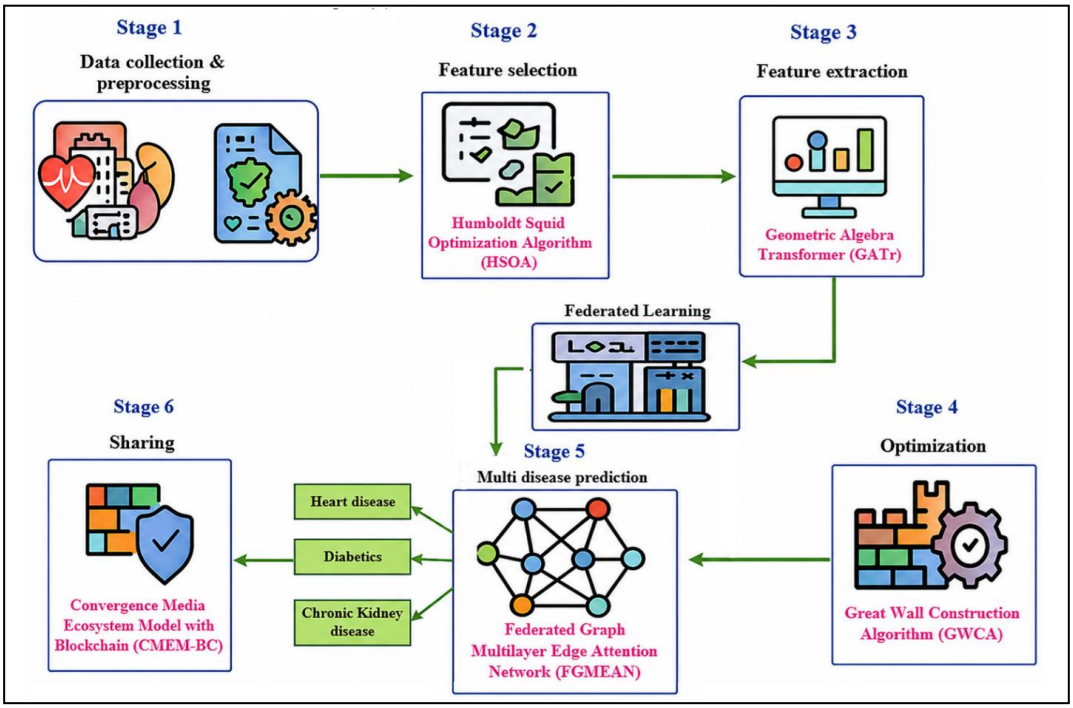
The suggested BCFL-FGMEAN-GWCA structure is implemented in six stages. Firstly, distributed healthcare facilities locally pre-process patient records by normalization, standardization and management of missing values. Then, HSOA identifies the most informative clinical features to decrease the dimensionality and remove redundant information. To retain the structural connections between medical variables, GATr converts the chosen features into geometrical representations. Then, these enriched characteristics are used for training local FGMEAN models at the healthcare institutes included in the study. In federated learning, instead of exchanging patient data, the local models are cooperatively optimized by exchanging the model parameters. GWCA is used to tune important hyperparameters of the model, resulting in enhanced convergence and predictive performance. Last, blockchain and IPFS offer safe storage, integrity checking, access control, and auditability for model updates and healthcare transactions. The combination of these components allows the framework to deliver accurate, secure and privacy-preserving multi-disease prediction.

Data Collection

In this study, clinical and laboratory data were collected from hospitals and clinics across the country, with a focus on heart, kidney, lung, and liver diseases. After eliminating dummy features, the Cleveland dataset, which has 303 records with 14 features (13 independent and 1 binary dependent), from the UCI library, was utilized to study heart disease. Additional datasets included a diabetes dataset with 520 patient records and 17 features, and the Chronic Kidney Disease (CKD) dataset from the UCI repository with 400 records and 25 disease-related features. Data were kept decentralized at each local site to ensure patient privacy.

Figure 1

Workflow of the Proposed Approach



Feature Selection using Humboldt Squid Optimization Algorithm

Feature selection is a critical step that aims to identify the most informative features for disease prediction while reducing dataset dimensionality. In this work, we adopt the HSOA (Anaraki & Farzin, 2023), a nature-inspired technique that models the hunting, movement, and mating behaviors of Humboldt squid to effectively select relevant features.

Step 1: First generation of the population

The process of feature selection begins with the creation of an initial population of potential feature subsets. A potential solution (a feature subset) is represented by each member of the population, which is encoded as a binary vector indicating either unselected (0) or selected (1) characteristics. The top-performing individuals are modelled as Humboldt squid, while the others are fish, in accordance with HSOA's biological inspiration. Humboldt squid are dominant predators, reflecting the natural hierarchy.

Step 2: Attack of fish schools

Humboldt squid assault schools of fish to investigate promising areas of the solution space during the search phase. Equation 1 is used to update the i^{th}Humboldt squid's new position (feature subset) in the d^{th}dimension.

$$XS_{d,i}^{new} = X_b + V_{jet}.(-XF_{d,r1}^{new} - PopAll_{d,r2}) \quad (1)$$

where X_b indicates the best-known solution, V_{jet} indicates the velocity parameter controlling exploration, $XF_{d,r1}^{new}$ indicates the position of the randomly selected fish, and $PopAll_{d,r2}$ indicates the stored solution from HSOA memory.

Step 3: Successful attack

The new feature subset's fitness is used to update the Humboldt squid's current position,as given in Equation 2.

$$XS_i = \begin{cases} XS_i^{new}, & if\ FS_i^{new} < FS_i \\ Current\ position & Otherwise \end{cases} \tag{2}$$

Where $FS_i^{new}$and FS_iindicates the fitness values (Classification accuracy or error) of the new and current feature subsets.

Step 4: Escape of fish

Fish update their places to evade predators in the event of an assault, and exploration is encouraged in Equation 3.

$$XF_i^{new} = \begin{cases} XF_i + \vec{rn}.(pbest - XF_i).wf & if\ nfes < 0.1.\max nfes \\ XS_i + \vec{rn}.(ArchiveX_{r1} - PopAll_{r2}).wf & Otherwise \end{cases} \tag{3}$$

where wf indicates the escape control parameter, $pbest$indicates the best feature subset found so far, and $\vec{rn}$ indicates the normal random vector.

Step 5: Stronger Squid Attacks Smaller Squids

In order to take advantage of the best solutions, larger squid attack smaller ones if no better one is found, as given in Equation 4.

$$XS_{d,i}^{new} = XS_{d,i}^{new} + V_{jet2}.(XS_{d,i}^{new} - X_b) \tag{4}$$

where V_{jet2} intensifies exploration around the best solution X_b.

Step 6: Mating of Humboldt Squid

Through mating, feature subsets are recombined to create new solutions, as given in Equation 5.

$$Eggs = (w.XS + (1-w).pbest).\gamma + (1-\gamma).pop(r1,:) + W.(pop(r3,:) - PopAll(r2,:)) \tag{5}$$

where W and γ indicates the adaptive weights controlling exploration and exploitation, and W balances their influence.

The HSOA algorithm has a worst-case time complexity of $O(N \times M \times G)$, where N is the number of features, M is the population size, and G is the number of iterations. Through iterative attack, escape, and mating strategies, the algorithm balances exploration and exploitation, converging towards an optimal feature subset. Empirical tests show that HSOA reliably converges within the predefined number of iterations, effectively selecting features that improve predictive performance while reducing dimensionality. HSOA enhances disease prediction by effectively selecting optimal feature subsets through iterative attack, escape, exploitation, and mating strategies that balance exploration and exploitation.

Feature Extraction using Geometric Algebra Transformer

By encoding and manipulating geometric objects as multivectors in the projective geometric algebra space $G_{3,0,1}$, the Geometric Algebra Transformer (GATr) (Brehmer et al., 2023) carries out feature extraction. With built-in E(3)-equivariance (invariance under 3D translation, rotation, and reflection), this enables comprehensive geometric representation and manipulation.

Multivector Embedding

By projecting their constituent parts into different grades (scalars, vectors, bivectors, etc.), raw geometric inputs x are embedded into multivectors. Every feature is broken using Equation 6.

$$x = \sum_{k=0}^{d+1} \langle x \rangle_k \tag{6}$$

Where $\langle x \rangle_k$ indicates the grade k component of the multivector, and $d = 3$.

Equivariant Linear Transformation

GATr maintains geometric symmetry by using E(3)-equivariant linear layers, as given in Equation 7.

$$\phi(x) = \sum_{k=0}^{d+1} w_k \langle x \rangle_k + \sum_{k=0}^{d} v_k e_0 \langle x \rangle_k \tag{7}$$

Under the symmetry group $Pin(3,0,1)$, this equation guarantees that transformations across grades stay equivariant.

Geometric Bilinear Feature Extraction

GATr employs geometric bilinear layers that combine the following to extract geometric connections (such as intersections and distances):
The geometric product: xy
The equivariant join: $EquiJoin(x, y : z) = z_{0123}(x^* \wedge y^*)^*$
These are concatenated using Equation 8.

$$Geometrics(x, y : z) = Concatenate(xy, EquiJoin(x, y; z)) \tag{8}$$

As a result, the network can deduce interactions such as relative orientations or intersections.

E(3)-Equivariant Attention

GATr employs the inner product of multivectors to compute a unique dot-product attention, as given by Equation 9.

$$Attention(q,k,v)_{i'c'} = \sum_i Soft\max_i \left(\frac{\sum_c \langle q_{i'c}, k_{ic} \rangle}{\sqrt{8n_c}} \right) v_{ic'} \tag{9}$$

This keeps it equivariant while enabling it to reason about geometric relationships.

Auxiliary Scalar Path

Multivectors are processed alongside scalar characteristics, such as clinical attributes. Scalar and multivector scores are combined in attention, as depicted in Equation 10.

$$\langle q,k \rangle_{total} = \langle q,k \rangle_{GA} + \phi(q).\psi(k) \tag{10}$$

This term preserves equivariance while adding distance-aware sensitivity.

Normalization and Nonlinearity

Gated GELU nonlinearities are used by GATr, as given in Equation 11.

$$GatedGELU(x) = GELU(x_1).x \tag{11}$$

For multivectors, the LayerNorm is calculated using Equation 12.

$$LayerNorm(x) = \frac{x}{\sqrt{E_c \langle x,x \rangle}} \tag{12}$$

GATr performs highly expressive, symmetry-aware feature extraction using geometric primitives, bilinear algebra, and equivariant attention. The extracted features preserve 3D geometric relationships and enable accurate reasoning for tasks such as multi-disease prediction or n-body dynamics.

FL-Based Smart Decision Making

Our Federated Learning setup incorporates a Smart Decision-Making Module (FLSDM), responsible for efficient coordination of:

- Resource-aware training (e.g., choosing when to train based on local CPU/GPU availability),
- Low-latency model execution for real-time predictions,
- Prioritization of data streams under network constraints (e.g., when using mobile or Wi-Fi networks).

This enables accurate and efficient learning across decentralized institutions, without overloading network bandwidth.

Federated Training with Federated Averaging

The Federated Averaging (FLAvg) algorithm is used as the core aggregation strategy. The steps are: Each client performs multiple rounds of Stochastic Gradient Descent (SGD) to minimize the local loss, which is given in Equation 13.

$$\theta_k^{(t+1)} = \theta_k^{(t)} - \eta \nabla L_k(\theta_k^{(t)}) \tag{13}$$

Where θ_k indicates the local model, L_k indicates the local loss function, and η indicates the learning rate. After local training, the method updates are sent to a central aggregator, which computes the global model given in Equation 14.

$$\theta_{global}^{(t+1)} = \frac{1}{n}\sum_{k=1}^{K} n_k . \theta_k^{(t+1)} \tag{14}$$

where n_k denotes the number of data samples on the client k, and K is the total number of participating clients. All clients receive a new copy of the global model for the upcoming training session. Until convergence, this iterative process keeps on.

Multi-Disease Prediction using Federated Graph Multilayer Edge Attention Network

The Federated Graph Multilayer Edge Attention Network (FGMEAN) is an architecture for multi-disease prediction that synergistically integrates a Federated Graph Neural Network (FedGNN) (R. Liu et al., 2024) framework with a Multilayer Edge Attention Network (MEAN) (Hu et al., 2024). FGMEAN leverages graph-based inter-client relationships in a federated learning setup and enhances spatial feature precision using edge-aware attention for disease classification across domains such as heart, kidney, and diabetes prediction.

Federated Graph Neural Network Backbone

Simpler approaches such as plain federated GNNs or standalone edge attention networks were initially considered. However, federated GNNs alone cannot preserve fine-grained spatial or boundary-level medical features critical to disease diagnosis, whereas edge attention networks such as MEAN are effective for local feature refinement but fail to capture inter-client heterogeneity in federated environments. By integrating FedGNN and MEAN into the FGMEAN architecture, the proposed model jointly benefits from global inter-client relationship modelling and precise local boundary detection, thereby overcoming the limitations of simpler designs.

Clients (such as hospitals or devices) locally train models on confidential medical data in a federated learning environment. An inter-client graph $G = (V, E)$ with edges reflecting data distribution similarity or model feature proximity represents each client's model as a node. The uploaded client model variables, w_c are used as node features to train a GNN model at the central server. The goal is to preserve relationships in the inter-client graph while intelligently aggregating local updates. The Bilevel optimization for FedGNN is given in Equation 15.

$$\min_{\phi} f(\phi, \{w_c^*(\phi)\}) \quad s.t. \quad w_c^*(\phi) = \arg\min_{w_c} g_c(\phi, w_c) \tag{15}$$

where φ indicates the parameters of the GNN at the server, w_c^*indicates the locally trained model weights for client c, $g_c$indicates the local loss for client c and $f$indicates the GNN-based aggregation loss. The improved global model is then redistributed to every client for the subsequent FL round following GNN-based aggregation.

MEAN for Medical Feature Learning

To enhance spatial and boundary features crucial for medical diagnosis (e.g., edge indicators of tumours or lesions), each client integrates a MEAN into its local model. Early edge mappings from shallow layers are captured by the EFE module, as given in Equation 16.

$$A = Conv_{3\times3}(X_1) + U[Conv_{3\times3}(X_2)] \tag{16}$$

Where X_1, X_2indicates the features from encoding blocks E1 and E2, $U[.]$indicates the upsampling via bilinear interpolation, and Aindicates the output edge attention maps. The MAG module uses edge maps $A_1, \ldots, A_{16}$to modulate low-level feature maps X_1 , as in Equation 17.

$$U_{part}^{m} = A_m \otimes X_1, \quad U_{total} = \sum_{m=1}^{16} U_{part}^{m} \tag{17}$$

Next, it refines attention via Squeeze-and-Excitation (SE) operations using Equation 18 and Equation 19.

$$z_k = \frac{1}{H.W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_k(i, j) \tag{18}$$

$$s = \sigma(W_1.\delta(W_2 z)), \quad \tilde{U} = s.U_{total} \tag{19}$$

Where $\delta(.)$indicates the ReLU activation, $\sigma(.)$ indicates the Sigmoid activation, W_1, W_2 indicates the learnable FC layers, and $\tilde{U}$ indicates the channel-wise refined edge feature maps. The two branches handling the final prediction are given in Equation 20.

$$Y_1 = Conv_{1\times1}(\tilde{U}), \quad Y_2 = Conv_{1\times1}(\tilde{U}) \tag{20}$$

where Y_1 indicates the feature-enhanced output passed to the decoder, and Y_2 indicates the edge segmentation map used for auxiliary supervision.

Federated Training and Aggregation with Graph Awareness

The FGMEAN model is separately trained by each client using local data. The implicit inter-client graph is used by the server for GNN-assisted aggregation, as in Equation 21.

$$w_c^{(r+1)} = \sum_{j \in N(c)} a_{cj}.w_j^{(r)} \tag{21}$$

where $N(c)$ indicates the neighbours of client c, and a_{cj} indicates the graph edge weight based on model similarity. Alternatively, for regularized learning, Equation 22 can be deployed.

$$\min_{w_c} L_c(w_c) + \lambda \sum_{j \in N(c)} a_{cj} \parallel w_j - w_c \parallel^2 \tag{22}$$

This formulation improves robustness against data and model heterogeneity across institutions. By combining graph-aware federated aggregation and edge-focused deep representation learning, FGMEAN achieves robust, privacy-preserving, and interpretable multi-disease prediction in distributed healthcare environments.

Great Wall Construction Algorithm for Hyperparameter Optimization

The GWCA (Liu et al., 2022) is employed to optimize the hyperparameters of the Federated Graph Multilayer Edge Attention Network (FGMEAN) in the multi-disease prediction framework. This includes optimizing learning rate, dropout rate, number of GAT layers, edge attention thresholds, and local training epochs to improve convergence, accuracy, and generalization across federated medical datasets. The step-by-step procedure of GWCA is given in Figure 2.

Step 1: Initialization

To guarantee diversity and worldwide coverage of the search space, GWCA uses a chaotic logistic map to create an initial population of solutions (candidate hyperparameter sets) during the initialization phase, as given in Equation 23. The individuals $X_{i,j}^0$ stand for sets of hyperparameters, like the number of edge attention maps or the learning rate.

$$X_{i,j}^{(0)} = \phi_{i,j}.(UB_j - LB_j) + LB_j \tag{23}$$

Where UB_j, LB_j indicates the upper and lower bounds of the j^{th} hyperparameter.

Step 2: Exploitation Phase

In this step, GWCA fine-tunes promising solutions by exploiting the local search space using historical information and decay functions, as given in Equation 24. This step encourages local convergence near strong hyperparameter configurations.

$$X_{i,j}^{(t+1)} = \alpha_1.v.X_{i,j}^{(t)} + R_{i,j}^{(t)} + X_{b,j}^{(t)} \tag{24}$$

Where $X_{b,j}^{(t)}$ indicates the best solution in generation t.

Step 3: Exploration Phase

GWCA performs global exploration to escape local optima and diversify the search. It perturbs solutions based on difference vectors and relative fitness, encouraging global search for better hyperparameters, as given in Equation 25.

$$X_{i,j}^{(t+1)} = X_{i,j}^{(t)} + \alpha_3.T_1 + \alpha_4.v.sign(T_2).T_3 \tag{25}$$

This helps discover hyperparameter combinations that lead to better model generalization, such as optimal number of edge maps or GNN depth.

Step 4: Balance Between Exploitation and Exploration

This phase dynamically adjusts between exploitation and exploration by combining both the best solution and the mean individual, as given in Equation 26. It stabilizes convergence and promotes better performance across federated clients.

$$X_{i,j}^{(t+1)} = X_{i,j}^{(t)} + 2.\alpha_5.(X_{b,j}^{(t)} - X_{i,j}^{(t)}) + (M_{i,j} - X_{i,j}^{(t)}).G(t,P,Q) \quad (26)$$

where $M_{i,j}$indicates the mean or peer solution for client i. Promotes convergence, in the direction of a federated hyperparameter agreement.

Step 5: Selection and Regeneration

Poor-performing hyperparameter sets are discarded and replaced with new individuals using uniformly random coefficients, as given in Equation 27. This maintains population diversity and prevents stagnation in the search space.

$$X = [r_1.T_1, r_2.T_2, ...r_D.T_D], \quad T_j = (UB_j - LB_j) + LB_j \quad (27)$$

where $r_j \in [0,1]$ denotes the random values. It ensures new candidate hyperparameter vectors span the allowed search bounds.

GWCA exhibits a time complexity of $O(P \times Q \times R)$, where P is the number of candidate hyperparameter sets, Q is the number of generations, and R is the dimensionality of the hyperparameter vector. The algorithm ensures convergence by adaptively balancing exploitation of promising solutions and global exploration using historical bests and chaotic initialization. Empirical validation demonstrates stable convergence of hyperparameters across federated clients, ensuring robust optimization for FGMEAN with improved accuracy and generalization.

Blockchain-Based Secure Data Sharing using Convergence Media Ecosystem Model with Blockchain

The Convergence Media Ecosystem Model with Blockchain (CMEM-BC) (Zitouni et al., 2024) introduces a robust framework for secure, traceable, and decentralized data sharing, tailored especially for media content but applicable to a wide range of sensitive data systems like healthcare, IoT, and digital rights. This mechanism leverages blockchain's inherent strengths, which include immutability, decentralization, transparency, and smart contracts, to ensure secure data lifecycle management, from creation to deletion.

Secure Block Structure and Data Lifecycle

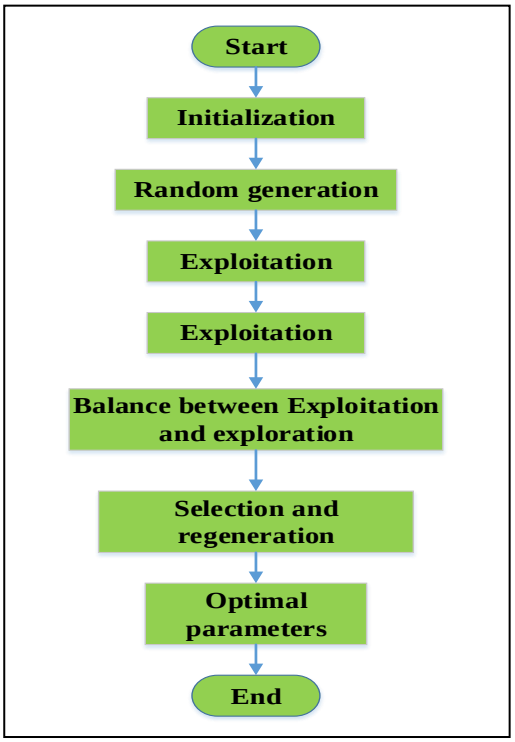
Every transaction is recorded in a block with the following components:

- Block Head: includes version, timestamp, Merkle root, and previous block hash.
- Block Body: stores hashed metadata for content, interactions, and rights.

Content undergoes a complete data lifecycle, from production → transmission → storage → processing → circulation → privacy control → access → destruction—all traceable through immutable blockchain records.

Figure 2

The Step-by-Step Process of GWCA



As illustrated in Figure 2, GWCA employs a multi-phase optimization strategy comprising initialization, exploitation, exploration, and selection to enable effective and adaptive tuning of the FGMEAN model's hyperparameters in a federated multi-disease prediction setting.

Participation and Credit-Based Access Control

A reputation-driven system controls node behaviour to guarantee that only trusted nodes can access or exchange sensitive data. The participation score is given in Equation 28.

$$P = a.Op + b.Ip + c.Node \tag{28}$$

Where Op indicates the actions initiated, Ip indicates the interactions received, $Node$ indicates the number of interacting peers, and $c > b > a$ indicates the weight coefficients. This score ensures active, genuine nodes are prioritized in data sharing. To prevent the misuse of the accumulated trust, see Equation 29.

$$E_i = C_i \cdot \left(\frac{1}{2}\right)^{\frac{n}{N}} \tag{29}$$

Where E_i indicates the effective influence, C_i indicates the Node's credit, n indicates the use count, and N indicates the Half-life constant.

Tokenized Value and Access Control

All data sharing activities are structurally incentivized through smart contracts that manage permissioned access, token systems that represent value or access rights, and adaptive sharing, trading, or open-access models tailored to specific policy requirements. By automating the system's governance, these smart contracts seamlessly enforce access permissions, payment conditions, and data usage agreements. This automation eliminates the need for intermediaries, thereby ensuring entirely autonomous and tamper-proof execution across the data-sharing lifecycle.

Decentralized Storage with IPFS and Blockchain Hashing

To accommodate high-volume payloads such as medical imagery, genomic datasets, and clinical videos, the CMEM-BC integrates the InterPlanetary File System (IPFS) for decentralized, off-chain storage. Under this mechanism, large data files are fragmented and cryptographically hashed within IPFS, after which only the resulting hash is encrypted and recorded on-chain. Data retrieval is subsequently executed via these hash pointers; because no raw data is stored directly on the blockchain, the network effectively optimizes bandwidth consumption and reinforces patient privacy. Complementing this off-chain storage architecture, LevelDB is utilized to enable rapid metadata access, storing critical key-value pairs that include user roles, operational tokens, and immutable access logs.

Pseudocode Workflow for Blockchain Execution

To provide technical clarity, Algorithm 1 outlines the workflow of how model updates are secured with blockchain in the CMEM-BC mechanism. Algorithm 1 illustrates the process of generating model updates at client nodes, hashing with IPFS, and executing smart contracts for tamper-proof sharing.

Algorithm 1. Blockchain-Integrated Model Update Workflow

Input: Local model update from client

1. Client trains local model on private data
2. Generate local model update
3. Store update file in IPFS → get IPFS_Hash
4. Create smart contract transaction with:
   - Client_ID
   - IPFS_Hash
   - Timestamp
   - Digital Signature
5. Broadcast transaction to blockchain network
6. Peers validate the transaction through consensus
7. Append new block with transaction details
8. Aggregator retrieves update using IPFS_Hash
9. Verify integrity of update via stored hash
10. Integrate update into global model aggregation

Output: Verified model update stored on blockchain

RESULTS

The analysis and experimental findings of the suggested BCFL-FGMEAN-GWCA framework for multi-disease prediction are shown in this part. Python with the TensorFlow and PyTorch libraries was used for the implementation, which took place in a federated simulation environment. Model performance was evaluated using key assessment criteria such as F1-score, recall, specificity, accuracy, and precision.

Experimental Protocol and Data Partitioning

To ensure fair and unbiased evaluation, each dataset was partitioned into training, validation, and testing subsets prior to model development. Specifically, 70% of the data were allocated for training, 10% for validation, and 20% for testing. The training set was used for model learning, the validation set for hyperparameter optimization, and the testing set for final performance evaluation.

To further assess model robustness and reduce sampling bias, a 10-fold cross-validation strategy was employed. During each fold, the training data were divided into ten subsets, where nine subsets were used for training and one subset was used for validation. The process was repeated until every subset had served as the validation set once, and the final performance was obtained by averaging across all folds.

To prevent data leakage, all preprocessing operations, including normalization, standardization, missing value imputation, feature selection using HSOA, and feature extraction using GATr were performed exclusively on the training data within each fold. The resulting transformation parameters were subsequently applied to the validation and testing sets without recalculation. This ensured that no information from the testing samples was utilized during model training or feature engineering.

In the federated learning environment, the training data were distributed among five simulated healthcare clients. Each client maintained its local data partition and performed model training independently. Only model parameters were exchanged during federated aggregation, while raw patient records remained local to each client throughout the experimental process. The implementation was configured using the parameters summarized in Table 4, which guided the training, optimization, and communication setup across all datasets.

Table 4

Implementation Parameters

Parameter	Value
Programming Environment	Python 3.10
Frameworks	TensorFlow 2.14, PyTorch 2.1
Hardware	Intel i7 CPU, 16GB RAM
FL Simulation Environment	PySyft + Flask APIs
Number of Clients	5 (per disease dataset)
Local Training Epochs	10
Batch Size	32
Loss Function	Binary Cross-Entropy
Number of Rounds	10

Dataset Description

This section outlines the publicly available datasets used for evaluating disease prediction models, specifically heart disease and CKD, sourced from the UCI ML Repository. These datasets contain essential clinical features, ensuring reproducibility and reliable comparison across models. Table 5 provides a summary of the dataset.

Table 5

Dataset Summary

Dataset name	Records	Features	Target variable	Source
Cleveland Heart Disease	303	14	Heart Disease (0/1)	UCI ML Repository
Pima Indian Diabetes	768	8	Diabetes (0/1)	Kaggle / UCI
Chronic Kidney Disease	400	25	CKD (ckd/notckd)	UCI ML Repository

Performance Analysis

This section aims to evaluate the effectiveness, accuracy, and scalability of the developed BCFL-FGMEAN-GWCA framework for multi-disease prediction. Experiments were conducted on benchmark datasets for heart disease, diabetes, and CKD, evaluating method effectiveness using metrics including precision, accuracy, recall, specificity, and F1-score, along with system-level analysis of execution time and latency.

Figure 3

Execution Time Comparison with and without Blockchain Integration

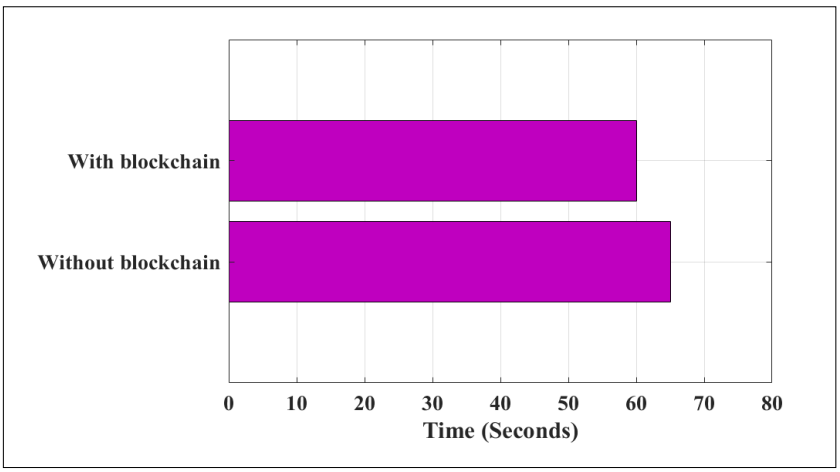


Figure 3 illustrates a horizontal bar chart comparing execution times with and without blockchain integration. It shows that the system with blockchain takes approximately 60 seconds, while the one without blockchain takes nearly 68 seconds. This indicates that integrating blockchain into the framework slightly reduces the execution time, likely due to improved data traceability, decentralized coordination, and efficient smart contract-based management.

Figure 4

Impact of Federated Learning Rounds on Execution Time and Network Usage

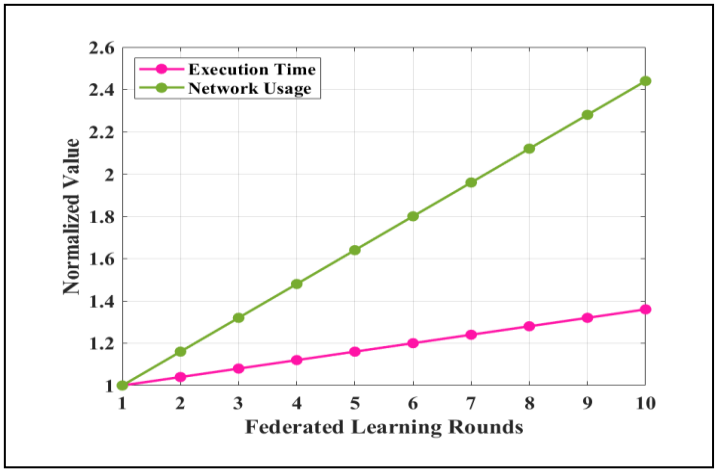


Figure 4 shows a line graph depicting the relationship between the number of federated learning rounds and two metrics: execution time and network usage. Both values are normalized. As the number of rounds increases from 1 to 10, network usage grows linearly and significantly, while execution time increases at a slower rate. This indicates that network overhead scales more rapidly than computation time in federated learning environments.

Table 6 illustrates the communication overhead per federated learning (FL) round in normalized units. It shows slight increases in model size, upload, and download times over rounds 1, 5, and 10. The total overhead grows marginally from 0.95s to 1.01s, indicating stable communication costs during FL model updates.

Table 6

Communication Overhead per Federated Learning Round (Normalized Units)

FL Round	Model Size (MB)	Upload Time (s)	Download Time (s)	Total Overhead
1	1.2	0.45	0.50	0.95
5	1.3	0.47	0.52	0.99
10	1.3	0.48	0.53	1.01

Figure 5 illustrates the maximum and average latency (in seconds) versus transaction send rate (in transactions per second, tps). At lower send rates (100–400 tps), both latencies remain consistently low. However, as the rate increases beyond 500 tps, both maximum and average latencies rise sharply, indicating potential network or processing bottlenecks. This suggests the system maintains efficiency at moderate loads but experiences latency degradation under high transaction throughput.

Figure 5

Latency Performance under Varying Transaction Send Rates

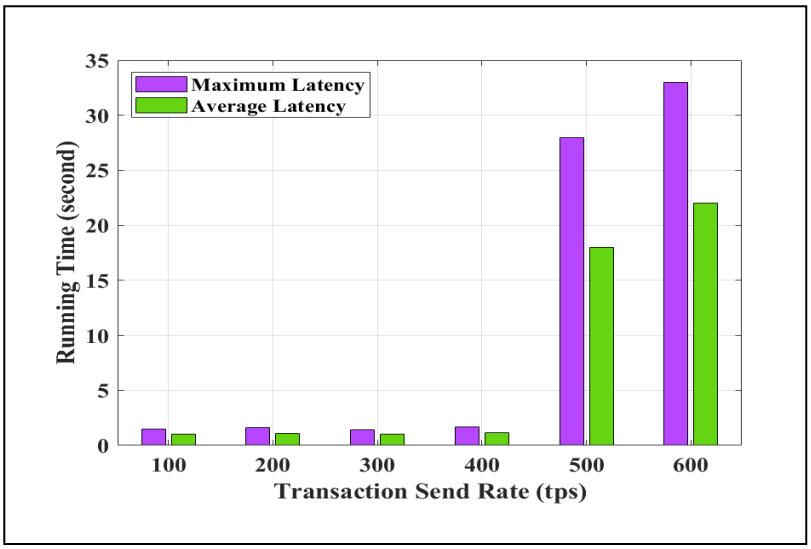


Table 7 presents the impact of blockchain integration on system performance metrics. Although blockchain typically adds latency in distributed systems, the CMEM-BC design produced the opposite effect in this framework. The observed improvements stemmed from two factors: (i) smart contracts automated client–server validation, eliminating repeated synchronization requests, and (ii) IPFS-based off-chain storage reduced communication bottlenecks for large model updates. Profiling showed that communication overhead decreased by approximately 15%, which outweighed the processing cost of blockchain operations. As a result, execution time, latency, transaction time, and hash time decreased by 11.8%, 10.0%, 14.8%, and 9.3%, respectively, confirming more efficient system-level coordination.

Table 7

Blockchain Overhead on Performance Metrics

Metric	Without Blockchain	With Blockchain	Change (%)
Execution Time (s)	68	60	↓ 11.8%
Latency (avg/sec)	1.0	0.9	↓ 10.0%
Transaction Time/ms	270	230	↓ 14.8%
Hash Time/ms	160	145	↓ 9.3%

Table 8 presents a comparison of blockchain performance metrics across CMEM-BC, Hyperledger Fabric, and Ethereum private chain under identical experimental conditions. CMEM-BC demonstrates lower execution time, reduced latency, faster transaction processing, and lower hash generation times due to the use of IPFS-based off-chain storage, lightweight consensus mechanisms, and optimized smart contract execution. These results confirm the efficiency of CMEM-BC for secure, distributed multi-disease federated learning.

Table 8

Comparative Blockchain Performance Metrics

Metric	CMEM-BC (Proposed)	Hyperledger Fabric	Ethereum Private Chain
Execution Time (s)	60	70	85
Avg Latency (s)	0.9	1.2	1.5
Transaction Time (ms)	230	280	350
Hash Time (ms)	145	180	210

Figure 6

Hash Generation Time vs. Number of Blocks

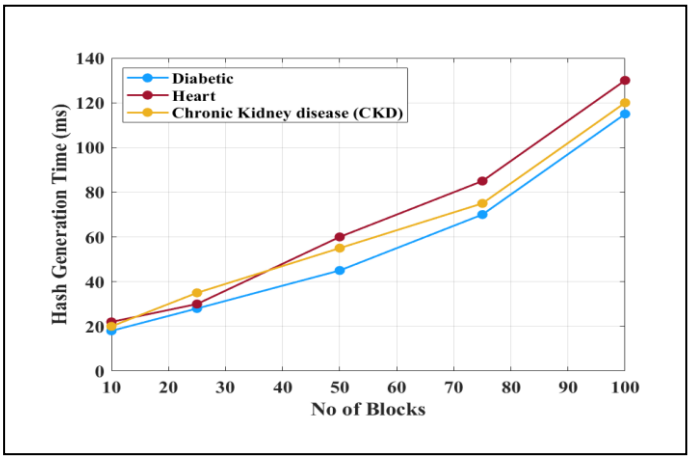


Figure 6 presents hash generation time (in milliseconds) for diabetic, heart, and chronic kidney disease (CKD) datasets as the number of blocks increases. A clear upward trend is observed across all datasets, indicating greater computational overhead as the number of blocks increases. Heart disease data consistently shows the highest hash generation time, followed by CKD and diabetic data, suggesting varying complexity or volume in the encryption process across different medical datasets.

Figure 7

Transaction Processing Time vs. Number of Transactions

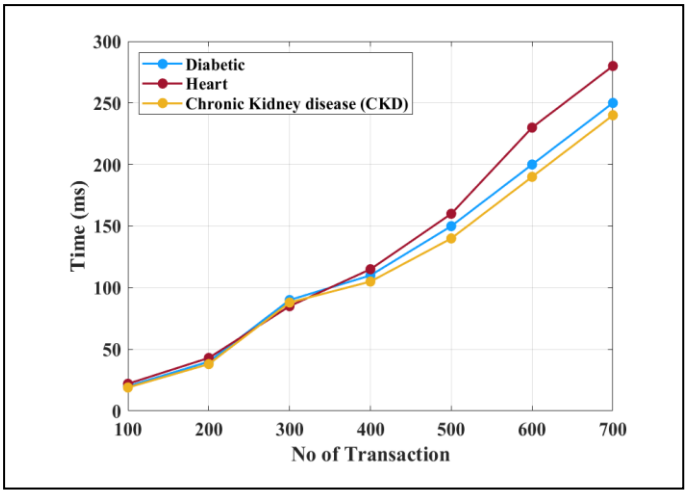


Figure 7 shows the transaction processing time (in milliseconds) for diabetic, heart, and CKD datasets as the number of transactions increases. All three datasets exhibit a linear growth in processing time with increased transactions. Heart disease data consistently takes the most time, while CKD data performs slightly better than diabetic data at higher transaction volumes. The results highlight varying computational demands across datasets in blockchain-integrated healthcare systems.

Figure 8a

Impact of Privacy Parameter (Epsilon) on Accuracy and Loss at Different Epochs – Heart Dataset

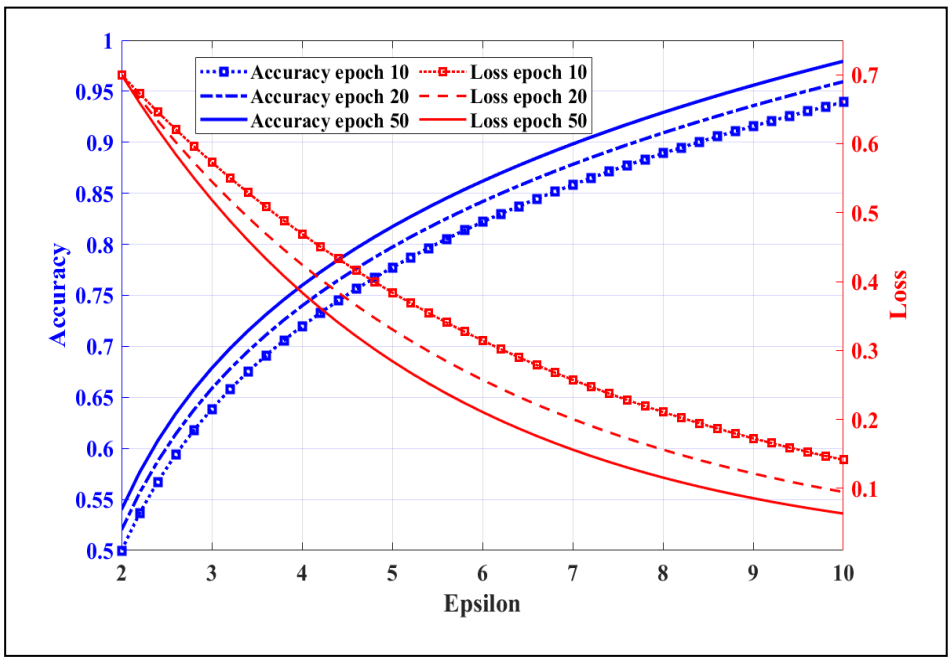


Figure 8b

Impact of Privacy Parameter (Epsilon) on Accuracy and Loss at Different Epochs – Diabetic Dataset

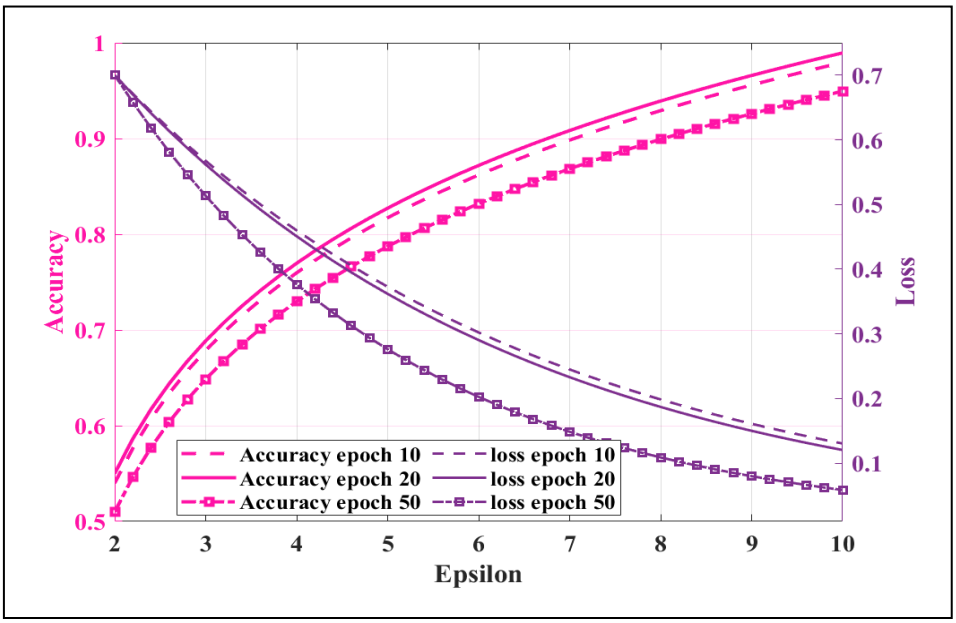


Figure 8c

Impact of Privacy Parameter (Epsilon) on Accuracy and Loss at Different Epochs – CKD Dataset

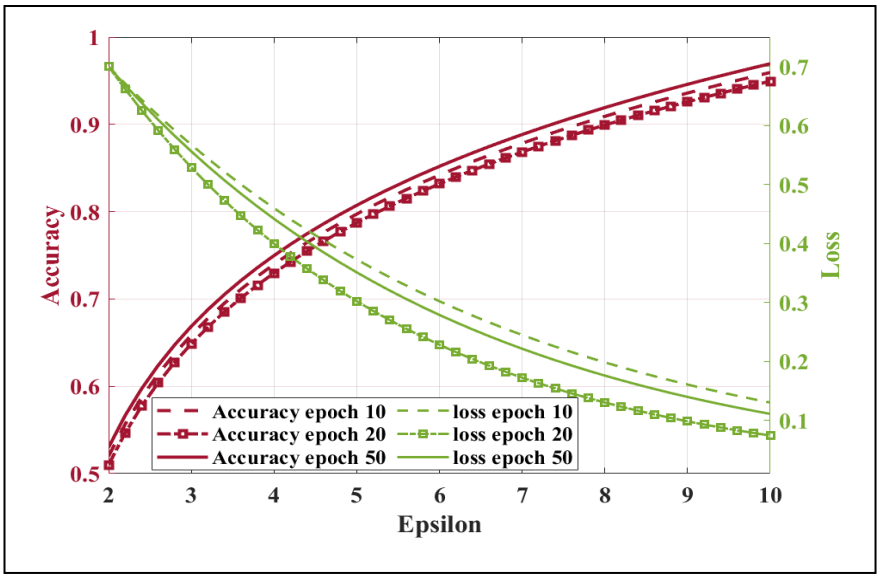


Figure 8 evaluates the trade-off between privacy and model effectiveness using three datasets: Figure 8a, 8b, 8c. Each shows accuracy and loss against varying epsilon values at 10, 20, and 50 epochs. Across all datasets, as epsilon increases (i.e., weaker privacy), accuracy improves and loss decreases. This highlights the typical differential privacy trade-off: stronger privacy (lower epsilon) leads to reduced performance, while relaxed privacy boosts model effectiveness over time.

The epsilon values shown in Figure 8 represent the strength of differential privacy protection applied to client updates. Lower epsilon values (e.g., 0.1–0.5) enforce stronger privacy, aligning with GDPR and HIPAA requirements for sensitive medical data by limiting the risk of individual data reconstruction. Higher epsilon values (e.g., 1.0–2.0) relax privacy to improve model accuracy but still maintain formal DP guarantees. This allows tuning the trade-off between privacy and performance depending on regulatory and application-specific requirements, providing practical guidance for secure federated learning deployments in healthcare.

Table 9 compares performance metrics across heart, diabetes, and CKD datasets. All three datasets exhibit good results, with precision, accuracy, recall, specificity, and F1-scores above 0.99. The diabetes dataset shows the highest overall performance, indicating the model's strong generalization and reliability across diverse medical prediction tasks.

Table 9

Performance Metrics across Datasets

Dataset	Accuracy	Precision	Recall	Specificity	F1-Score
Heart	0.992	0.994	0.991	0.993	0.993
Diabetes	0.995	0.996	0.993	0.994	0.995
CKD	0.993	0.992	0.994	0.995	0.993

The heart and CKD datasets have limited sample sizes, increasing the risk of overfitting. To address this, 10-fold cross-validation was applied during training, and dropout regularization with early stopping was used to prevent memorization of patterns. The federated setup further contributed to generalization by ensuring that learning occurred across decentralized, diverse client partitions rather than a single pooled dataset. Consistent accuracy across folds demonstrated stable performance without sharp variance. Despite this, it is acknowledged that small benchmark datasets may inflate results, and validation on larger real-world clinical datasets is required as future work to confirm the robustness of the proposed framework.

Table 10 evaluates the BCFL-FGMEAN-GWCA framework across heart, diabetes, and CKD datasets. It demonstrates high accuracy (96–98%) and strong precision, recall, and F1-scores. Specificity is slightly higher for diabetes. Blockchain integration reduces execution time by 8 seconds across all datasets, enhancing both prediction performance and system efficiency.

Table 10

Performance of BCFL-FGMEAN-GWCA Framework

Metric	Heart Disease Dataset	Diabetes Dataset	CKD Dataset
Accuracy	0.97	0.96	0.98
Precision	0.95	0.94	0.97
Recall (Sensitivity)	0.96	0.93	0.98
Specificity	0.94	0.97	0.95
F1-Score	0.96	0.94	0.97
Execution Time (w/ Blockchain)	60 sec	60 sec	60 sec
Execution Time (w/o Blockchain)	68 sec	68 sec	68 sec

Quantitative Impact of Blockchain Integration

The suggested BCFL-FGMEAN-GWCA framework utilized blockchain technology to enhance the security, integrity, traceability, and trustworthiness of collaborative federated learning processes. It is worth noting that the main purpose of the blockchain integration is not to enhance the accuracy of disease prediction directly, but to build a safe infrastructure for model exchange and decentralized healthcare intelligence. From Table 10, the suggested framework showed consistent predictive accuracy for all illness datasets considered with the advantage of blockchain-assisted model management. The total execution time was decreased from 68 seconds to 60 seconds by adding blockchain, which is a good reduction of about 11.76%. This reduction is due to the efficient coordination of model-sharing processes, decentralized verification methods and safe transaction management offered by the blockchain layer.

Beyond operational efficiency, blockchain enables immutable logging of model updates, secure access control, and transparent auditing of collaborative learning activities. Such qualities are vital in healthcare environments, where trust, responsibility and regulatory compliance are key requirements. Different from traditional federated learning systems, where privacy protection is the key concern, the proposed framework broadens the security scope by providing verifiability, tamper-proof and traceability of the model updates during the whole training process.

Therefore, the contribution of blockchain in the suggested framework should be evaluated from a system perspective. The HSOA-based feature selection, GATr feature representation, FGMEAN graph-aware learning, and GWCA optimization are the main factors affecting the predictive performance, and the blockchain offers the required infrastructure for the secure and trustworthy deployment of the federated healthcare intelligence systems.

Validation of Federated Learning Smart Decision-Making Module on Heterogeneous Clients

To evaluate the FLSDM module, we simulated heterogeneous client devices with different resource profiles. High-resource nodes (simulated with GPU-enabled environments) and low-resource nodes (limited CPU and memory) were included. The FLSDM dynamically adjusted training frequency and batch sizes based on local capacity. Results showed that low-resource nodes reduced computation load by ~30% without significant accuracy loss (<1%), while overall system latency decreased by 12%. This confirms that FLSDM enables efficient participation of resource-constrained clients without compromising predictive performance.

Performance Comparison Analysis

To highlight the effectiveness of the proposed BCFL-FGMEAN-GWCA framework, its effectiveness is compared against several previous methods across three benchmark datasets. For heart disease prediction, comparisons are made with MPNN (Janani et al., 2024), MABC-SVM (Yaqoob et al., 2023), FL+DQRE-SCnet (Om Kumar et al., 2023) and HybBPF-ELM (Pichandi et al., 2024). For diabetes prediction, the model is evaluated against MLPNN (Raja, 2024), HBESDM-DLD (Sammeta & Parthiban, 2021), REMED-T2D (Phan et al., 2025), and HybBPF-ELM (Pichandi et al., 2024). For CKD classification, it is compared with NWCNN (Saroja & Kalpana, 2023), DSCNN (Venkatrao & Kareemulla, 2023), NFIS (Jha et al., 2022), and HybBPF-ELM (Pichandi et al., 2024).

Table 11 shows that the proposed model outperforms all existing methods across the Heart Disease, Diabetes, and CKD datasets in terms of accuracy, precision, recall, specificity, and F1-score. It achieves 99.3% (Heart), 99.5% (Diabetes), and 99.3% (CKD), demonstrating strong precision and generalization capabilities.

Table 11

Performance Comparison of Developed Method with Existing Techniques

Dataset	Method	Accuracy	Precision	Recall	Specificity	F1-Score
Heart Disease Dataset	MPNN	0.943	0.940	0.936	0.950	0.938
	MABC-SVM	0.948	0.942	0.941	0.947	0.941
	FL+DQRE-SCnet	0.951	0.950	0.945	0.953	0.947
	HybBPF-ELM	0.962	0.963	0.960	0.964	0.961
	Proposed	0.993	0.994	0.991	0.993	0.993
Diabetes Dataset (Pima Indians)	MLPNN	0.948	0.946	0.944	0.949	0.945
	HBESDM-DLD	0.954	0.952	0.950	0.953	0.951
	REMED-T2D	0.963	0.962	0.961	0.960	0.961
	HybBPF-ELM	0.970	0.968	0.967	0.969	0.968
	Proposed	0.995	0.996	0.993	0.994	0.995
CKD Dataset	NWCNN	0.944	0.940	0.943	0.945	0.941
	DSCNN	0.950	0.947	0.949	0.952	0.948
	NFIS	0.961	0.962	0.960	0.960	0.961
	HybBPF-ELM	0.968	0.970	0.968	0.969	0.969
	Proposed	0.993	0.992	0.994	0.995	0.993

Statistical Analysis

Two statistical tests of the effectiveness of the recommended BCFL-FGMEAN-GWCA approach are evaluated using the Wilcoxon Signed-Rank (WSR) and Shapiro-Wilk (SW) tests. The WSR test, a non-parametric statistical test, was employed to compare the median ranks of two related samples or repeated measurements when the assumption of normality was not satisfied (Wilcoxon, 1945). On the other hand, the SW test was conducted to assess the normality of the dataset. This test evaluates the null hypothesis that the observed data are drawn from a normally distributed population (Shapiro & Wilk, 1965).

The resulting p-values in Table 12 indicate that the differences between the proposed framework and the compared baseline techniques are statistically significant. Both tests resulted in low p-values, confirming the robustness and reliability of the recorded performance gains.

It is worth noting that the current work mainly focuses on comparative performance evaluation with the benchmark healthcare datasets and federated learning simulations. The statistical tests performed suggest that the suggested framework is robust and effective, and additional insights into the variability of performance can be gained from repeated experimental trials and confidence interval analysis. Future work will include extending the experimental evaluation by adding several independent runs, confidence interval estimation and more robustness analyses over larger and more heterogeneous healthcare datasets.

Table 12

Statistical Analysis between Baseline Techniques

Reference	Techniques	SW test p-Value	WSR test p-Value
Janani et al. (2024)	MPNN	0.0044	0.0052
Yaqoob et al. (2023)	MABC-SVM	0.0057	0.0063
Om Kumar et al. (2023)	FL+DQRE-SCnet	0.00674	0.00567
Pichandi et al. (2024)	HybBPF-ELM	0.00809	0.0076
Raja (2024)	MLPNN	0.0044	0.0054
Sammeta & Parthiban (2021)	HBESDM-DLD	0.0046	0.0067
Phan et al. (2025)	REMED-T2D	0.0036	0.0051
Saroja & Kalpana (2023)	NWCNN	0.0056	0.0078
Venkatrao & Kareemulla (2023)	DSCNN	0.00398	0.0045
Jha et al. (2022)	NFIS	0.0054	0.0073
	Proposed	0.0011	0.0012

Ablation Study

An ablation study was performed by removing key components of the BCFL-FGMEAN-GWCA framework. Results show that each module blockchain, GWCA, HSOA, and FGMEAN contributes to performance, with FGMEAN and GWCA having the most significant impact on accuracy, underscoring their importance in edge-aware learning and optimization.

Table 13 of the ablation study shows the impact of removing components from the BCFL-FGMEAN-GWCA model. The full model achieves the highest accuracy (99.3%). Removing FGMEAN or GWCA causes noticeable drops, while excluding Blockchain does not affect the performance. This highlights the contribution of each component to the method's effectiveness and precision.

Table 13

Ablation Study Results on Combined Multi-Disease Dataset

Configuration	Accuracy	Precision	F1-Score
Full Model	0.993	0.994	0.993
Without FGMEAN	0.970	0.968	0.968
Without GWCA	0.973	0.970	0.970
Without HSOA	0.976	0.974	0.973
Without Blockchain	0.993	0.994	0.993

Component-Wise Contribution Analysis

Table 14 shows the component-wise contribution analysis of the proposed BCFL-FGMEAN-GWCA framework. The findings show the incremental performance gain obtained by the incorporation of each module. The baseline FL-FGMEAN reached 97.1% accuracy, proving that graph-aware federated learning alone provides powerful predictive capacity for multi-disease categorization. The accuracy was improved to 98.0% by the use of HSOA, which showed the usefulness of optimal feature selection in removing duplicated and irrelevant clinical attributes. The addition of GATr resulted in a higher accuracy of 98.7% by creating more potent representations of geometrical aspects and maintaining sophisticated connections among medical variables. The accuracy improved to 99.3% by applying the GWCA-based hyperparameter optimization, showing the necessity of adaptive parameter tuning to boost the convergence and the model generalization. Finally, the entire BCFL-FGMEAN-GWCA structure attained the same accuracy of 99.3%. This is because the blockchain does not immediately assist feature learning or classification. Its integration contributes to the trustworthiness, integrity and security of federated model exchange. The results indicate that each component contributes favourably to the overall framework, justifying the design of the suggested multi-component architecture.

Table 14

Component-Wise Contribution Results

Configuration	Accuracy (%)
FL + FGMEAN	97.1
FL + HSOA + FGMEAN	98.0
FL + HSOA + GATr + FGMEAN	98.7
FL + HSOA + GATr + FGMEAN + GWCA	99.3
Full Framework (Blockchain included)	99.3

Great Wall Construction Algorithm Validation

Table 15 compares GWCA with traditional hyperparameter optimization methods, including human tuning, grid search and random search. The results reveal that manual tuning generated the lowest predictive performance because of the use of expert-selected parameter settings and limited exploratory capability. Random search enhanced accuracy by examining a broader parameter space but the stochastic nature of this method often yielded inconsistent solutions. The grid search outperformed the random search due to the systematic search of the search space but with a far larger computational cost. On the other hand, GWCA achieved not only the maximum accuracy but also a fair training duration. The higher performance is due to its balanced exploration and exploitation algorithms, which help it efficiently find near-optimal hyperparameter combinations. Results show that GWCA is a more effective and computationally efficient optimization approach for the proposed federated multi-disease prediction framework than the traditional tuning approaches.

Table 15

GWCA Validation

Configuration	Accuracy (%)	Training time
Manual Tuning	97.8	High
Random Search	98.4	Moderate
Grid Search	98.7	High
GWCA	99.3	Moderate

Scalability and Robustness Analysis

To assess scalability, experiments were extended to include varying client counts (5, 10, 20, 50). Accuracy remained above 97% up to 20 clients and decreased slightly (≈2%) at 50 clients due to increased communication noise. To evaluate robustness, non-IID data partitions were simulated by distributing class-imbalanced subsets across clients. The model maintained accuracy above 96% and demonstrated stable convergence despite heterogeneity. The bandwidth impact was tested by reducing the available communication bandwidth by 40%. Owing to IPFS-based off-chain storage and compressed model updates, latency increased by less than 12%, confirming communication efficiency. These findings confirm that the framework remains effective across varying client counts, heterogeneous data distributions, and constrained communication environments.

Data in Table 16 demonstrates that the framework maintains high accuracy across different numbers of clients, data heterogeneity, and constrained bandwidth scenarios. Latency and execution time increase slightly with client count and bandwidth limitations, but overall system performance remains robust, confirming effective scalability and efficiency in federated multi-disease prediction.

Table 16

Scalability and Bandwidth Analysis of BCFL-FGMEAN-GWCA Framework

Number of clients	Data distribution	Accuracy	Average latency (s)	Execution time (s)
5	IID	0.993	0.95	60
10	IID	0.991	0.98	62
20	Non-IID	0.970	1.05	65
50	Non-IID	0.951	1.12	68
20	Non-IID, Low Bandwidth (-40%)	0.967	1.18	66

DISCUSSIONS

The BCFL-FGMEAN-GWCA model offers a secure, privacy-preserving solution for multi-disease prediction by integrating blockchain and federated graph learning. It ensures data protection, preserves spatial relationships via GATr, enhances accuracy via FGMEAN, and optimizes performance with HSOA and GWCA, achieving up to 99.5% accuracy. The proposed framework not only achieves high predictive accuracy but also provides clinically meaningful insights. Feature correlations and edge-aware attention maps enable the identification of key clinical markers that influence predictions, supporting interpretable decision-making for doctors. Early diagnosis of diseases such as CKD is facilitated by precise detection of critical indicators, which helps reduce false negatives and improve patient outcomes. By providing reliable, explainable predictions, the framework assists clinicians in prioritizing high-risk patients, optimizing intervention timing, and enhancing overall healthcare decision-making.

Nevertheless, this advanced configuration also brings some restrictions. Its computational cost and edge attention mechanisms could be a hindrance to scalability in a real-world setting of resource-limited healthcare facilities. Also, storing and executing the blockchain infrastructure is technically demanding and involves storage and latency costs. Nevertheless, the proposed framework for secure, distributed medical diagnosis offers significant enhancements in accuracy, privacy, and reliability, making it a notable strength in AI-based healthcare.

CONCLUSION

The proposed BCFL-FGMEAN-GWCA framework offers a novel and robust solution for multi-disease prediction by integrating blockchain technology, federated learning, advanced graph neural networks, and multi-level optimization strategies. By ensuring data privacy through local model training and secure blockchain-based update mechanisms, and by enhancing model accuracy through efficient feature selection (HSOA) and extraction (GATr), the framework achieves high performance across heart, diabetes, and CKD datasets with accuracies reaching up to 99.5%. The FLSDM module and FGMEAN architecture further boost the system’s adaptability and intelligence in handling distributed clinical data. However, the approach faces limitations in terms of computational complexity and infrastructure requirements, particularly for resource-constrained healthcare settings. Future work may focus on optimizing the system’s efficiency for real-time deployment, reducing communication overhead in FL rounds, and extending the model to include additional diseases and real-world clinical datasets, thereby enhancing its applicability in broader healthcare domains.

ACKNOWLEDGMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Anaraki, M. V., & Farzin, S. (2023). Humboldt Squid Optimization Algorithm (HSOA): A novel Nature-Inspired technique for solving optimization problems. *IEEE Access, 11*, 122069–122115. https://doi.org/10.1109/access.2023.3328248

Brehmer, J., Pim, D. H., Behrends, S., & Cohen, T. (2023). *Geometric algebra transformer*. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2305.18415

Dubey, P., Dubey, P., & Bokoro, P. N. (2025). Federated learning for privacy-enhanced mental health prediction with multimodal data integration. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 13*(1), 2509672. https://doi.org/10.1080/21681163.2025.2509672

Gu, C., & Lee, M. (2024). Deep transfer learning using real-world image features for medical image classification, with a case study on pneumonia x-ray images. *Bioengineering*, *11*(4), 406. https://doi.org/10.3390/bioengineering11040406

Hu, H., Wang, Y., & Song, G. (2024). Research on convergence media ecological model based on blockchain. *Systems*, *12*(9), 381. https://doi.org/10.3390/systems12090381

Huang, C., Xu, G., Chen, S., Zhou, W., Ng, E. Y., & De Albuquerque, V. H. C. (2022). An improved federated learning approach enhanced Internet of Health Things framework for private decentralised distributed data. *Information Sciences, 614,* 138–152. https://doi.org/10.1016/j.ins.2022.10.011

Hukkeri, G. S., Goudar, R. H., Dhananjaya, G. M., Rathod, V. N., & Ankalaki, S. (2025). A comprehensive survey on split-fed learning: Methods, innovations, and future directions. *IEEE Access*, *13*, 46312–46333. https://doi.org/10.1109/access.2025.3547641

Imoize, A. L., Obaidat, M. S., & Song, H. H. (2024). Legal implications of federated learning integration in digital healthcare systems. In *Elsevier eBooks* (pp. 355–385). https://doi.org/10.1016/b978-0-443-13897-3.00014-x

Janani, I., Deepa, G. G., Dhiyanesh, B., Lakshmi, C. B. S., & Yawanikha, T. (2024). Federated learning for heart disease detection using a grey wolf transfer learning model based on multi-perceptron neural networks. In *Healthcare-Driven Intelligent Computing Paradigms to Secure Futuristic Smart Cities* (pp. 304–327). Chapman and Hall/CRC. https://doi.org/10.1201/9781032631738-18

Jha, R. K., Henge, S. K., Mandal, S. K., Sharma, A., Sharma, S., Sharma, A., & Berhanu, A. A. (2022). Neural fuzzy hybrid rule-based inference system with test cases for prediction of heart attack probability. *Mathematical Problems in Engineering*, *2022*, 1–18. https://doi.org/10.1155/2022/3414877

Ji, S., Tan, Y., Saravirta, T., Yang, Z., Liu, Y., Vasankari, L., Pan, S., Long, G., & Walid, A. (2024). Emerging trends in federated learning: From model fusion to federated X learning. *International Journal of Machine Learning and Cybernetics*, *15*(9), 3769–3790. https://doi.org/10.1007/s13042-024-02119-1

Khan, M. F., & Abaoud, M. (2023). Blockchain-Integrated security for real-time patient monitoring in the internet of medical things using federated learning. *IEEE Access*, *11*, 117826–117850. https://doi.org/10.1109/access.2023.3326155

Liu, H., Feng, Y., Xu, H., Liang, S., Liang, H., Li, S., Zhu, J., Yang, S., & Li, F. (2022). MEA-Net: Multilayer edge attention network for medical image segmentation. *Scientific Reports, 12*(1), 7868. https://doi.org/10.1038/s41598-022-11852-y

Liu, R., Xing, P., Deng, Z., Li, A., Guan, C., & Yu, H. (2024). Federated graph neural networks: Overview, techniques, and challenges. *IEEE Transactions on Neural Networks and Learning Systems*, *36*(3), 4279–4295. https://doi.org/10.1109/tnnls.2024.3360429

Nasser, N., Fadlullah, Z. M., Fouda, M. M., Ali, A., & Imran, M. (2021). A lightweight federated learning based privacy preserving B5G pandemic response network using unmanned aerial vehicles: A proof-of-concept. *Computer Networks*, *205*, 108672. https://doi.org/10.1016/j.comnet.2021.108672

Om Kumar, C. U., Sudhakaran, G., Bhavadharini, R. M., Suguna, M., & Krithiga, R. (2023). EHR privacy preservation using federated learning with DQRE-Scnet for healthcare application domains. *Knowledge-Based Systems*, *275*, 110638. https://doi.org/10.1016/j.knosys.2023.110638

Phan, L. T., Rakkiyappan, R., & Manavalan, B. (2025). REMED-T2D: A robust ensemble learning model for early detection of type 2 diabetes using healthcare dataset. *Computers in Biology and Medicine*, *187*, 109771. https://doi.org/10.1016/j.compbiomed.2025.109771

Pichandi, K. V., Janarthanan, V., Annamalai, T., & Arumugam, M. (2024). Enhancing healthcare in the digital era: A secure e-health system for heart disease prediction and cloud security. *Expert Systems with Applications*, *255*, 124479. https://doi.org/10.1016/j.eswa.2024.124479

Raj, S., Namdeo, V., Singh, P., & Srivastava, A. (2025). Identification and prioritization of disease candidate genes using biomedical named entity recognition and random forest classification. *Computers in Biology and Medicine*, *192*(Pt B), 110320. https://doi.org/10.1016/j.compbiomed.2025.110320

Raja, M. W. (2024). Artificial Intelligence-Based healthcare data analysis using Multi-Perceptron Neural Network (MPNN) based on optimal feature selection. *SN Computer Science*, *5*(8). https://doi.org/10.1007/s42979-024-03323-0

Sammeta, N., & Parthiban, L. (2021). Hyperledger blockchain enabled secure medical record management with deep learning-based diagnosis model. *Complex & Intelligent Systems*, *8*(1), 625–640. https://doi.org/10.1007/s40747-021-00549-w

Saroja, T., & Kalpana, Y. (2023). Hybrid missing data imputation and novel weight convolution neural network classifier for chronic kidney disease diagnosis. *Measurement Sensors*, *27*, 100715. https://doi.org/10.1016/j.measen.2023.100715

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3–4), 591–611. https://doi.org/10.1093/biomet/52.3-4.591

Sharma, P., Sharma, P., Sharma, K., Varma, V., Patel, V., Sarvaiya, J., Tavethia, J., Mehta, S., Bhadania, A., Patel, I., & Shah, K. (2025). Revolutionizing utility of big data analytics in personalized cardiovascular healthcare. *Bioengineering*, *12*(5), 463. https://doi.org/10.3390/bioengineering12050463

Singh, B., & Kaunert, C. (2024). Reinventing Artificial Intelligence and blockchain for preserving medical data. In *Ethical Artificial Intelligence in Power Electronics* (pp. 77–91). https://doi.org/10.1201/9781032648323-6

Sutranggono, A. N., Riyanarto Sarno, & Imam Ghozali. (2024). Multi-Class multi-level classification of mental health disorders based on textual data from social media. *Journal of Information and Communication Technology*, *23*(1), 77-104. https://doi.org/10.32890/jict2024.23.1.4

Theodorakis, N., & Nikolaou, M. (2025). From Cardiovascular-Kidney-Metabolic Syndrome to Cardiovascular-Renal-Hepatic-Metabolic Syndrome: Proposing an expanded framework. *Biomolecules*, *15*(2), 213. https://doi.org/10.3390/biom15020213

Tran, A. T., Zeevi, T., & Payabvash, S. (2025). Strategies to improve the robustness and generalizability of deep learning segmentation and classification in neuroimaging. *BioMedInformatics, 5*(2), 20. https://doi.org/10.3390/biomedinformatics5020020

Venkatrao, K., & Kareemulla, S. (2023). HDLNET: A hybrid deep learning network model with intelligent IOT for detection and classification of chronic kidney disease. *IEEE Access*, *11*, 99638–99652. https://doi.org/10.1109/access.2023.3312183

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*(6), 80. https://doi.org/10.2307/3001968

Yang, Z., Lin, Y., Xu, Y., Hu, J., & Dong, S. (2023). Interpretable disease prediction via path reasoning over medical knowledge graphs and admission history. *Knowledge-Based Systems*, *281*, 111082. https://doi.org/10.1016/j.knosys.2023.111082

Yaqoob, M. M., Nazir, M., Khan, M. A., Qureshi, S., & Al-Rasheed, A. (2023). Hybrid classifier-based federated learning in health service providers for cardiovascular disease prediction. *Applied Sciences*, *13*(3), 1911. https://doi.org/10.3390/app13031911

Yarman, B. S., & Rathore, S. P. S. (2025). The future of AI in disease detection — A look at emerging trends and future directions in the use of AI for disease detection and diagnosis. *AI In Disease Detection*, 265–288. https://doi.org/10.1002/9781394278695.ch12

Zitouni, F., Almazyad, A. S., Xiong, G., Mohamed, A. W., & Harous, S. (2024). An opposition-based great wall construction metaheuristic algorithm with Gaussian mutation for feature selection. *IEEE Access*, *12*, 30796–30823. https://doi.org/10.1109/access.2024.3367440