International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 24s, 2025
https://theaspd.com/index.php

Machine Learning-Driven Clinical Decision Support Systems For Healthcare Optimization

[1]Ismail Almsallati, [2]Hari Mohan Singh
[1]Dept. Of Computer Science &IT, SHUATS, Sam Higginbottom university of agriculture technology and sciences, Prayagraj, aalmslate@gmail.com
[2]Assistant Professor, Dept. Of Computer Science &IT, SHUATS. hari.singh@shiats.edu.in

Abstract
In the era of digital healthcare, the exponential growth of sensitive medical data presents a dual challenge: leveraging data for AI-driven insights while preserving patient privacy. Traditional centralized AI models pose significant risks related to data breaches, regulatory compliance, and trust among data providers. To overcome these issues, this paper suggests a federated learning (FL) model that incorporates blockchain technology to facilitate the implementation of secure and decentralized training of models in concert among various healthcare facilities. This study aims at creating a system whereby local models are trained on-site using clinical, research and behavioral data, and at most only model parameter is shared and not the raw data. A consensus mechanism to guarantee integrity and traceability of model updates is a blockchain-based method, in which validated updates are permanently recorded with the help of smart contracts. The methodology has been organized into three levels, namely, data extraction and storage, data management, and data application. The findings show that the proposed FL-CNN-HMChain model outperforms traditional CNN and VGG16 architectures, achieving AUC scores above 85% across all tasks, with a maximum of 98.9% on Organ AMNIST. This approach demonstrates both high accuracy and strong privacy preservation, paving the way for scalable and ethical AI deployment in healthcare.
Keywords: *Federated Learning, Blockchain, Privacy-Preserving AI, Healthcare Data Security, Smart Contracts*

1. INTRODUCTION

The proliferation of health-related data sources has markedly escalated over the last decade, driven by advancements in the size, performance, and usability of hardware technologies and software applications [1]. A substantial volume of data is produced and retained inside the infrastructure of diverse organizations across many sectors, including health insurance, pharmacy, patient management, and healthcare. The present circumstances enable data science to explore and profoundly influence decision-making processes, providing essential support in healthcare quality. Machine learning applied to health-related data and data-driven insights encounters several hurdles due to the inherent nature of the data created. They are characterized by high dimensionality, heterogeneity, time dependence, sparsity, and irregularity [3]. The distinct attributes of health-related data, including complexity, temporal volatility, heterogeneity, interoperability, sparsity, and isolation, provide problems in the establishment of effective data analysis and data science methodologies [4]. Conversely, advancements in hardware and software applications are augmenting data volume, therefore capturing researchers' attention and broadening the potential for effect. Over the last decade, several approaches and tools for predictive analysis have been widely accessible, enhancing data insights and facilitating the decision-making process [5]. The difficulty of preserving privacy in the management of health-related data is a significant ethical issue that requires meticulous deliberation. Safeguarding patient confidentiality is essential, since any violation of privacy may result in significant repercussions for the affected persons. This difficulty stems from the need to handle substantial quantities of data produced by the healthcare sector, including electronic health records, patient histories, and diagnostic findings. The vast volume of data complicates the safeguarding of each individual piece of information, necessitating a strong architecture and severe rules to assure continuous privacy protection [6]. Traditional machine learning involves a centralized server that aggregates data sets from several sources, does data pre-processing on the aggregated data, and utilizes this for training certain models. A significant obstacle for institutions in adapting to this architecture is the legal norms and legislation governing the handling of private and sensitive data [7]. These constraints hinder the prospective advancement of machine learning initiatives inside organizations that possess this data type. The data necessary for training machine learning models is often inadequate

inside a particular institution. Data sets are often constrained and influenced by external environmental factors, introducing bias that leads to models that lack generalization and exhibit subpar performance when applied to atypical feature values. Federated learning presents novel opportunities and pathways by addressing two primary concerns: privacy and sparsity [8]. The notion denotes a collaborative methodology whereby several entities use their data and computational resources to iteratively train a machine learning model, sharing just the updates and the refined model with other participants.
Healthcare privacy preservation by FL is defined in figure 1. Focusing on healthcare factors using technology-based clarifications in offering illness forecast and enhancing patient satisfaction. This enabled the growth in medicine; however, it has its own cons including privacy at risk [10]. FL also solves several of the privacy concerns associated with centralized healthcare systems such as breaches, unauthorized access to information, and lack of control over patient data. The centralized systems are risky because they are putting essential data at a single location. The adoption of new laws such as as GDPR and HIPAA has necessitated the use of decentralized systems such as FL to ensure safe handling of data. The care provision evidence has been increasing in healthcare as a result of the modernization of electronic medical databases and widespread use of wearable expertise [11]. This surge in the availability of data has increased concerns in the areas of openings, unlawful infiltration, and covert theft of involved data [12]. In order to establish these anxieties, FL is practiced in an attempt to preserve confidentiality, altering out-dated methods of investigating data [13].

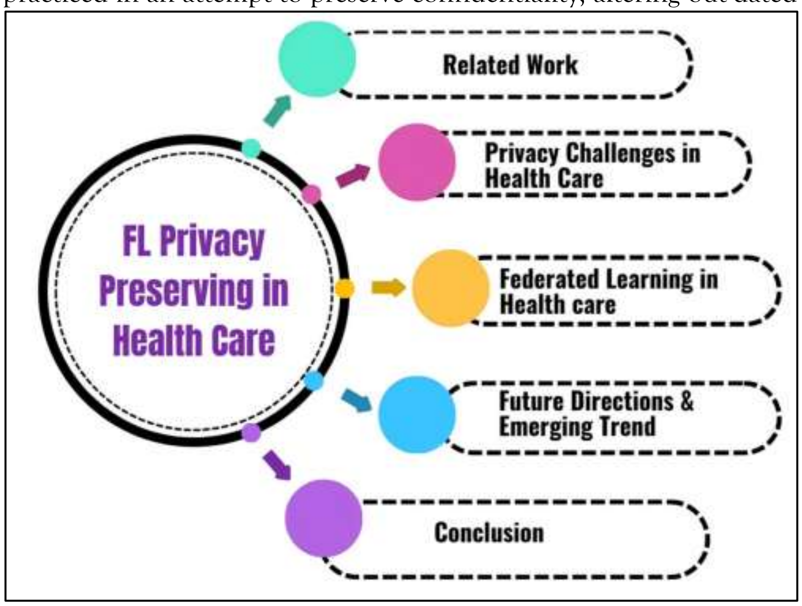

Figure 1. Healthcare privacy preservation by FL [14]
The present study adds a new privacy-preserving model that integrates federated learning models with blockchain technology to train AI models in a decentralized way without involving the exchange of sensitive patient information in multiple healthcare facilities. It provides a stratified approach with data extraction, management, and application, which is scalable, modular, and trusted. By integrating blockchain-based consensus and smart contracts, the system ensures tamper-proof validation and transparent recording of model updates.

2. REVIEW OF LITERATURE

Chen et al., (2025)[15] examined comparative analysis of innovative agent system design using modular agents to estimate laboratory data, vital signs, and clinical context in order to make predictions and verify them. They combine the agent system with eICU database and begin with the implementation of lab analysis, vitals-only interpreter agent, and the contextual reasoner agent, and with the dispersion of the memory to the integration agent, prediction agent, transparency agent, and validation agent. The conclusions are that the MAS was more efficient in predicting mortality (59% vs. 56%), mean error of length of stay (LOS).
Singh et al., (2024)[16] suggested the Gravitational Search Optimization Algorithm (GSOA) approach for meta heuristic-based feature selection. Glaucoma has been chosen as the focus of study because to its rapid

global proliferation; projections indicate 111 million cases by 2040, an increase from 64 million in 2015. It results in extensive visual impairment. Optic nerve fibers may undergo degeneration and cannot be subsequently regenerated in this condition. Initially, retinal fundus photographs of individuals with glaucoma and healthy individuals were used, from which 36 characteristics were extracted from both public benchmark datasets and a proprietary dataset. Six machine learning models are trained for categorization based on the subset of characteristics supplied by the GSOA. The suggested feature selection approach improves classification performance by identifying the most significant characteristics. The eight metrics for measuring statistical performance, together with execution time, are computed. The training and testing were conducted utilizing a split methodology (70:30), with 5-fold and 10-fold cross-validation (CV). The suggested method attained an accuracy of 95.36%. Owing to its favorable efficacy, physicians may use the suggested approach to get a second opinion, so alleviating the pressure on competent medical professionals and preventing patients from experiencing eyesight loss.

Bandi et al., (2024)[17] examined an osteoporosis is a skeletal disorder that diminishes bone density and elevates the likelihood of fractures. The identification of osteoporosis has been done through the application of machine learning. However, prediction of osteoporosis illness has not been better and the time taken has not been reduced. The quantitative data analysis indicates that the accuracy of illness prediction, precision, recall and F1 score of RCCEDCGAN method are 5% and specificity increase by 9% as compared to the RR model and modified GP classifier. The p-value of RCCEDCGAN was below 0.05 and the confidence interval of 95 percent of obtaining osteoporosis. The prediction time was shortened by 8 percent of the RR model and new GP classifier algorithms through RCCEDCGAN method. The RCCEDCGAN, therefore, is an effective tool in early diagnosis and risk reduction of osteoporosis in order to prevent and manage it.

Takale et al., (2023)[18] introduced a critical care unit (ICU) innovative machine learning-based clinical decision support system. The system will focus on forecasting the mean arterial pressure (MAP) values on-the-fly and providing decision-making assistance to medical personnel. The suggested LSTM+HTM model is potentially useful to critical care nurses and physicians since it is most accurate, precise, and recalls, F1 score, and AUC-ROC when predicting MAP levels. The solution offered will assist in early detecting the change in MAPs values and offer the health care professional with timely decision making, thereby enhancing patient outcome within the critical care units. The system's effectiveness requires evaluation on a larger dataset, and the feasibility of its integration into existing clinical procedures must be examined more thoroughly.

Mustapha et al., (2022)[19] seen to use an innovative technique that integrates artificial intelligence with a multi-criteria decision-making process for a more comprehensive assessment of machine learning models. The last is the least popular, that is the naive Bayes classifier with a net flow of -0.0766. The findings of this research demonstrate the efficacy of the suggested strategy in facilitating optimal decision-making for the selection of the best suitable machine learning model.

Knapic et al., (2021)[20] discussed the potential of Explainable Artificial Intelligence techniques for decision assistance in medical image analysis contexts. Three explainable methods were applied to the same set of medical pictures to make judgments provided by the CNN more interpretable. The explanations that resulted were humanized. They conducted three user studies on the basis of explanations generated by LIME, SHAP and CIU. A series of tests were conducted that involved individuals who represented different non-medical domains in a web-survey context and explained how they experienced the provided explanations and what their perception about the same was. Three types of explanations were performed quantitative analysis in three groups of users (n = 20, 20, 20). The findings have confirmed the assumption that CIU-explainable approach was more efficient than the LIME and SHAP to support the human decision-making process, as well as be more transparent and comprehensible to the users.

Mansour et al., (2021)[21] introduced a novel illness detection model for smart healthcare systems, based on the convergence of AI and IoT technologies. The main goal of the paper is to create a disease diagnostic model of the heart disease and diabetes with the help of AI and IoT convergence. The suggested model consists of a great number of steps data gathering, pre-preparation, categorization, and adjustment of parameters. The wearables and sensors are IoT applications that provide constant data collection and the AI methodologies utilize the statistics to identify diseases. The suggested process relies on a CSO-CLSTM process of disease diagnosis based on a Crow Search Optimization algorithm. In addition, the Isolation Forest (iForest) method also eliminates outliers in this study. The CSO-LSTM model was estimated in terms of the effectiveness of healthcare data. In both experiments, CSO-LSTM has achieved the highest accuracy of 96.16 and 97.26 in the diagnosis of heart diseases and diabetes respectively. Consequently, the suggested CSO-LSTM model may serve as an effective diagnostic instrument for intelligent healthcare systems.

Fitriyani et al., (2020)[22] suggested an effective heart disease prediction model (HDPM) for a CDSS which consists of Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to detect and eliminate the outliers, a hybrid Synthetic Minority Over-sampling Technique-Edited Nearest Neighbor (SMOTE-ENN) to balance the training data distribution and XGBoost to predict heart disease. Two publicly available datasets (Statlog and Cleveland) were used to build the model and compare the results with those of other models (naive bayes (NB), logistic regression (LR), multilayer perceptron (MLP), support vector machine (SVM), decision tree (DT), and random forest (RF)) and of previous study results. The findings suggested that the proposed model had a better performance in comparison with other models and the findings of previous researches in terms of capacity to run with high accuracies of Statlog and Cleveland datasets of 95.90 and 98.40 respectively.

Yahyaoui et al., (2019)[23] suggested a Decision Support System (DSS) for diabetes prediction using Machine Learning (ML) approaches. They compared the different machine learning strategy with that of the deep learning strategies. Using the traditional machine learning framework, we have considered the most common types of classifiers SVM and RF. On the other hand, they employed an entirely CNN in predicting and recognizing diabetes patients in DL. The database in the approach provided is the Pima Indians Diabetes database, which is publicly accessible and is present as a total of 768 samples with 8 characteristics per sample. There were five hundred non-diabetic and two hundred and sixty-eight the diabetes individuals. The overall accuracy of the DL, SVM and RF was 76.81, 65.38 and 83.67 respectively. The outcomes of the experiments display that RF outperformed deep learning and SVM methods in the prediction of diabetes.

Yang et al., (2019)[24] delineated the design and field assessment of an innovative kind of DST. It autonomously produces presentations for physicians' decision-making sessions, using discreetly integrated machine prognostics. This design draws inspiration from the concept of Unremarkable Computing, suggesting that by enhancing users' routines, technology and AI may play a substantial role in users' lives while being inconspicuous. The field assessment indicates that doctors are more inclined to encounter and use such a decision support tool. Based on their findings, they examined the significance and complexities of achieving the appropriate degree of inconspicuousness in DST design, and convey insights gained from prototyping essential AI systems as a contextual experience.

3. Problem statement

Despite the previously described blockchain and federated learning-based methodologies, systems, platforms, and frameworks that facilitate data storage, transmission, processing, and access, the majority fail to achieve an equilibrium between the efficacy and security of data applications. In other studies, the effectiveness of data services was under consideration but there was no attention to data fusion, storage, and sharing security. The various security controls have been presented, but it is not clear on the effectiveness of information sharing and collaboration. To exemplify this, a system that is based on efficiency may fail to provide the security to withstand any attack thus exposing data to illegal access, manipulation, or exposure. Conversely too strict security requirements may impair communication and cooperation and thus lower the effectiveness of the system. Moreover, the services of data and data security or other of a combination of the two are performed with the help of blockchain or federated learning. Federated learning based on blockchain has been studied widely, and blockchain based federated learning is under studied which means that the gains are unevenly spread between the two. In this respect, we are attentive to the improvement of blockchain by training the federated learning model, implying a platform of offering privacy-sensitive healthcare and medical data collaboration services. It combines verified data on various sources, transactions, using the node consensus of the blockchain, allowing complex data storage and transfer to users with a variety of medical application services.

4. RESEARCH METHODOLOGY

Here, we are integrating a federated learning model with blockchain technology that will allow the training of AI models on sensitive healthcare data of various medical institutions, in a decentralized and secure way. Each of the institutions does not share the raw patient data but they each train a local model using their own data and only send the model updates (parameters) to a central server. Such updates are then verified by a blockchain-based consensus mechanism, after the successful verification, the update is stored as irreversible blocks using smart contracts to guarantee data integrity, trust and traceability. This joint model enables institutions to create robust, privacy-sensitive AI models to use in clinical research, biomedical development, neuroscience, and genome engineering without violating data security and ethical guidelines. Figure 2 depict the framework of blockchain-based consensus mechanism as shown below.

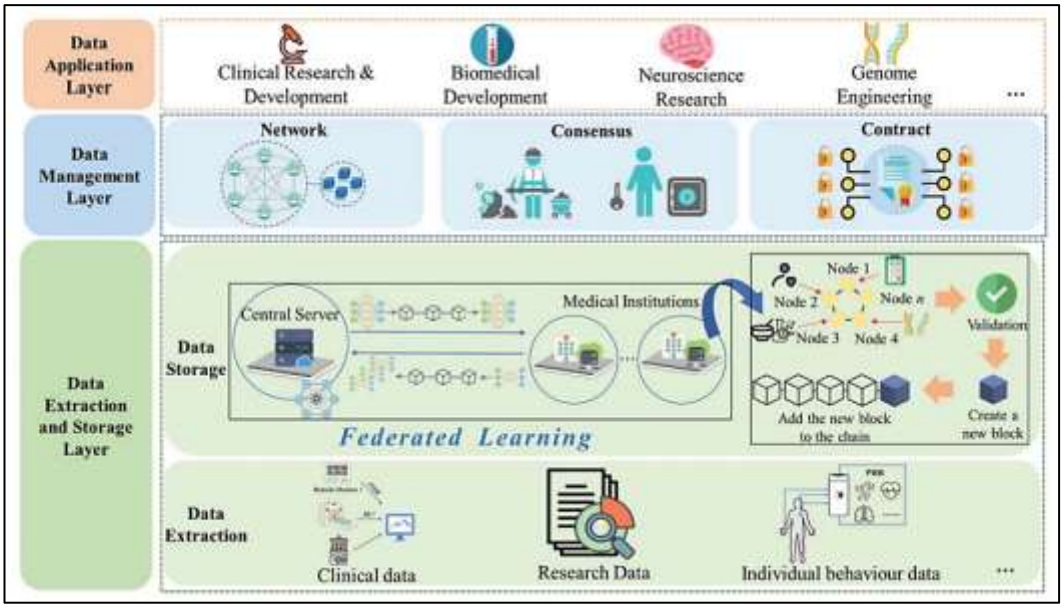

Figure 2. Framework for blockchain-based consensus mechanism

4.1 Data extraction and storage layer

The data extraction and storage layer is responsible for providing efficiency and dependability in the fusion and storage of data. The techniques for the production of blocks and of transactions have been provided in order to store data in a safe and dependable manner. "Federated learning (FL) has the capability to bring about both local and global model training, as well as to update the parameters of the learned model on the blockchain without the need to share raw data.

4.1.1 Data extraction

The characteristics of multi-source, cross-regional, multi-relations, heterogeneity, and other characteristics such as these are present in healthcare and medical data that has been acquired from various mobile devices, platforms, and systems. Specifically, this system is responsible for the extraction of three categories of data from a wide variety of healthcare and medical organizations. These categories include clinical, research, and individual behavior data:

- **Clinical Data:** This information is primarily gathered as the hospital information system (HIS), that is, a variety of sub-systems: the doctor station, the picture archiving and communication system (PACS), the laboratory information system (LIS), the microbial experiments, the pharmaceutical administration system, the pathological information system, etc. EMRs have been searched to retrieve clinical text, picture and multimedia content related to diagnosis, treatment, prescriptions, tests, exams and other related topics.
- **Research Data:** These data include clinical trial data, gene sequence data, and medical data, most of such data are gathered by the systems or platforms of devices research and development companies, outsourcing companies, genetic testing companies or by the scientific research institution.
- **Individual Behavior Data:** This information is live-time and gathered on a per-activity basis that can be seen on the mobile devices, which include smart phones, smart watches, and fitness gadgets. Through

mobile devices, the activity of an individual is monitored daily on the real time like the sleeping patterns, calories burned, body temperature and heart rate. The rest of the relevant information is found in many online healthcare and medical systems or platforms most of which are stored in the EHR.

4.1.2 Data storage

The data storage layer contains two different types of data that are stored within the blockchain. Personal behavioral data, clinical data, research data, and information received in various institutions, on mobile devices, systems, or the platform. The other data set is the parameters obtained during global and local training of the model in a federated learning framework of medical data and healthcare cooperation service. A blockchain is developed to share, become transparent, and be vulnerable to manipulation of various healthcare and medical information, and model equations. Each block contains an index, a timestamp, a data, a current hash value and a previous block hash value and each block is linked by pointers to the other blocks. These interconnected blocks constitute a blockchain, which provides the integrity of healthcare and medical data and model parameters and also their tamper-resistance. Blockchain is based on a decentralized registry. A node of a healthcare network initiates a transaction and replicates to other nodes. This network would then generate one-of-a-kind hash through a safe algorithm which is called the safe Hash Algorithm 256 (SHA-256). Lastly, each hash is connected to the one that precedes it forming an irreversible network of transactions, which can only be validated by a node or a smart contract. This unchanging registry can have more blocks of trade added to it. The establishment of a decentralized network will facilitate the verification of certain completed transactions, including contracts and encryption. "Algorithm 1 serves as block storage for extracted data and model parameters.

Algorithm 1: Storage of Blockchain

Input: Block ()
1. Index, previous_hash
2. Data = healthcare and medical data, or parameters from model training
3. Timestamp = time ()
4. Hash = SHA-256 () // hash value generation function

Input: AddBlock ()

5. Block = Block ()
6. If blok, previous_hash = previous_block.hash then
7. Chain. Append (block)
8. End if

Input: Verification ()

9. For i in range (1, len (chain)) do
10. Block = chain [i]
11. Previous_block = chain [i-1]
12. If block. Previous_hash ! = previous_block. Hash then
13. Invalid block
14. End if
15. End for

4.2 Data management layer

Peer-to-peer (P2P) networks, consensus procedures, and smart contracts are the primary technologies that are referred to by this layer.

4.2.1 Peer-to-Peer network

Blockchain nodes must agree on transactions using various consensus processes inside a peer-to-peer network. The dissemination of information in blockchain systems among interconnected peers relies on a P2P network.

4.2.2 Consensus mechanism

This enhanced consensus technique has five phases: A cohort of master nodes is designated to oversee the validation of local gradients and the production of blocks. The masters are chosen dynamically at the start of each round, contingent upon the performance scores of nodes from the preceding round. The remaining client nodes in the network do local training on their individual datasets, producing local updates and transmitting < REQUEST, o, t, c > to the master node along with signatures. The 'o' denotes the particular operation requested. The timestamp is denoted as 't', while 'c' represents the identification of the client node. The REQUEST encompasses the message content and the message summary d(m).

- Upon receiving the request, the master nodes authenticate it with their data as a validation set and allocate it a sequence number n. Each master executes the learning model using the local updates, assesses the validation accuracy on their dataset, and allocates the accuracy to each local client as the score s. The master node subsequently disseminates << PRE PREPARE, v, n, d >, m, s > to additional client nodes and then signs and records the transaction. v, d, and m denote the view number, the client message digest, and the message content, respectively.
- Upon receiving the Pre-Prepare message, the ith node aggregates and consolidates the scores from many master nodes. The median of these scores is used to ascertain the ultimate score for each local update client. Subsequently, the ith node transmits a signed < PREPARE, v, n, d, i, s > to the master nodes and records the Pre-Prepare and Prepare messages. This method reduces the impact of outliers and guarantees an equitable evaluation.
- The primary nodes validate the accuracy of Prepare and disseminate. Upon receiving a validated Prepare broadcast from all master nodes, node i transmits a signed < COMMIT, v, n, d, i > message to the master nodes and logs the commit message.
- Ultimately, if master nodes receive commit messages and the score exceeds a predetermined threshold, they accept the request operation o and return < REPLY, s > to the local update client, either uploading the parameters from the client node to the blockchain or downloading the updated parameters from the blockchain. When the update client gets uniform reply messages, the request achieves consensus throughout the whole network.

At the conclusion of each round, a new master node group is selected based on the scores of nodes from the preceding round. Nodes with elevated scores are more probable to be selected, guaranteeing that the masters consist of dependable and high-performing nodes."

4.2.3 Smart contracts

The concept of smart contracts is established in a blockchain to provide secure technological infrastructure, which ensures healthcare and medical information integrity, enhances the quality of services, end user coordination, and efficiency of cooperation of data.

4.2.4 Data Application Layer

The healthcare and medical data services covered under its data application layer include electronic health records (EHR) service and electronic medical records (EMR) service, clinical research and development, biomedical development, neuroscience research, genome engineering as well as medical image identification as depicted in Figure 3. Such a network is referred to as a consortium blockchain, whereby, a subset of the nodes, which all players can access, are involved in the consensus process, but a subset of the nodes is, at the permission of named parties, obligated to participate. To illustrate the application services in detail, we dwell upon the sample case: the services of data cooperation of clinical researches and development.

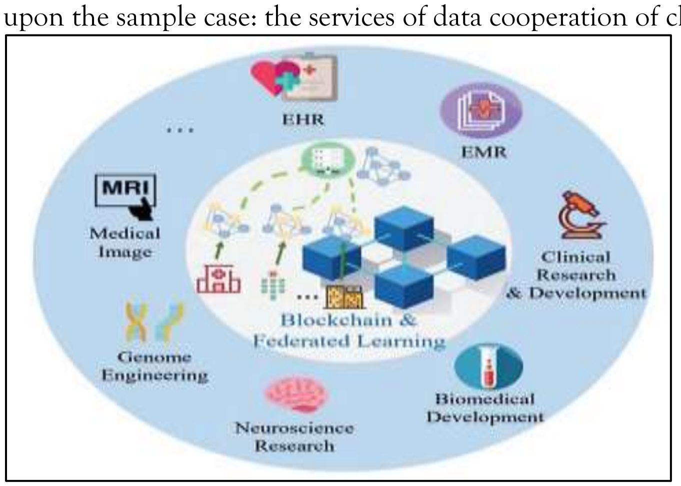

Figure 3. Data collaboration services in FL-HM Chain

4.2.5 CNN-based Federated learning model in healthcare and medical Blockchain

Through the development of a Convolutional Neural Network-based Federated Learning Model in the Healthcare and Medical Chain (FL-CNNHMChain) (depicted in Figure 4), we are able to validate the efficacy and dependability of FL-HMChain. This model is comprised of the following components:

4.2.6 CNN model constructing

This federated learning structure involves training local models on a six-layer CNN model and aggregating the server into a single server. The kernel size of this Convolutional CNN model is 3 x 3 in the first convolutional layer. The second dimension of convolutional layer is 3 3 and a maximum pooling layer of 2 2 and a stride of 2. The maximum pooling layer reduces the spatial dimension of the input feature map as well as preserving the most important features of the input feature map. The size of the kernels of the convolutional layers 3 and 4 is 3 x 3. The fifth convolutional layer is 3 x 3 having a maximum pooling layer of 2 x 2 and stride of 2. The BatchNorm2D is applied to every convolutional layer to accelerate the convergence rate and avoid over fitting.

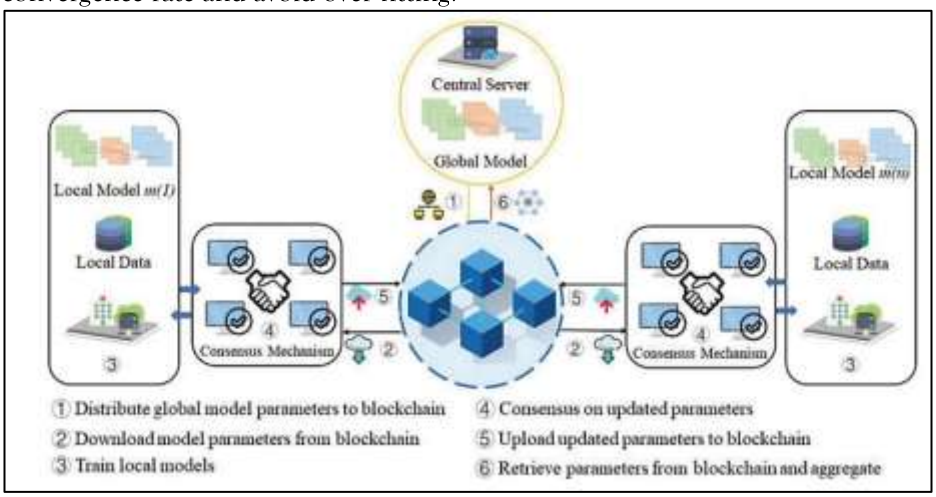

Figure 4. The framework of FL-CNN-HMChain

The final interdependent layer provides the prediction outcomes. The activation of all layers is set to ReLU. The adopted optimizer is stochastic gradient descent (SGD) that is applied to update the model parameters, after every iteration, to minimize the objective function in question.

5. Experiment

For the purpose of contrasting the FL-CNN-HMChain model with a generic CNN model, we make use of the 2D image data sets that are included in MedMNIST v2. These data sets include PneumoniaMNIST, BreastMNIST, TissueMNIST, and OrganAMNIST. Large-scale standardized biomedical pictures are included in the MedMNIST version 2 database.

5.1 Dataset

Before any of the two-dimensional pictures are handled, they are all preprocessed into 28 × 28 pixels with labels that are already known. A total of 5856 pediatric chest X-ray pictures serve as the foundation for the PneumoniaMNIST. These photos are divided into three categories: 4708 training data, 524 validation data, and 624 test data. In addition to the normal courses, it has pneumonia. BreastMNIST is particularly concerned with breast ultrasonography which has 780 samples of which 546 samples are training, 78 are validation and 156 are testing. It is divided into malignant and benign (normal) too and has 165,466 training data, 23,640 validation data, 47, 280 test data. It is connected to eight classes the collecting duct and connecting tubule, the distal convoluted tubule, the glomerular endothelial cells, the interstitial endothelial cells, the leukocytes, the podocytes, the proximal tubule segments, and the thick ascending limb. The OrganAMNIST dataset consists of 58,850 abdominal computer tomography (CT) samples. The total number of training data stands at 34,581, validation data is 6491 and the test data amount to 17,778. This is divided into eleven groups, the following; the spleen, the pancreas, the femur-left, the femur-right, the heart, the kidney-left, the kidney-right, the liver, the lung-left and lung-right."

5.2 Results

The evaluation metric used was AUC, which provides a robust measure of model accuracy across different threshold values.

Table 2 presents the AUC performance of three models—VGG16, CNN, and FL-CNN-HMChain—across four medical imaging datasets: Pneumonia MNIST, Breast MNIST, TissueMNIST, and Organ AMNIST. VGG16 shows the weakest performance, with AUC values mostly below 75%, and particularly low scores in TissueMNIST and Organ AMNIST, indicating poor generalization to medical datasets. The CNN model performs significantly better, achieving high AUCs (up to 97.6) on most datasets, though its performance slightly drops under federated settings, likely due to decentralized data access. FL-CNN-HMChain outperforms both VGG16 and CNN across all tasks, with AUC values consistently above 85%, reaching up to 98.9 for Organ AMNIST, demonstrating the robustness of federated learning combined with hybrid model chaining. Overall, FL-CNN-HMChain proves to be the most effective and reliable approach for secure and accurate classification in healthcare imaging tasks.

Table 2. Model comparison on AUC and ACC with different datasets

Model	Metric	Pneumonia MNIST	Breast MNIST	Tissue MNIST	Organ AMNIST
VGG16	ACC	54.1	50.3	49.5	46.9
	AUC	62.5	73.1	32.1	18.5
CNN	ACC	92.1	77.6	86.1	97.6
	AUC	83.7	73.1	50.4	80.8
FL-CNN-HMChain	ACC	96.2	87.5	89.6	98.9
	AUC	87.5	79.5	62.1	86.8

Figure 5 illustrates the AUC performance of three models—VGG16, CNN, and FL-CNN-HMChain—across four medical imaging datasets: Pneumonia MNIST, Breast MNIST, TissueMNIST, and Organ AMNIST. Each model is evaluated in two settings, likely centralized and federated, with AUC scores represented by colored bars. VGG16 shows the lowest performance, with AUC values often below 60 and particularly poor results on Organ AMNIST and TissueMNIST. CNN performs significantly better, achieving high AUCs, especially on Pneumonia and Organ AMNIST, though a performance dip is observed in one setting. FL-CNN-HMChain outperforms both models consistently across all datasets, maintaining AUC scores above 85 and reaching up to 98.9, indicating its superior accuracy and robustness, especially in federated or privacy-sensitive environments.

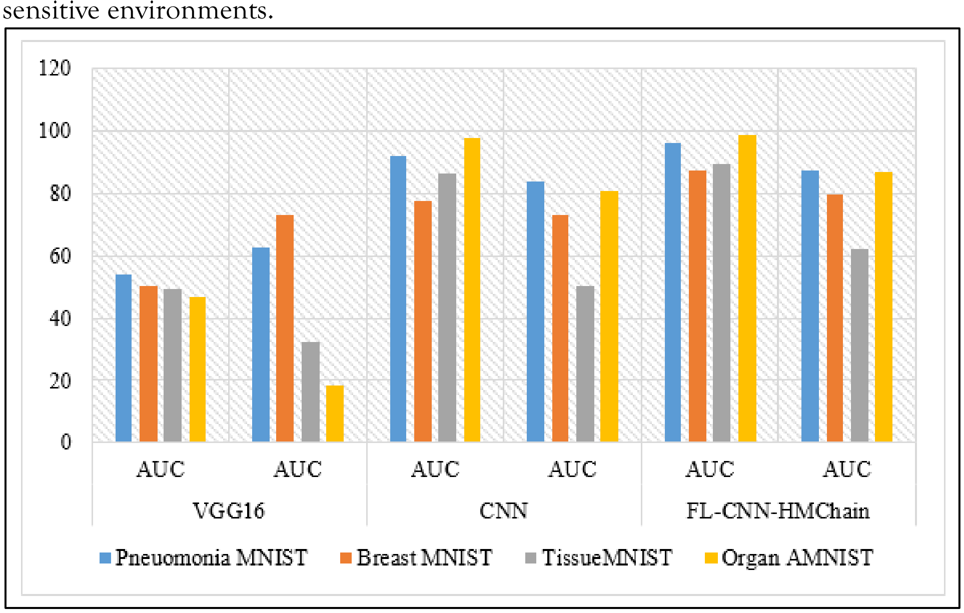

Figure 5. Comparative analysis of different algorithms

6. CONCLUSION AND FUTURE WORK

This study is a powerful and privacy-conscious AI architecture that is a federated learning and blockchain system that enables privacy-preserving and secure usage of medical information. The proposed system can be practical in the minimization of the negativity of the traditional centralized AI models, such as the threat of data privacy, institutional mistrust, and non-observation of regulations. The architecture is transparent, data integrity and trusted since it allows different healthcare institutions to train AI models, collaboratively without sharing raw data, and through blockchain to verify model updates under consensus and smart contracts. The results obtained through experiments with various healthcare datasets, such as Pneumonia MNIST, Breast MNIST, TissueMNIST, and Organ AMNIST show that the FL-CNN-HMChain model is much more effective than baseline models, such as CNN and VGG16, in terms of AUC since they manage to achieve up to 98.9% accuracy. These results support the fact that this architecture has a potential to advance safe and high-performance AI solutions in practice within the medical and research context. Federated learning can be made more efficient in terms of communication efficiency in the future by using model compression methods including pruning and quantization to minimize the bandwidth and latency.