JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A BLOCKCHAIN-INTEGRATED FRAMEWORK FOR SECURE FEDERATED LEARNING OF ENCRYPTED EHR DATA WITH HOMOMORPHIC ENCRYPTION.

[1]Goranti Ramya, [2]Ganji Mahithi, [3]Dr.Shivani Yadao

[12]PG Scholars, [3]Asso.Professor
[1,2,3]Department of Computer Science Engineering,
[1,2,3]Stanley College of Engineering and Technology for Women, Hyderabad, India.

Abstract : The rapid digital transformation of the healthcare sector has led to widespread adoption of Electronic Health Records (EHRs), which contain sensitive patient information and play a crucial role in modern medical systems. Although machine learning techniques have significantly improved disease prediction and healthcare analytics, sharing data across healthcare organizations remains a major challenge due to concerns related to privacy and security. Federated Learning (FL) has emerged as a promising solution by enabling collaborative model training without the need to exchange raw data; however, existing FL approaches are often centralized and susceptible to risks such as data leakage during model updates and potential misuse by malicious participants. To address these limitations, this study proposes a secure framework that combines federated learning with blockchain technology and homomorphic encryption. In this approach, EHR data is encrypted prior to model training, allowing computations to be performed directly on encrypted data, thereby preserving confidentiality. A permissioned blockchain network is utilized to manage and record model updates, ensuring transparency, traceability, and trust through the implementation of smart contracts. The framework is evaluated using the Chronic Kidney Disease dataset, and the results demonstrate that it achieves high prediction accuracy while maintaining strong standards of data privacy, security, and reliability, making it a robust solution for secure healthcare analytics.

***Index Terms* - Federated Learning, Blockchain, Homomorphic Encryption, Electronic Health Records (EHR), Privacy-Preserving Machine Learning, Healthcare Data Security.**

I. INTRODUCTION

The continuous advancement of digital technologies in healthcare has significantly increased the adoption of Electronic Health Records (EHRs) for storing vital patient information. With the growing volume of healthcare data, machine learning techniques have become highly valuable for improving disease prediction, diagnosis, and treatment outcomes. However, the sensitive nature of medical data raises serious concerns regarding privacy and security, especially when data needs to be shared across multiple healthcare organizations. Traditional machine learning approaches typically rely on collecting and storing patient data in a centralized system, which can lead to potential violations of healthcare regulations and increased risk of data breaches. To address these issues, Federated Learning (FL) has emerged as an effective alternative, enabling multiple organizations to collaboratively train models without exchanging raw data. In this approach, each organization trains a local model using its own private data and shares only the model parameters with a global system, thereby preserving data privacy through decentralization. Despite these advantages, conventional FL systems still face challenges, such as the risk of sensitive information leakage through model updates, dependence on a central server that can become a single point of failure, and vulnerability to malicious participants who may disrupt the learning process. To enhance the security of federated learning, integrating advanced protection mechanisms has become an important area of research. One such powerful technique is homomorphic encryption, which allows computations to be performed directly on encrypted data without exposing the original information, thereby ensuring a higher level of data confidentiality and security.

II. LITERATURE SURVEY

Recent studies have increasingly focused on developing privacy-preserving approaches for collaborative healthcare analytics by combining federated learning, encryption techniques, and blockchain technology. Federated Learning (FL) has gained significant attention as it enables multiple healthcare institutions to collaboratively build machine learning models without directly sharing sensitive patient data, thereby supporting privacy requirements and regulatory compliance. In this approach, each institution trains models locally using its own data, and only the model parameters are shared and aggregated to form a global model, ensuring a decentralized learning environment. To further strengthen security, researchers have explored the integration of advanced cryptographic methods within federated learning frameworks. Among these, homomorphic encryption has emerged as a highly effective technique, as it allows computations to be performed directly on encrypted data without exposing the original information, thus preserving the confidentiality of Electronic Health Records (EHRs) and reducing the risk of data leakage during model training and aggregation. In addition, blockchain technology has been incorporated into federated learning systems to enhance transparency, trust, and security. By providing a decentralized and tamper-resistant ledger, blockchain enables secure recording and verification of model updates across participating institutions, while also reducing reliance on a central server. Several research contributions highlight the effectiveness of these combined approaches. For instance, studies by Firdaus et al. (2025), Yang et al. (2024), and Munusamy et al. (2025) demonstrate the use of blockchain-based federated learning frameworks to improve privacy, reliability, and trust in healthcare data sharing. Similarly, Wani et al. (2025) introduced a decentralized FL-based framework for secure healthcare analytics, while Zhang et al. (2025) emphasized the importance of secure aggregation techniques using homomorphic encryption. Kumar et al. (2025) provided a comprehensive review of privacy-preserving FL methods and cryptographic techniques in decentralized healthcare systems. Other contributions, such as those by Liang et al. (2023), Kumar et al. (2021), Walskaar et al. (2023), and Salim et al. (2024), further highlight the role of integrating blockchain and encryption methods with federated learning to achieve secure, scalable, and efficient healthcare analytics. Overall, these studies collectively demonstrate the growing importance of combining federated learning, cryptography, and blockchain to address privacy, security, and trust challenges in modern healthcare systems.

III . PROPOSED SYSTEM

The proposed framework presents a secure and privacy-focused approach for collaborative healthcare analytics by integrating Federated Learning, Homomorphic Encryption, and Blockchain technology. Its main objective is to enable multiple healthcare institutions to jointly train machine learning models on Electronic Health Records (EHRs) without directly sharing sensitive patient data, thereby maintaining privacy while supporting accurate predictive analysis. The process begins with the collection of EHR datasets from participating organizations, which typically include important patient details such as medical history, symptoms, laboratory results, and diagnoses. Before model training, the data undergoes preprocessing steps, including handling missing values, converting categorical data into numerical form, and normalizing features to improve model performance. To ensure data confidentiality, CKKS-based homomorphic encryption is applied, allowing computations to be performed on encrypted data without exposing the original information. The encrypted data is then distributed among multiple federated clients, each representing a different healthcare organization. Each client independently trains a machine learning model using XGBoost, chosen for its strong performance and efficiency in classification tasks, with training performed solely on local data. Once training is completed, instead of transmitting model parameters to a centralized server, the updates are securely recorded on a blockchain network. Smart contracts are used to validate and store these updates, with model parameters treated as transactions within the blockchain. These transactions are organized into blocks, where each block is cryptographically linked to the next through hash pointers, ensuring data integrity, transparency, and resistance to tampering throughout the collaborative learning process. The figure illustrates the overall workflow of the proposed system, outlining each stage of the process in a structured manner. It begins with the collection of Electronic Health Record (EHR) datasets, which include essential patient information such as medical history, symptoms, and diagnostic results. This data is then preprocessed to improve its quality and suitability for model training by handling missing values, converting categorical features into numerical form, and applying normalization techniques. To ensure patient privacy, the processed data is encrypted using the CKKS homomorphic encryption scheme, allowing secure computations without exposing sensitive information. The encrypted data is subsequently distributed among multiple federated clients, each representing a different healthcare institution. These clients independently train machine learning models using the XGBoost algorithm based on their local data. Instead of sharing raw data, the trained model parameters are transmitted to a blockchain network, where they are verified, securely recorded, and aggregated in a privacy-preserving manner to form a global federated model. Finally, the global model is deployed back to the participating clients, where it is used to predict disease outcomes on test data. The system then generates results that reflect its effectiveness, including metrics such as prediction accuracy and computational efficiency. The decentralized nature of the system allows multiple healthcare organizations to contribute to model development without compromising data ownership or control. Overall, this approach creates a balanced solution that addresses key challenges in modern healthcare systems, enabling secure, efficient, and trustworthy predictive analytics while maintaining high standards of data confidentiality.

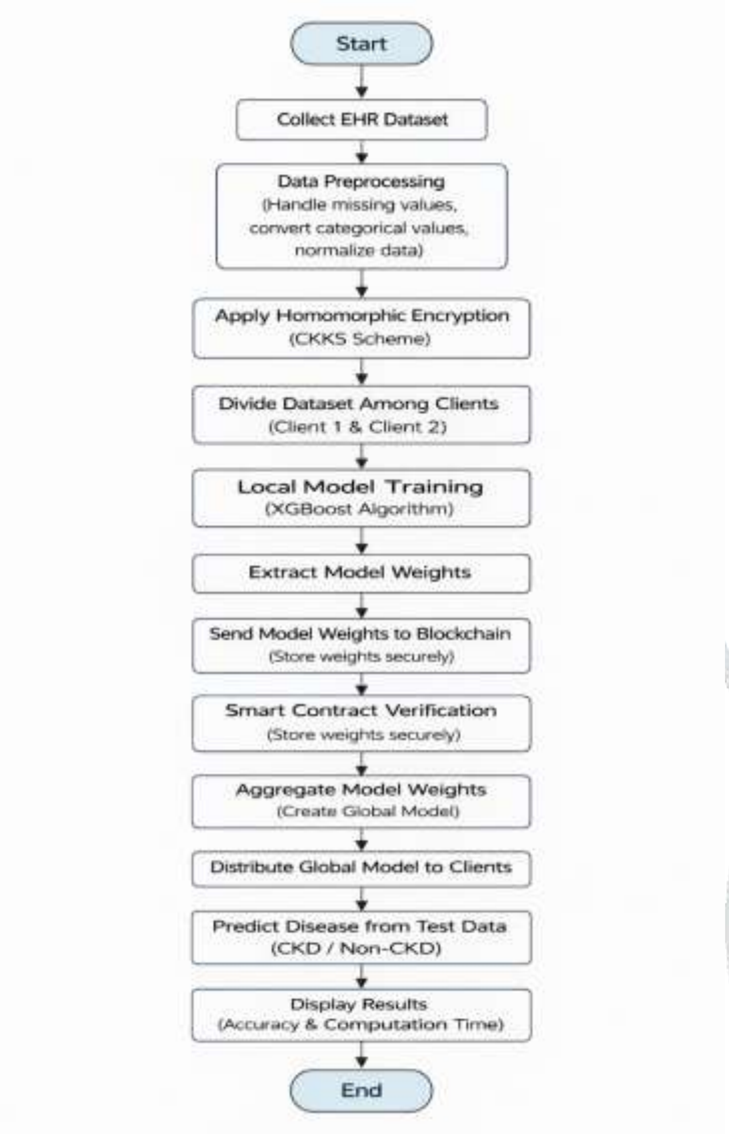

Start
Collect EHR Dataset
Data Preprocessing (Handle missing values, convert categorical values, normalize data)
Apply Homomorphic Encryption (CKKS Scheme)
Divide Dataset Among Clients (Client 1 & Client 2)
Local Model Training (XGBoost Algorithm)
Extract Model Weights
Send Model Weights to Blockchain (Store weights securely)
Smart Contract Verification (Store weights securely)
Aggregate Model Weights (Create Global Model)
Distribute Global Model to Clients
Predict Disease from Test Data (CKD / Non-CKD)
Display Results (Accuracy & Computation Time)
End

Fig 1 : Secure federated learning for EHR system

ALGORITHMS:

XGBoost Algorithm : It is a powerful supervised learning algorithm widely used for both classification and regression problems due to its high performance and efficiency. It is an advanced implementation of the gradient boosting technique, where multiple decision trees are built sequentially to improve prediction accuracy by reducing errors from previous models. The core idea behind XGBoost is to combine several weak learners into a strong predictive model that can capture complex patterns within the data. In the proposed framework, XGBoost is used to train local models at each participating healthcare institution within the federated learning setup. This algorithm is chosen because of its speed, scalability, and capability to effectively handle missing data and intricate relationships among features. Additionally, it incorporates regularization techniques that help prevent overfitting, making the model more robust. During training, XGBoost iteratively optimizes the model by minimizing a defined loss function using gradient-based methods, which enhances overall prediction accuracy. Upon completion of the training process, the model generates parameters that encapsulate the learned patterns, which are then utilized in the collaborative learning framework.

Homomorphic Encryption (CKKS Scheme) : Homomorphic Encryption (HE) is an advanced cryptographic technique that allows computations to be performed directly on encrypted data without the need to decrypt it, making it highly suitable for privacy-sensitive domains such as healthcare. In this study, the CKKS (Cheon–Kim–Kim–Song) homomorphic encryption scheme is adopted due to its ability to support approximate arithmetic operations, which are particularly useful for machine learning applications involving numerical data. Before the training process begins, the sensitive information contained in Electronic Health Records (EHRs) is encrypted using the CKKS scheme, ensuring that the data remains confidential at all times. This approach enables machine learning models to operate on encrypted inputs, allowing computations such as training and aggregation to be carried out securely without exposing the underlying data. As a result, even during collaborative processes like federated learning, patient information remains protected, thereby maintaining data privacy while still enabling effective model development.

Federated Learning Algorithm : Federated Learning (FL) is a collaborative machine learning approach that enables multiple participants to build a shared global model without exchanging their raw data. Instead of sending sensitive data to a central server, each participant, or client, trains a local model using its own dataset and shares only the learned model parameters. In the proposed framework, this concept is applied to connect multiple healthcare institutions, allowing them to jointly develop a predictive model while preserving data privacy. Each client trains an XGBoost model on its locally encrypted data and submits the resulting model updates to a blockchain network for secure handling. These updates are then aggregated to form a global model that reflects the collective learning of all participants. This iterative process is repeated over multiple rounds, continuously refining the model until the desired level of accuracy and performance is achieved, ensuring both effective learning and strong data protection.

Blockchain Smart Contract Algorithm : In the proposed framework, blockchain technology is utilized to ensure decentralization, security, and immutability of model updates within the federated learning process. Smart contracts, developed using Solidity, are employed to regulate and automate the interactions between federated clients and the blockchain network. After each client completes its local training and obtains the model parameters, these updates are submitted to the blockchain as transactions rather than being sent to a centralized server. The smart contract is responsible for validating the authenticity and integrity of the submitted parameters before they are accepted. Once verified, the updates are permanently recorded on the blockchain, where each transaction is assigned a unique cryptographic hash, making it resistant to tampering or unauthorized modification. The stored model parameters are then securely aggregated through the blockchain mechanism to form a global model. By removing dependence on a central authority, this approach enhances transparency, strengthens trust among participating entities, and ensures a secure and reliable collaborative learning environment.

IV .SYSTEM IMPLEMENTATION

The proposed system integrates Federated Learning, Homomorphic Encryption, and Blockchain technology to enable secure and privacy-preserving analytics on healthcare data. It is implemented using the Python programming language, combining machine learning techniques with blockchain mechanisms to allow model training directly on encrypted Electronic Health Records (EHRs) without revealing sensitive patient information. The overall implementation follows a structured workflow that includes data preprocessing, encryption of sensitive information, decentralized model training through federated learning, secure management of model updates using blockchain, and performance evaluation. Each stage is designed to ensure data confidentiality, system reliability, and effective predictive analysis within a distributed healthcare environment

Data Collection and Preprocessing : The process begins with the collection of the Chronic Kidney Disease (CKD) Electronic Health Records dataset, which contains important clinical attributes such as blood pressure, glucose levels, and hemoglobin values that are essential for identifying the presence of kidney disease. To ensure the data is suitable for model training, a preprocessing phase is carried out to enhance its quality and consistency. During this stage, missing values are handled by replacing them with appropriate statistical measures, such as the mean of the respective features. Categorical variables are then transformed into numerical representations using label encoding techniques, allowing them to be processed by machine learning algorithms. Finally, normalization is applied to scale all features to a uniform range, which helps improve the efficiency and performance of the learning model. This structured preprocessing ensures that the dataset is clean, consistent, and ready for effective analysis.

Homomorphic Encryption Implementation : To protect patient confidentiality, the system incorporates Homomorphic Encryption as a key security mechanism. Specifically, the CKKS encryption scheme is implemented using the TenSEAL library to enable secure data processing. In this approach, the healthcare data is encrypted before it is used for model training, ensuring that sensitive information is never exposed in its original form. The CKKS scheme supports arithmetic operations on encrypted data, allowing the machine learning model to be trained without decrypting the underlying values. This ensures that privacy is maintained throughout the computation process while still enabling effective analysis and model development.

Federated Learning Model Training : In the next phase, the encrypted dataset is distributed among multiple clients, with each client representing a separate healthcare institution. These clients utilize their respective local data to train machine learning models independently, ensuring that sensitive information does not leave their environment. The XGBoost algorithm is employed for this purpose due to its high accuracy, computational efficiency, and strong performance on structured healthcare data. Each client trains its model locally on the encrypted data and then shares only the learned model parameters, rather than the raw data, with the overall system. This decentralized training approach ensures that patient data remains private while still enabling collaborative model development across institutions.

Blockchain Integration : To ensure secure and reliable sharing of model parameters, the system incorporates a blockchain network based on Ethereum. Smart contracts are utilized to manage and automate the interactions within this network. After completing local model training, each client submits its trained model parameters to the blockchain instead of sending them to a centralized server. These parameters are then validated by the smart contract to ensure their authenticity and integrity before being accepted. Once verified, they are recorded on the blockchain as transactions, each containing details such as transaction hash and block identifiers, which provide traceability and security. This decentralized approach ensures that model updates are stored in a tamper-resistant manner, eliminates dependence on a central authority, and significantly enhances the overall security and trustworthiness of the system.

Global Model Aggregation : After all participating clients have submitted their locally trained model parameters to the blockchain, these updates are securely combined to generate a unified global model. This aggregation process integrates insights learned from multiple healthcare institutions, resulting in a more robust and generalized model while still preserving the uniqueness of local data contributions. The aggregated global model is then distributed back to the clients, where it can be used for further training iterations or for making predictions. This collaborative cycle continues over multiple rounds, progressively improving the model’s performance until the desired level of accuracy and reliability is achieved.

Disease Prediction : Once the global model is established, it is used to perform predictions on encrypted test data to ensure that patient information remains confidential throughout the process. The model evaluates the input features and classifies each case as either Chronic Kidney Disease (CKD) or non-CKD. By analyzing the prediction outcomes, it can be observed that the model effectively captures underlying disease patterns and provides reliable classification results. This demonstrates the capability of the proposed system to deliver accurate predictions while maintaining strict data privacy and security standards.

V. RESULT & ANALYSIS

For evaluating the performance of the proposed system, the Chronic Kidney Disease (CKD) Electronic Health Records dataset is utilized, which includes key clinical attributes such as blood pressure, hemoglobin levels, and glucose concentration to determine the presence of kidney disease in patients. The dataset is divided into training and testing subsets to assess the effectiveness of the model. Initially, the data is encrypted using the CKKS homomorphic encryption scheme to ensure privacy. The encrypted data is then distributed across multiple clients within a federated learning environment, where each client independently trains a local model using the XGBoost algorithm. After training, the model parameters are securely transmitted to a blockchain network, where they are verified and aggregated to form a global model. The performance of the system is evaluated using standard classification metrics, including accuracy, precision, and recall, which demonstrate the model's ability to make reliable and consistent predictions while maintaining data security throughout the process.

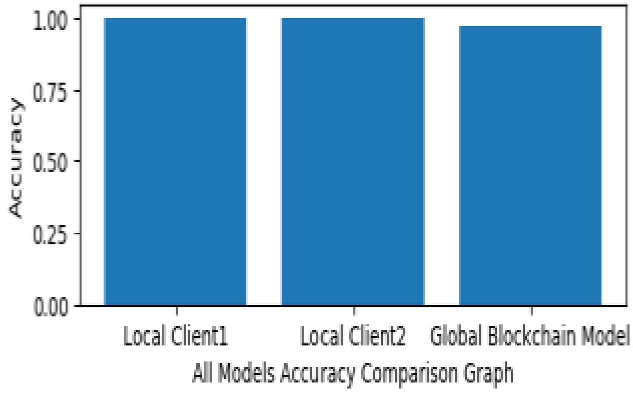
Model Type	Algorithm Used	Accuracy (%)	Description
Client 1 Local Model	XGBoost	97%	Model trained on first half of the encrypted dataset
Client 2 Local Model	XGBoost	92%	Model trained on second half of the encrypted dataset
Global federated Model	XGBoost	96%	Model generated by aggregating weights from both clients

Table 1 : Accuracy comparision of local and global models

Accuracy : Accuracy measures the number of correct predictions relative to the total number of cases. Accuracy is measured as shown below:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Where:
TP = True Positives (CKD)
TN = True Negatives (Non-CKD)
FP = False Positives (Non-CKD predicted as CKD)
FN = False Negatives (CKD predicted as Non-CKD)



Precision : Precision measures the number of actual CKD cases relative to the total number of cases predicted as CKD. Precision is measured as shown below:

Precision = TP / (TP + FP)

Precision measures the reliability of the model when predicting CKD cases.

Recall : The recall measure indicates the ability of the model to recall all the actual positive values. It can be calculated in the following way:

Recall = TP / (TP + FN)

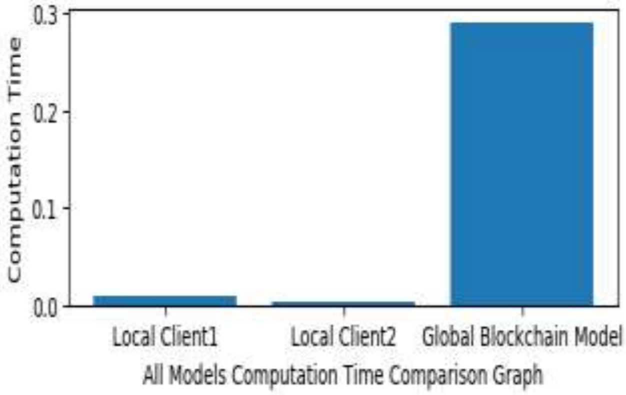
F1 Score : The F1 Score combines the precision and recall to create a comprehensive view of the model. This is useful in situations where there is class imbalance in the data.

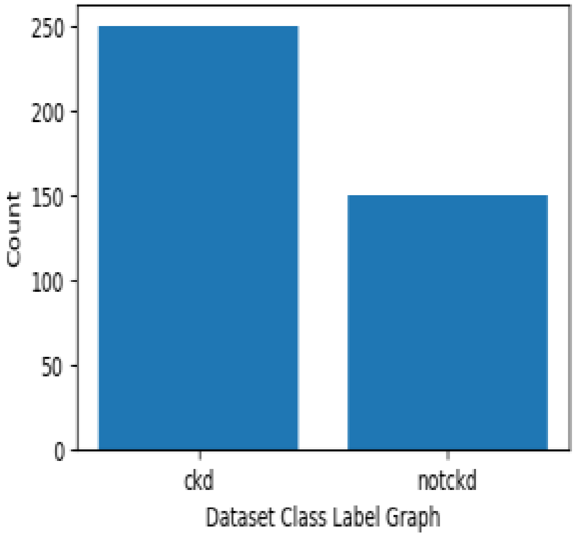
F1 Score = 2 * (Precision * Recall) / (Precision + Recall)

Computation time : In addition to accuracy, we have also considered the computation time taken by the local models and the global federated model. Computation time is calculated in the following way:

Computation Time = End Time - Start Time

From the results, we can see that the local models are accurate, and the global federated model, in addition to accuracy, offers better security and privacy. Although the addition of blockchain and homomorphic encryption complicates the model, they offer better health data security.





VI. FUTURE WORK

Although the proposed blockchain-enabled federated learning framework demonstrates strong potential for privacy-preserving healthcare analytics, several enhancements can be explored in future work to further improve its effectiveness. One possible direction is the integration of advanced deep learning techniques such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which are well-suited for handling large and complex medical datasets and can improve predictive performance. Additionally, while homomorphic encryption provides robust data security, it introduces computational overhead; therefore, future research can focus on optimizing existing encryption schemes or developing more efficient alternatives to reduce processing time. The current implementation considers a limited number of clients to simulate the federated environment, but expanding this to include a larger number of healthcare institutions would make the system more practical and scalable. Improvements can also be made by adopting more efficient blockchain consensus mechanisms to enhance system performance. Furthermore, the integration of Internet of Medical Things (IoMT) devices could enable real-time health monitoring by continuously collecting patient data while maintaining privacy through federated learning. Another important area for future development is the incorporation of explainable AI techniques, which can provide clearer insights into model predictions and support better decision-making in healthcare applications.

VII. CONCLUSION

The proposed framework presents a secure and privacy-focused approach for collaborative healthcare analytics by integrating Federated Learning, Homomorphic Encryption, and Blockchain technologies into a unified system. It effectively addresses key challenges associated with healthcare data sharing, including privacy, security, and trust, by enabling model training directly on encrypted Electronic Health Records within individual organizations. This approach eliminates the need to share raw data, thereby reducing the risk of data leakage during the learning process. Homomorphic Encryption ensures that all computations are performed on encrypted data, preserving confidentiality at every stage, while blockchain technology facilitates decentralized management of model updates through smart contracts, removing reliance on a central authority and enhancing transparency among participating entities. Experimental results demonstrate that the framework achieves high predictive accuracy while maintaining strong privacy protection. Although the inclusion of encryption and blockchain introduces additional computational overhead, the benefits in terms of data security, integrity, and reliable collaboration make it a practical solution for real-world healthcare applications. Overall, the proposed system establishes a secure foundation for medical data analysis and supports the advancement of intelligent, privacy-preserving healthcare systems.

VIII. REFERENCES

[1]Kumar, R., Marchang, N., & Tripathi, R. (2022). Blockchain-based federated learning framework for secure healthcare data sharing. Computers in Biology and Medicine, 146, 105604.

[2]Chang, Y., Weng, W., & Chen, C. (2021). A blockchain-based federated learning method for smart healthcare. Wireless Communications and Mobile Computing.

[3]Firdaus, M., Rauf, A., Kumar, R., & Khan, R. (2025). Blockchain-enabled federated learning for healthcare systems. Journal of Information Security and Applications.

[4]Wang, X., Li, Y., & Zhang, H. (2026). Securing federated learning with blockchain in healthcare applications. Journal of Medical Internet Research.

[5]Pati, S., Singh, P., & Sharma, V. (2024). Privacy preservation techniques for federated learning in healthcare systems. Artificial Intelligence in Medicine.

[6]Liang, X., Shetty, S., & Tosh, D. (2023). Architectural design of blockchain-enabled federated learning in healthcare. Journal of Medical Internet Research.

[7]Bhasker, B., & Sharma, K. (2025). Federated blockchain IoT framework for secure healthcare monitoring systems. Sensors.

[8]Zhang, L., Xu, J., Vijayakumar, P., Sharma, P., & Ghosh, U. (2022). Homomorphic encryption-based privacy-preserving federated learning in IoT healthcare systems.

[9]Walskaar, I., Wibawa, F., & Borgaonkar, R. (2023). Practical implementation of privacy-preserving federated learning using homomorphic encryption. *Cryptography Journal*.

[10]Kurniawan, H., & Ahmad, F. (2022). Homomorphic encryption-based federated privacy preservation for healthcare data. *IEEE Access*.

[11]Khan, S., & Hussain, M. (2024). Advancing medical innovation through blockchain and federated learning integration. *Health Informatics Journal*.

[12]Qammar, A., Karim, F., & Rehman, S. (2023). Securing federated learning with blockchain: A systematic review. *Artificial Intelligence Review*.

[13]Rehman, A., Ullah, I., & Khan, S. (2022). Secure healthcare systems using blockchain and federated learning. *Computers in Biology and Medicine*.

[14]Ahmed, S. T., Kim, Y., & Lee, S. (2024). Blockchain-based federated learning approach for AI-enabled healthcare devices. *BMC Medical Imaging*.

[15]Al-Janabi, A. A., & Al-Shourbaji, I. (2024). Privacy-preserving federated transfer learning for healthcare data analysis.

[16]Alqazzaz, A., & Rahman, M. (2025). Federated learning with efficient homomorphic encryption for secure data analytics. *Journal of Big Data*.

[17]Nehal, M., & Hussain, A. (2025). Secure federated learning in healthcare using blockchain and secure multi-party computation.

[18]Limbepe, Z., & Kouam, J. (2025). Blockchain-based privacy-enhancing federated learning for healthcare systems. *Future Internet*.

[19]Choi, G., Kim, D., & Lee, J. (2024). Survey of medical applications of federated learning. *Healthcare Informatics Research*.

[20]Hegde, M. G., & Patil, S. (2025). Secure federated learning with homomorphic encryption for privacy protection.

[21]Wibawa, F., Borgaonkar, R., & Kanhere, S. (2022). Privacy-preserving machine learning with homomorphic encryption in healthcare.

[22]Rangwala, M., & Shah, S. (2025). Blockchain-enabled federated learning: Architecture and security challenges.

[23]Samantray, B. S., & Mishra, A. (2025). Hybrid blockchain-federated learning framework for smart healthcare systems.

[24] Sun, R., Wang, Z., Zhang, H., Jiang, M., Wen, Y., & Liu, E. (2024). Multi-continental healthcare modelling using blockchain-enabled federated learning.

[25]Waheed, N., Rehman, A., Nehra, A., Farooq, M., Tariq, N., Jan, M. A., & Nanda, P. (2023). FedBlockHealth: Federated learning and blockchain for secure IoT-enabled healthcare systems.