ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 2, 2026

Peer Reviewed, Referred & Indexed Journal

A BLOCKCHAIN-INTEGRATED PRIVACY-PRESERVING MACHINE LEARNING MODEL FOR CHRONIC KIDNEY DISEASE RISK ASSESSMENT

Mrs.Md. Parveen Begum[1],Maddi Venkata Naga Anjani Sai Sanjana[2], Manam Srilakshmi[3], Meda Lakshmi Triveni[4],Mogili Annamani[5]

[1]Assistant Professor, Department of Computer Science and Engineering,
KKR & KSR Institute of Technology and Sciences, Vinjanampadu, Vatticherukuru Mandal, Guntur, Andhra Pradesh 522017
Email: mohammedparveen@gmail.com[1]
[2345]UG Scholar, Department of Computer Science and Engineering,
KKR & KSR Institute of Technology and Sciences, Vinjanampadu, Vatticherukuru Mandal, Guntur, Andhra Pradesh 522017
Email: 22jr1a0576@gmail.com[2], 22jr1a0577@gmail.com[3], 22jr1a0581@gmail.com[4], 22jr1a0582@gmail.com[5]

Abstract: Chronic Kidney Disease (CKD) is a major global health concern that often progresses silently and leads to severe complications if not detected early. Machine learning techniques have shown promising results in predicting CKD; however, traditional systems rely on centralized data storage, which raises significant concerns regarding data privacy, security, and transparency. This research proposes a privacy-preserving decentralized framework for CKD prediction by integrating machine learning, blockchain technology, and the InterPlanetary File System (IPFS). In the proposed system, sensitive medical records are encrypted using AES and RSA algorithms before storage and processing to ensure confidentiality. Multiple machine learning algorithms, including Decision Tree, Random Forest, Logistic Regression, and K-Nearest Neighbors, are employed to analyze clinical parameters and predict CKD risk accurately. Blockchain technology ensures immutability, transparency, and secure transaction logging, while IPFS provides decentralized and reliable storage of encrypted medical data. Experimental evaluation demonstrates improved prediction accuracy, enhanced data security, and scalable healthcare analytics for trustworthy AI-driven medical diagnosis.

Keywords— *Chronic Kidney Disease (CKD), Machine Learning, Blockchain, InterPlanetary File System (IPFS), Healthcare Data Security, Decentralized Systems*

I. INTRODUCTION

Chronic Kidney Disease (CKD) is a progressive medical condition that affects the kidneys' ability to filter waste and excess fluids from the blood. Globally, millions of people suffer from CKD, and its prevalence continues to increase due to factors such as diabetes, hypertension, obesity, and aging populations. One major challenge is that CKD often develops silently in its early stages, showing minimal symptoms until severe kidney damage occurs. Therefore, early detection and timely treatment are essential to prevent complications such as kidney failure and cardiovascular diseases.

Artificial Intelligence (AI) and Machine Learning (ML) techniques have recently shown strong potential in improving disease prediction and clinical decision-making. ML algorithms such as Decision Tree, Random Forest, Logistic Regression, and K-Nearest Neighbors can analyze large medical datasets and identify patterns related to CKD risk factors. However, many existing systems rely on centralized data storage, which raises concerns about data privacy, security vulnerabilities, and single points of failure.

To address these challenges, this research proposes a privacy-preserving CKD prediction framework integrating machine learning, blockchain, IPFS-based storage, and encryption techniques like AES and RSA for secure and reliable healthcare data management.

II. Literature Survey

Several studies have explored the use of machine learning techniques for predicting Chronic Kidney Disease (CKD) using clinical data. Algorithms such as Decision Tree, Random Forest, Logistic Regression, and K-Nearest Neighbors have demonstrated promising results in early disease detection by analyzing medical parameters like blood pressure, serum creatinine, and glucose levels. However, most existing CKD prediction systems rely on centralized data storage, which raises concerns regarding data privacy, security, and transparency. Recent research has proposed integrating blockchain technology and decentralized storage solutions like IPFS to address these challenges. These approaches improve data security, ensure tamper-proof record management, and enable secure sharing of medical data among healthcare institution.

III. PROPOSED WORK

The proposed work presents a privacy-preserving and decentralized framework for predicting chronic kidney disease (CKD) by integrating machine learning techniques with blockchain technolmodel enables healthcare institutions to collaborate without directly sharing raw patient daogy and the InterPlanetary File System (IPFS). The main objective of this approach is to ensure secure handling of sensitive patient data while maintaining high prediction accuracy and transparency in healthcare analytics. Unlike traditional centralized systems, the proposed ta, thereby protecting patient privacy.In this framework, patient medical records containing clinical attributes such as blood pressure, serum creatinine, albumin levels, and glucose values are first encrypted using AES and RSA encryption algorithms to ensure confidentiality during storage and transmission. The encrypted data is then stored in a decentralized storage network using IPFS, which generates a unique Content Identifier (CID) for each record. This CID is recorded on the blockchain, ensuring data integrity, traceability, and immutability. For disease prediction, multiple machine learning algorithms such as Decision Tree, Random Forest, Logistic Regression, and K-Nearest Neighbors are implemented to analyze patient data and classify CKD risk. Ensemble learning techniques are used to combine the predictions of these models to improve overall accuracy and reliability. Additionally, blockchain-based smart contracts manage data access permissions and maintain transparent transaction logs. This integrated architecture ensures secure data management, reliable disease prediction, and trustworthy AI-driven healthcare system.

IV. METHODOLOGY

The proposed methodology focuses on developing a secure and decentralized framework for Chronic Kidney Disease (CKD) prediction using machine learning, blockchain technology, and IPFS. The system is designed to ensure accurate disease prediction while protecting sensitive patient data. The methodology includes several stages such as data collection, preprocessing, model training, encryption, and decentralized storage.

4.1 Data Collection

The first step in the methodology involves collecting medical datasets related to Chronic Kidney Disease. The dataset contains various clinical parameters such as age, blood pressure, albumin, blood glucose level, serum creatinine, hemoglobin, and other laboratory measurements. These attributes are used as input features for the machine learning models to identify patterns associated with CKD and non-CKD cases.

4.2 Data Preprocessing

Data preprocessing is performed to improve the quality of the dataset before model training. This step includes handling missing values, removing duplicate records, and converting categorical data into numerical format. Normalization and scaling techniques are applied to maintain uniformity among features. Proper preprocessing ensures better model performance and reduces the chances of prediction errors.

4.3 Feature Selection

Feature selection is used to identify the most relevant attributes that influence CKD prediction.

Techniques such as correlation analysis and statistical tests are applied to remove irrelevant or redundant features. Important features like serum creatinine, blood pressure, and glucose levels are selected as key indicators for predicting kidney disease.

4.4 Machine Learning Model Development

Multiple machine learning algorithms are implemented to perform CKD prediction. These include Decision Tree, Random Forest, Logistic Regression, and K-Nearest Neighbors (KNN). Each model is trained using the prepared dataset and evaluated using performance metrics such as accuracy, precision, recall, and F1-score. Ensemble learning techniques are also applied to combine model outputs and improve prediction reliability.

4.5 Data Encryption

To ensure privacy and security of patient data, encryption mechanisms are applied before storing or transmitting medical records. AES (Advanced Encryption Standard) is used for encrypting the data, while RSA encryption secures the exchange of encryption keys. This dual-layer security ensures that sensitive healthcare information remains protected from unauthorized access.

4.6 Blockchain Integration

Blockchain technology is used to maintain a transparent and tamper-proof record of all transactions related to CKD prediction. Smart contracts control data access permissions and record prediction results on the blockchain ledger. This ensures data integrity, accountability, and trust among healthcare providers and patients.

4.7 IPFS-Based Decentralized Storage

The InterPlanetary File System (IPFS) is used for decentralized storage of encrypted medical records. When data is uploaded to IPFS, a unique Content Identifier (CID) is generated. This CID is stored on the blockchain and used to retrieve the data when required. The combination of blockchain and IPFS ensures secure, scalable, and reliable data management.

4.8 Prediction and Result Generation

Finally, the trained machine learning models analyze the patient data and generate predictions indicating whether the patient is at risk of CKD or not. The prediction results are securely stored and can be accessed by authorized healthcare providers. This approach enables accurate diagnosis while maintaining privacy and transparency in healthcare systems

V. ALGORITHMS

The proposed Chronic Kidney Disease (CKD) prediction system uses multiple machine learning and security algorithms to ensure accurate disease prediction and secure handling of patient data. These algorithms work together to analyze medical data, generate reliable predictions, and protect sensitive healthcare information within a decentralized environment.

5.1 Decision Tree Algorithm

The Decision Tree algorithm is a supervised machine learning method used for classification tasks. It works by splitting the dataset into different branches based on important attributes such as blood pressure, albumin level, and serum creatinine. Each node in the tree represents a decision condition, and the final leaf node represents the prediction result (CKD or Non-CKD). Decision Trees are easy to interpret because they follow simple "if–then" rules, making them useful in healthcare applications where transparency and explainability are important.

5.2 Random Forest Algorithm

Random Forest is an ensemble learning algorithm that combines multiple decision trees to improve prediction accuracy and stability. Instead of relying on a single tree, the algorithm builds several trees using different subsets of data and features. The final prediction is obtained by majority voting among all trees. Random Forest reduces overfitting and performs well with complex medical datasets, making it highly suitable for CKD prediction.

5.3 Logistic Regression

Logistic Regression is a statistical classification algorithm used to predict binary outcomes. In this system, it determines whether a patient has CKD or not based on medical parameters such as glucose levels, blood pressure, and hemoglobin. The algorithm uses a sigmoid function to convert linear outputs into probability values between 0 and 1. Due to its simplicity and interpretability, Logistic Regression is widely used in medical research.

5.4 K-Nearest Neighbors (KNN)

K-Nearest Neighbors is a supervised learning algorithm that classifies a new data instance based on similarity to existing data points. It calculates the distance between the new patient record and stored records using distance measures such as Euclidean distance. The algorithm then assigns the class most common among the nearest neighbors. KNN is simple, effective, and performs well with healthcare datasets containing mixed attributes.

5.5 AES Encryption Algorithm

The Advanced Encryption Standard (AES) is used to secure patient medical records before storage or transmission. AES encrypts data using a symmetric key, ensuring that sensitive healthcare information remains confidential. This encryption protects patient data from unauthorized access during the CKD prediction process.

5.6 RSA Encryption Algorithm

RSA is an asymmetric encryption algorithm used for secure key exchange. It uses a pair of keys—public and private—to encrypt and decrypt data. In this system, RSA protects the encryption keys used by AES, ensuring secure communication between users, healthcare providers, and storage systems.

5.7 Blockchain Technology

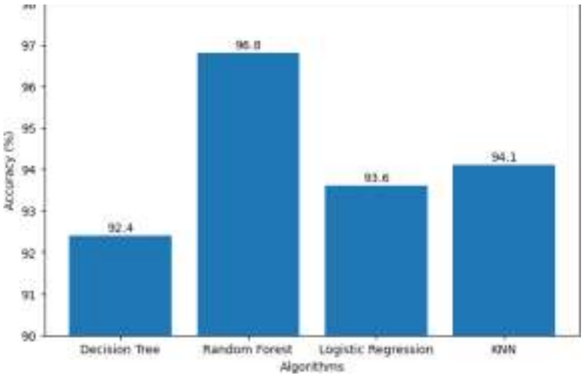
Blockchain technology is used to maintain a decentralized and tamper-proof record of CKD prediction transactions. Each transaction is stored as a block in a distributed ledger, ensuring transparency, security, and immutability. Smart contracts manage data access permissions and verify prediction records.

5.8 InterPlanetary File System (IPFS)

IPFS is a decentralized storage system used to store encrypted medical records and prediction results. When a file is uploaded to IPFS, it generates a unique Content Identifier (CID) based on the file's hash value. This CID is stored on the blockchain, allowing secure and efficient retrieval of patient data without relying on centralized servers

VI. RESULTS AND DISCUSSION

The proposed decentralized CKD prediction system was evaluated using several machine learning algorithms to measure prediction accuracy, performance efficiency, and system scalability. The models were trained using the preprocessed CKD dataset and evaluated using standard metrics such as accuracy, precision, recall, and F1-score. The experimental results show that ensemble-based methods such as Random Forest provide higher



accuracy compared to individual algorithms. Additionally, the integration of blockchain and IPFS ensures secure storage and transparent access to prediction records without compromising data privacy.The system was also tested for computational

performance and response time. Encryption mechanisms such as AES and RSA ensured secure data transmission, while blockchain guaranteed immutability of prediction logs. IPFS storage enabled decentralized file retrieval, eliminating single points of failure. The experimental results demonstrate that the proposed system can efficiently process medical data, generate accurate CKD predictions, and maintain strong data security.

Table 1: Machine Learning Model Accuracy Comparison

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Decision Tree	92.4	91.7	90.8	91.2

Random Forest	96.8	96.2	95.9	96.0
Logistic Regression	93.6	92.9	92.3	92.6
KNN	94.1	93.5	92.8	93.1

Random Forest achieved the highest accuracy due to its ensemble nature, which reduces overfitting and improves generalization.

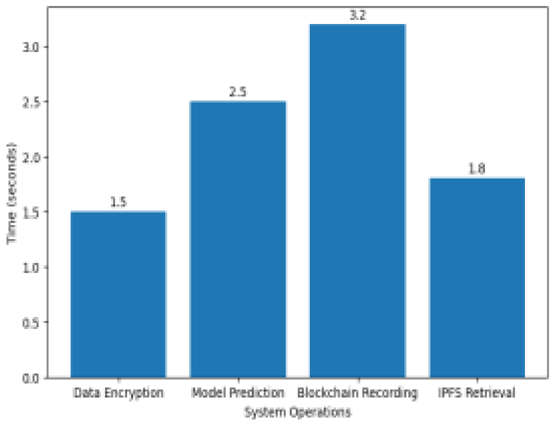


Figure 1: Model Accuracy Comparison of Machine Learning Algorithms

Figure 1 presents the accuracy comparison of different machine learning algorithms used for Chronic Kidney Disease (CKD) prediction. The X-axis represents the algorithms, including Decision Tree, Random Forest, Logistic Regression, and K-Nearest Neighbors (KNN), while the Y-axis represents the accuracy percentage. Among the evaluated models, the Random Forest algorithm achieved the highest accuracy of **96.8%**, outperforming the other models. Logistic Regression and KNN achieved accuracies of **93.6%** and **94.1%**, respectively, while Decision Tree obtained **92.4%** accuracy.

This comparison indicates that Random Forest provides better prediction performance for CKD detection.

Table 2: System Performance Metrics

Parameter	Value
Average Prediction Time	2.5 seconds
Blockchain Transaction Time	3.2 seconds
IPFS Retrieval Time	1.8 seconds
System Uptime	99.9%

Table 2 presents the performance evaluation of the proposed CKD prediction system in terms of speed, reliability, and system efficiency. The average prediction time of 2.5 seconds indicates that the machine learning models can quickly analyze patient data and generate results, which is important for real-time healthcare applications. The blockchain transaction time of 3.2 seconds represents the time required to securely record prediction results and data identifiers on the blockchain ledger, ensuring transparency and immutability. The IPFS retrieval time of 1.8 seconds shows the efficiency of decentralized data storage and retrieval. Additionally, the system uptime of 99.9% demonstrates high reliability and continuous availability of the healthcare prediction system.

Figure 2: System Response Time Analysis

Figure 2 illustrates the time required for different operations performed in the proposed CKD prediction system. The graph compares the response time of key processes including data encryption, machine learning prediction, blockchain transaction recording, and IPFS data retrieval. The results show that data encryption takes approximately 1.5 seconds, ensuring secureprotection of patient information before processing. The machine learning prediction process requires around 2.5 seconds, indicating efficient analysis of medical data for CKD detection. Recording the prediction results on the blockchain takes about 3.2 seconds, which ensures transparency and immutability of the transaction. Finally, IPFS retrieval requires around 1.8 seconds, demonstrating fast access to decentralized stored data. Overall, the graph indicates that the system maintains efficient processing speed while ensuring strong security and decentralized data management, making it suitable for real-time healthcare applications.

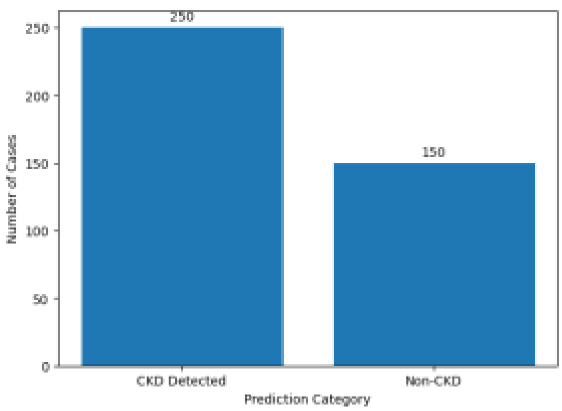
Table 3: Security Evaluation Results

Security Feature	Status

AES Data Encryption	Enabled
RSA Key Exchange	Secure
Blockchain Immutability	Verified
IPFS Data Integrity	Confirmed

Table 3 presents the security evaluation of the proposed CKD prediction system, highlighting the mechanisms used to protect sensitive patient data. The system uses AES encryption to secure medical records before storage or transmission, ensuring that unauthorized users cannot access the information. RSA encryption is applied for secure key exchange, providing an additional layer of protection during communication. The blockchain immutability feature ensures that once a transaction or prediction record is stored, it cannot be modified or tampered with, maintaining data integrity. Additionally, IPFS data integrity guarantees secure decentralized storage, allowing encrypted medical files to be safely stored and retrieved without relying on centralized servers.

Figure 3: prediction category



The CKD Prediction Distribution graph represents the number of cases predicted by the model as either CKD detected or Non-CKD. In this graph, the X-axis represents the prediction categories, while the Y-axis represents the number of cases in the dataset. The results show that 250 cases were predicted as CKD and 150 cases were predicted as Non-CKD. This distribution indicates that the model is capable of identifying patients who are at risk of Chronic Kidney Disease effectively. The graph helps visualize how the prediction model classifies patient records. It also demonstrates that the system can distinguish between diseased and healthy cases with reliable performance.

CONCLUSION

This research proposes a privacy-preserving and decentralized framework for Chronic Kidney Disease (CKD) prediction by integrating machine learning algorithms, blockchain technology, and the InterPlanetary File System (IPFS). The system addresses key challenges in traditional healthcare systems, such as data privacy risks, centralized storage, and limited transparency. By combining machine learning with decentralized technologies, the framework enables secure and reliable early detection of CKD.

Several machine learning algorithms, including Decision Tree, Random Forest, Logistic Regression, and K-Nearest Neighbors, were used to analyze patient medical data and predict CKD risk. Among them, ensemble-based models showed better accuracy and prediction performance. To protect sensitive medical information, AES and RSA encryption techniques are used during data storage and transmission.

Blockchain technology ensures transparency and immutability of prediction records, while IPFS provides decentralized storage for medical data. Experimental results demonstrate high prediction accuracy, improved security, and efficient performance, enabling secure collaboration among healthcare providers and supporting early diagnosis and better clinical decision-making for CKD patients.

REFERENCES

[1] J. M. López-Novoa, C. Martínez-Salgado, A. B. Rodríguez-Peña, and F. J. L. Hernández, "Common pathophysiological mechanisms of chronic kidney disease: Therapeutic perspectives," *Pharmacological Therapeutics*, vol. 128, no. 1, pp. 61–81, 2010.

[2] M. Cortinovis, N. Perico, P. Ruggenenti, A. Remuzzi, and G. Remuzzi, "Glomerular hyperfiltration," *Nature Reviews Nephrology*, vol. 18, no. 7, pp. 435–451, 2022.

[3] K. Kalantar-Zadeh, T. H. Jafar, D. Nitsch, B. L. Neuen, and V. Perkovic, "Chronic kidney

disease," *The Lancet*, vol. 398, no. 10302, pp. 786–802, 2021.
[4] D. Martinusen, J. G. Marin, E. Cheng, and W. Lau, "Chronic kidney disease and end-stage renal disease," in *Renal Medicine and Clinical Pharmacy*, Springer, 2020.
[5] C. P. Kovesdy, "Epidemiology of chronic kidney disease: An update," *Kidney International Supplements*, vol. 12, no. 1, pp. 7–11, 2022.
[6] E. F. Carney, "The impact of chronic kidney disease on global health," *Nature Reviews Nephrology*, vol. 16, no. 5, p. 251, 2020.
[7] A. Francis et al., "Chronic kidney disease and the global public health agenda: An international consensus," *Nature Reviews Nephrology*, vol. 20, no. 7, pp. 473–485, 2024.
[8] P. Rossing et al., "The impact of obesity on cardiovascular and kidney outcomes in patients with chronic kidney disease," *Diabetes, Obesity and Metabolism*, vol. 25, no. 10, pp. 2989–2998,
2023.
[9] R. A. Wahab, R. V. Cohen, and C. W. le Roux, "Recent advances in the treatment of patients with obesity and chronic kidney disease," *Annals of Medicine*, vol. 55, no. 1, 2023.
[10] D. Swain, U. Mehta, A. Bhatt, H. Patel, K. Patel, and D. Mehta, "A robust chronic kidney disease classifier using machine learning," *Electronics*, vol. 12, no. 1, p. 212, 2023.
[11] H. Iftikhar et al., "A comparative analysis of machine learning models for predicting chronic kidney disease," *Sustainability*, vol. 15, no. 3, p. 2754, 2023.
[12] F. Sanmarchi et al., "Predict, diagnose, and treat chronic kidney disease with machine learning: A systematic review," *Journal of Nephrology*, vol. 36, no. 4, pp. 1101–1117, 2023.
[13] Todupunuri, A. (2025). The Role of Human-Centric AI in Building Trust in Digital Banking Ecosystems. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5120605
[14] Babburi, S. Privacy-Preserving Collaborative Framework with Auditable Federated Learning.
[15] Gaddam, S. Integrating Analytics into the Development Process: Bridging the Gap between Data Insights and Design Execution.
[16] Bajarang Bhagwat, V. (2023). Optimizing Payroll to General Ledger Reconciliation: Identifying Discrepancies and Enhancing Financial Accuracy. JOURNAL OF ADVANCE AND FUTURE RESEARCH,1(4). https://doi.org/10.56975/jaafr.v1i4.501636
[17] S. M. K. P. (2025). Cryptography in iOS: A Study of Secure Data Storage and Communication Techniques. International Journal on Science and Technology,16(1). https://doi.org/10.71097/ijsat.v16.i1.1403
[18] Doragacharla, V. R. (2026). AI-Enabled Commerce Platforms in Cloud Computing Environments: An Architectural and Socio-Economic Analysis. Journal of Computational Analysis & Applications, 35(1).
[19] Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.
[20] Poojari, R. Frameworks for Data Management and Lineage in Large-Scale Healthcare Data Systems.
[21] Uday Kumar Kalae. (2025). AN AUTOMATED SYSTEM FOR MANAGING HIGH-AVAILABILITY CLOUD INFRASTRUCTURE THROUGH INFRASTRUCTURE-ASCODE (IAC) PRACTICES. American Journal of AI Cyber Computing Management, 5(2), 42–50. https://doi.org/10.64751/ajaccm.2025.v5.n2.pp42-50
[22] Kalae, U. K. (2023). Enhancing deployment efficiency through CI/CD pipelines and containerization with Docker and Kubernetes. International Journal of Communication Networks and Information Security, 15(4), 728–736.
[23] Banda Saikumar. (2025). Integrating azure network rules for storage account through terraform in CI/CD pipelines: automating storage account access restrictions to public IP. Journal of Scien+B112ce & Technology, 10(2), 15–22. https://doi.org/10.46243/jst.2025.v10.i02.pp15-22
[24] Vasagam, M., Kumar, A., & Garg, A. (2026). Learning Execution Plan Embeddings for Multi-Dimensional Query Resource Prediction. IEEE Access.
[25] Patel, S., & Patyrykin, K. (2025). Strategic Impacts of Salesforce Automation on

Organisational Competitive Advantage in Emerging Markets. Journal of Posthumanism, 5(12), 357–372. https://doi.org/10.63332/joph.v5i12.3782

[26] Patyrykin, K. (2025). CANCEL CULTURE PROBLEM. Lex Localis: Journal of Local Self-Government, 23.