International Journal of Computer Applications (0975 - 8887)
Volume 187 - No.100, April 2026

A Blockchain-based Patient-Centric Electronic Health Record System with Secure IPFS Storage and Machine Learning-Driven Health Analytics

Aryan Deepak Saraf
Department of Information Technology
Vivekanand Education Society's Institute of Technology
Mumbai, India

Ganesh Sanjay Pandhre
Department of Information Technology
Vivekanand Education Society's Institute of Technology
Mumbai, India

Vivek Sachchelal Gupta
Department of Information Technology
Vivekanand Education Society's Institute of Technology
Mumbai, India

Yash Shivdas Naikwadi
Department of Information Technology
Vivekanand Education Society's Institute of Technology
Mumbai, India

Anushree Prabhu
Department of Information Technology
Vivekanand Education Society's Institute of Technology
Mumbai, India

ABSTRACT

The electronic health record (EHR) has been an indispensable tool for healthcare organizations in recent years. However, problems remain associated with EHRs, including fragmented data storage, security vulnerabilities, and limited patient control over personal health information. In this paper, MedLink, a blockchain-powered health record management system, is proposed to provide safe and efficient management of patients' medical data. The system adopts blockchain architecture to grant access permissions via smart contracts, with encrypted health records being stored externally by using the InterPlanetary File System (IPFS).
In addition to security and access permission mechanisms, MedLink incorporates customized machine learning (ML) algorithms to analyze medical data and conduct early diagnosis for diseases such as diabetes, heart conditions, and cancers. Different from the purely general-purpose artificial intelligence generation (AI) model, the proposed method concentrates on creating a customized ML model specifically for particular tasks using existing medical data sources. Experiments have been conducted to evaluate the prototype, revealing promising results regarding data integrity, traceability, and analytical capabilities.

General Terms

Blockchain, Healthcare Security, Decentralized Systems, Machine Learning

Keywords

Blockchain, Electronic Health Records, Decentralized Storage, IPFS, Smart Contracts, Machine Learning, Healthcare Data Security, Patient-Centric Systems

1. INTRODUCTION

The healthcare sector has experienced an accelerated development and implementation of digital systems for storing patient information. Nevertheless, the architecture itself remains to contain some issues. For instance, in most cases, patient health records are scattered around different hospitals and clinics, making the information fragmented. Such a situation leads to ineffective treatment due to delays in starting it, repetition of some tests, and an increased risk of making mistakes during a medical procedure [2]. Also, currently used EHR systems use centralized databases, which become easy targets for various cyberattacks. Consequently, compromising of the database means compromising of personal and health information of many patients [5].

The suggested system, MedLink, provides a new patient-oriented approach to storing and managing access to medical records. Blockchain technology is used as an immutable and transparent access controller. All medical data, on the other hand, is stored on the IPFS using encryption to ensure maximum security of information. Moreover, MedLink is not only about providing a reliable and secure platform for storing data but it can provide valuable information regarding patients' health based on machine learning techniques. Unlike other health data management systems, the sug-

gested solution uses specific machine learning models rather than generative AI for analyzing health data, in order to diagnose illnesses such as diabetes, cancer, and others [13].

2. PROBLEM STATEMENT

Despite the extensive use of EHR technology, there are several factors that limit its potential within the current healthcare system. First, the interoperability between EHR systems of different healthcare providers remains a challenge due to the lack of standardized practices and guidelines. Therefore, medical information collected from patients by one hospital or clinic is inaccessible and cannot be applied by another organization [3].

Security is also one of the main concerns associated with the operation of EHR solutions. They store medical data in centralized databases that are open to unauthorized access, data theft, and alterations. Moreover, they often have insufficient audit trails that fail to track data modification and usage processes [10].

In terms of analysis, existing EHR systems operate primarily as data storage platforms without utilizing the data to analyze and predict potential diseases in patients. This prevents them from using their potential to detect diseases early and make evidence-based clinical decisions. Besides, some modern AI-based healthcare applications depend on LLMs or generic generative AI models to make diagnoses and provide consultations to physicians. Although such models facilitate interactions between healthcare professionals and patients, they might not be highly accurate due to their limited understanding of medical knowledge [9].

3. LITERATURE SURVEY

There has been an increased interest in combining blockchain and AI as means to overcome security and privacy issues associated with healthcare data systems. Various studies investigate possibilities for integrating the two technologies and improving management of Electronic Health Records (EHRs). In particular, Sun et al. (2024) developed a federated learning framework, where the learning of AI models is performed using EHR data from several sites without compromising patient privacy [1]. In other words, one can derive insights from medical data without storing it at a single site [1].

Similarly, Chamola et al. (2022) analyzed the possibility of using blockchain for improving the security and integrity of health record databases through the help of AI-based analytics [3]. Blockchain can contribute to maintaining the integrity of healthcare records.

Moreover, systematic literature reviews by Tariq et al. (2025) and Chandak and Chandak (2026) revealed the importance of using specific machine learning tools for the purposes of analyzing healthcare data [5, 7]. As stated by Khalid et al. (2023), privacy-preserving AI algorithms play a crucial role in gaining the trust of patients, which is vital for the development of diagnostic systems [9].

Thus, considering the above evidence and the findings from the reviewed literature, MedLink is supposed to use the blockchain network for managing patient access control and a separate machine learning engine that will use real-world clinical datasets.

4. PROPOSED SYSTEM ARCHITECTURE

MedLink is structured into four primary layers to ensure modularity, security, and intelligence.

4.1 User Layer

The User Layer acts as the key interface through which patients and healthcare professionals interact with the system. It is built using React.js for developing components in the user interface and Tailwind CSS for maintaining consistent style across various screens. The user interface provides cryptographic authentication of users, meaning that no conventional password-username mechanism is required for logging in.

Using the simple user dashboard, patients are able to add, manage, and view their medical records, whereas healthcare professionals are capable of requesting medical information from these records after obtaining permission from the patients. The interface presents health insights offered by the system in a simple manner.

4.2 Application and Intelligence Layer

The Application and Intelligence Layer involves analyzing medical data and drawing meaningful conclusions from the patient record. First, when a user uploads a medical report either in PDF format or as an image file, the system uses an OCR and NLP pipeline to extract useful textual and numerical information from the medical document [15]. In essence, the medical report is converted into a structured data set using machine learning techniques.

Secondly, the extracted data is forwarded to a custom-designed machine learning engine, which has different classifiers that can analyze the structured data set. Classifiers used include Random Forest and SVM. In the development of these machine learning algorithms, they are fed with real-world medical datasets, and they can predict diseases like breast cancer, diabetes, and kidney disease. The use of ML algorithms ensures that the system performs more reliably compared to generic AI models.

Moreover, the Application and Intelligence Layer has a Generative AI assistant that offers a conversational interface to the user. The assistant enables patients to gain insight into the medical document and the results obtained from the ML algorithms in a simplified manner. However, the assistant acts as an explanation tool and does not make any medical recommendations to patients.

4.3 Security and Blockchain Layer

The Security and Blockchain Layer represents the key component of the access control and audit process at MedLink. It is based on Ethereum-based smart contracts that have been created and thoroughly checked within the Hardhat ecosystem. Unlike some other solutions, MedLink does not store medical information itself but uses blockchain technology for managing metadata, permissions, and access logs securely and immutably [10].

For the purpose of securely storing metadata of each medical record, the MedVault smart contract has been introduced. It contains crucial information about records, such as their IPFS Content Identifiers (CIDs), and allows hashing each document in order to detect its manipulation. The Guardian Contract, which provides patients with emergency access to their medical data, has been added in case something goes wrong. It operates in the following way: when a request appears, multiple parties must give their signatures for emergency access to be granted.

Finally, MedLink uses the blockchain layer to introduce fine-grained permission controls for medical data access. Patients can create limited access permissions to certain health organizations,

and all transactions associated with it will be stored on the blockchain.

4.4 Storage Layer

The Storage Layer ensures the safe and decentralized storage of medical records. In MedLink, the InterPlanetary File System (IPFS) protocol is used to store patient information in a decentralized way without storing them in central nodes that serve as single points of failure [6]. To guarantee the privacy of this information, all medical documents are encrypted using the AES-256 encryption algorithm locally on the user's device before sending them to IPFS.

Therefore, in case any storage node gets hacked, the data on it will be inaccessible to third parties because of the strong encryption algorithm applied. Only users with the proper credentials and encryption keys can access this information and make sense of it.

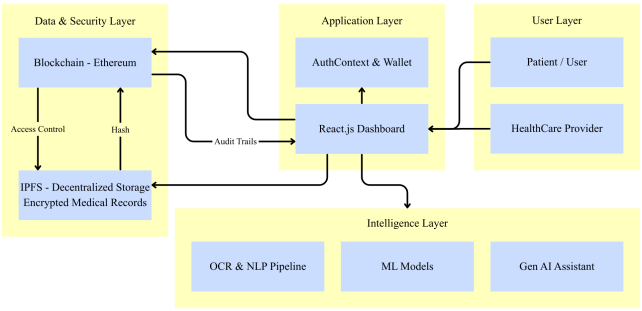

Fig. 1. System Architecture

5. IMPLEMENTATION DETAILS

5.1 Frontend Integration

For the front end of MedLink, a modern stack including React.js has been chosen to create a responsive application interface. The application's state management and session handling are performed by a custom AuthContext where login statuses and roles for both patients and healthcare professionals are tracked. Blockchain interactions are performed via BlockChainContext where the front end calls functions such as record registration, requesting permissions, and so on.

The front end has been made intuitive and easy to use for people without any special knowledge, since its goal is to help view health information easily and quickly. The responsiveness of the application allows both patients and doctors to access the records from mobile devices as well.

5.2 Machine Learning Engine

In terms of AI integration, MedLink focuses on the development of task-specific ML-based algorithms rather than the development of a generalized artificial intelligence system. This strategy helps the system deliver more meaningful results to users since ML algorithms are created specifically for each medical prediction task. Each model is trained on public, real-world medical data and serves as a tool for preliminary patient analysis, but does not make diagnostic decisions.

As for diabetes prediction, the system uses a Logistic Regression model trained on the PIMA Indians Diabetes Database containing common clinical indicators [16]. The analysis of heart diseases is conducted through a Random Forest algorithm that can account for multiple risk factors and complex relations between different variables in clinical data [17]. Besides table-based data analysis, MedLink also features a pipeline for preliminary image analysis that is based on the Convolutional Neural Network (CNN) [22] and MobileNetV2 for specific tasks like brain tumor and breast cancer detection [20, 21].

The output provided by this pipeline is represented as decision-making indicators, rather than final diagnoses.

5.3 Blockchain Workflow

In the blockchain flow of MedLink, registration, storage, and management of medical records have been elaborated on in a very effective way. After registering as a user, a Health ID is generated for the user upon interacting with a smart contract on the blockchain. This Health ID remains with the user throughout their stay on the platform.

Once a patient decides to upload a medical record, the record is first encrypted locally to protect its confidentiality before being uploaded to the InterPlanetary File System (IPFS). The encrypted medical record is stored in IPFS and a CID together with the SHA-256 hash of the encrypted record is anchored on the blockchain [12]. Any changes made to the record at any point in time will be detected very easily because of hashing and storing the hash on the blockchain.

Medical records are accessed via smart contracts. Once the healthcare provider requests access to the medical record, the patient has to approve the access through the blockchain. This access is time-bound in a way that gives the doctor time-limited access to view the medical record after decryption.

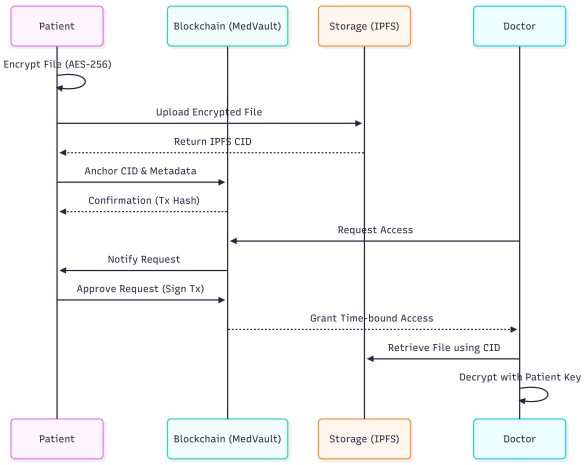

Fig. 2. Blockchain Workflow

6. RESULTS AND DISCUSSION

The MedLink prototype proves the viability of using blockchain technology and artificial intelligence to make EHR management

efficient, smart, and secure. The use of encryption and storage of medical records on the IPFS platform together with anchoring of the metadata of those records in the blockchain makes this approach highly viable from both security and decentralization perspectives.

6.1 Machine Learning Model Performance

The proposed ML integration solutions were exhaustively tested using publicly accessible medical datasets, producing encouraging results for preliminary disease screening. Task-based models are more effective and precise for processing structured datasets than large language models, as evidenced by the presented results. The most important thing to remember is that all of these predictions are meant to facilitate early detection and not be considered any form of diagnosis.

As for diabetes detection, a Logistic Regression model trained on the PIMA Indians Diabetes Dataset achieves an accuracy rate of approximately 76% [16]. The detection of heart disease, carried out with the help of a Random Forest classifier, produced accuracy in the range of 72% [17]. In addition, the image analysis component, which relies on the CNN model, achieved an accuracy of approximately 93% for pneumonia detection from chest X-ray pictures [22].

Table 1. Disease Prediction Models and Accuracy

Disease	**Model Used**	**Approx. Accuracy**
Diabetes	Logistic Regression	76%
Blood Pressure	Random Forest	92%
Liver Disease	Random Forest	73%
Chronic Kidney Disease	Logistic Regression	100%
Image Based:		
Pneumonia (Chest X-ray)	CNN	93%
Brain Tumour (MRI)	MobileNetV2	96%
Breast Cancer	MobileNetV2	90%
Diabetic Retinopathy	CNN	72%
Skin Cancer	MobileNetV2	92%

6.2 Blockchain and Security Performance

Regarding security and compliance, all access requests, approvals, and references to the records are immutable and logged in the blockchain, creating an auditable trail [10]. The characteristics fit well with HIPAA and GDPR requirements, including accountability, traceability, and secure data access. Ethereum-based smart contracts, developed using the Hardhat framework, ensure access permissions' integrity and immutability.

IPFS off-chain storage of AES-256 encrypted medical records ensures privacy, and even if one node in the IPFS network is hijacked, it will not be able to decrypt the information since only the patient has the private key [6]. Our design minimizes costs by keeping only the metadata in the blockchain, resulting in low transaction costs and increased throughput [12]. Although the retrieval time from IPFS is slower compared to centralized databases, it is optimized through efficient content addressing and local caching techniques.

6.3 User Interface and Experience

The frontend of the application, developed using React.js and Tailwind CSS, creates an interactive and user-friendly interface. The patient dashboard gives a visual representation of the health intelligence with simple navigation.

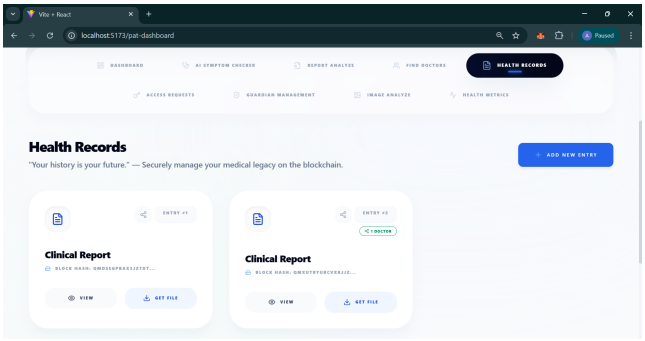

Fig. 3. Medical Patient Dashboard

The use of cryptographic wallet authentication makes the access procedure very easy, removing the traditional weaknesses of using passwords for secure access. The Generative AI assistant embedded into the system serves as a tool for explaining technical language to patients about their medical data and results of analysis, thus improving understanding without taking away responsibility from healthcare workers.

In order to better illustrate the structure and workflow of the system, two diagrams are provided. The first shows the architecture of the four main layers of the MedLink system along with their interactions. The second shows a step-by-step illustration of a transparent blockchain workflow starting from registration and uploading of encrypted files to IPFS, then proceeding with anchoring the metadata onto the blockchain, and eventually making an access request and getting the access permission from the healthcare provider.

7. LIMITATIONS

Despite its benefits, there are several drawbacks to the suggested design that need to be addressed. In particular, the machine learning algorithms utilized in the prototype were tested using open-source data, which can limit their effectiveness when applied to patient records within actual hospitals.

In addition, the system was designed for a proof-of-concept type of deployment and cannot take into account factors such as network latency and transaction cost when implemented at scale. Specifically, using blockchain-based access control can delay medical procedures in case of an emergency if the system is not properly configured. The data stored via IPFS remains dependent on the presence of a sufficient number of nodes supporting the process.

Legally, MedLink has not yet been implemented in a live healthcare environment, and thus, its full alignment with healthcare standards would require additional validation and testing. Finally, although machine learning algorithms are expected to assist with decision-making in the suggested framework, they are not intended to replace professional diagnoses and treatments.

8. CONCLUSION AND FUTURE WORK

The presented paper suggests the MedLink framework, an EHR-based system built around blockchain technology and leveraging task-oriented machine learning algorithms specifically designed for healthcare-related applications. Through the use of smart contracts providing access control combined with data storage via the encrypted off-chain IPFS platform, the suggested approach addresses many issues like data security, consistency, and user ownership. As opposed to solutions relying on generic AI algorithms, MedLink focuses on task-specific machine learning trained on medical datasets to provide trustworthy preliminary results of health analysis.

Although the presented prototype provides evidence for the possibility of a robust patient-focused EHR solution, there are several ways how its performance can be enhanced. Specifically, a future version of the system might involve the application of federated learning approaches allowing to train models collectively based on datasets of various medical facilities without disclosing patients' sensitive information [1]. Another way to increase the effectiveness and value of the system involves integration with IoT-based wearables for real-time continuous monitoring of individuals' health status and prediction of risks.

9. REFERENCES

[1] A. Althaf Ali et al., "Securing electronic health records using blockchain-enabled federated learning for IoT-based smart healthcare," *Clinical eHealth*, vol. 8, pp. 125–133, 2025, doi: 10.1016/j.ceh.2025.04.002.

[2] E. R. D. Villarreal et al., "Blockchain for healthcare management systems: A survey on interoperability and security," *IEEE Access*, vol. 11, pp. 5629–5652, 2023, doi: 10.1109/ACCESS.2023.3236505.

[3] V. Chamola et al., "Artificial intelligence-assisted blockchain-based framework for smart and secure EMR management," *Neural Computing and Applications*, vol. 35, pp. 22959–22969, 2023, doi: 10.1007/s00521-022-07087-7.

[4] S. Sabu et al., "Implementation of a secure and privacy-aware E-Health record and IoT data sharing using blockchain," *Global Transitions Proceedings*, vol. 2, no. 2, pp. 429–433, 2021, doi: 10.1016/j.gltp.2021.08.033.

[5] U. U. Tariq et al., "Blockchain-based secured data sharing in healthcare: A systematic literature review," *IEEE Access*, 2025, doi: 10.1109/ACCESS.2025.3547953.

[6] J. Sun et al., "Blockchain-based secure storage and access scheme for electronic medical records in IPFS," *IEEE Access*, vol. 8, pp. 59389–59401, 2020, doi: 10.1109/ACCESS.2020.2982964.

[7] A. Chandak et al., "Blockchain applications in electronic health records: A systematic review," *BMC Medical Informatics and Decision Making*, 2026, doi: 10.1186/s12911-026-03476-3.

[8] Z. Sun et al., "A blockchain-based secure storage scheme for medical information," *Journal of Wireless Communications and Networking*, 2022, doi: 10.1186/s13638-022-02122-6.

[9] N. Khalid et al., "Privacy-preserving artificial intelligence in healthcare: Techniques and applications," *Computers in Biology and Medicine*, vol. 158, 2023, doi: 10.1016/j.compbiomed.2023.106848.

[10] G. Wu et al., "Privacy-preserved electronic medical record exchanging and sharing," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 5, pp. 1917–1927, 2022, doi: 10.1109/JBHI.2021.3123643.

[11] L. Sun et al., "A blockchain-based e-healthcare system with provenance awareness," *IEEE Access*, vol. 12, pp. 110098–110112, 2024, doi: 10.1109/ACCESS.2024.3440170.

[12] J. Kaur et al., "Attribute-based access control scheme for secure storage and sharing of EHRs," *Cluster Computing*, vol. 27, pp. 1047–1061, 2024, doi: 10.1007/s10586-023-04038-2.

[13] V. Jackins et al., "AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes," *Journal of Supercomputing*, vol. 77, pp. 5198–5219, 2021, doi: 10.1007/s11227-020-03481-x.

[14] S. Jain et al., "Enhanced blockchain and IPFS based secure storage and sharing of electronic healthcare records," *Cluster Computing*, vol. 28, 2025, doi: 10.1007/s10586-025-05564-x.

[15] E. Hsu et al., "Deep learning-based NLP data pipeline for EHR-scanned document information extraction," *JAMIA Open*, vol. 5, no. 2, 2022, doi: 10.1093/jamiaopen/ooac045.

[16] H. Naz and S. Ahuja, "Deep learning approach for diabetes prediction using PIMA Indian dataset," *Journal of Diabetes & Metabolic Disorders*, vol. 19, no. 1, pp. 391–403, 2020, doi: 10.1007/s40200-020-00520-5.

[17] K. Sumwiza et al., "Enhanced cardiovascular disease prediction model using random forest algorithm," *Informatics in Medicine Unlocked*, vol. 41, 2023, doi: 10.1016/j.imu.2023.101316.

[18] D. A. Debal and T. M. Sitote, "Chronic kidney disease prediction using machine learning techniques," *Journal of Big Data*, vol. 9, 2022, doi: 10.1186/s40537-022-00657-5.

[19] K. Gupta et al., "Liver disease prediction using machine learning classification techniques," in *Proc. IEEE CSNT*, 2022, pp. 221–226, doi: 10.1109/CSNT54456.2022.9787574.

[20] S. Sumaiyya et al., "Breast cancer detection with MobileNetV2 and explainable AI," in *Proc. ICSES*, 2024, pp. 1237–1243, doi: 10.1109/ICSES63445.2024.10763178.

[21] R. Saini et al., "Brain tumor classification using VGG-16 and MobileNetV2 deep learning techniques on MRI," in *Communications in Computer and Information Science*, vol. 1788, 2023, doi: 10.1007/978-3-031-27609-5_24.

[22] A.-O. Asia et al., "Detection of diabetic retinopathy using CNN classification models," *Electronics*, vol. 11, no. 17, 2022, doi: 10.3390/electronics11172740.