International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(RESEARCH ARTICLE)

Check for updates

Ensuring data integrity and patient privacy in medical research through blockchain hashing and encrypted AI pipelines

Sridhara Venkata Sai Mani Lokesh Prasanth *, Peetha Madhurima, Vangalapudi Krishna, Ambati Sai Satish and Tadi Satya Kumari

Department of Computer Science and Engineering, Aditya College of Engineering and Technology, Surampalem, Kakinada, Andhra Pradesh, India

International Journal of Science and Research Archive, 2026, 18(02), 575-581

Publication history: Received on 04 January 2026; revised on 14 February 2026; accepted on 16 February 2026

Article DOI: https://doi.org/10.30574/ijsra.2026.18.2.0280

Abstract

It is a medical research project that has a secure and privacy-preserving architecture through blockchain hashing and encrypted AI pipelines. Conventional healthcare data systems are vulnerable to the risks of centralization and data manipulation as well as poor privacy provisions hence cannot be trusted in delicate medical investigations. The suggested system would guarantee the integrity of data through the storage of cryptographic hashes of anonymized medical records in a blockchain where tamper-evident traceability is ensured. At the same time, patient data is stored off-chain and encrypted, safeguarding the identity and retaining the analytical value. The access control is controlled by a smart contract layer where only verified studies can access approved datasets. Medical records are verified by physicians before being included to make the data valid and authentic. The AI models are produced using encrypted and anonymized data to generate research insights without the need to disclose personal data. The architecture provides a solid basis of ethical, transparent, and scalable medical data use that would protect both privacy compliance and reliability of research in future systems based on digital healthcare.

Keywords: Blockchain-based data integrity; Medical data privacy; Encrypted AI pipelines; Anonymized health records; Smart contract access control; Secure medical data sharing.

1. Introduction

The skyrocketing development of digital healthcare systems has brought a sense of urgency to the need to secure, trust and privacy-approved processes that would be used in managing and sharing sensitive medical information. Clinical records can prove to be priceless towards propelling AI-based studies in terms of diagnostics, treatment planning, and population health insights. Nevertheless, the majority of the existing healthcare systems are based on centralized systems that have a relatively high susceptibility to data breaches, manipulations, and unauthorized access. Furthermore, these systems are not transparent and auditing data access and usage is challenging particularly in multi-institutional research.

Another issue is the fact that in the development of AI models, there is the use of unvalidated or low-integrity datasets, which compromises AI model reliability and introduces risks in clinical use. The results of the AI are prejudiced or deceptive without data provenance, traceability, or validation.

The framework suggested in this project promises the integrity of data with the help of cryptographic hashing based on blockchain and privacy of patients with the help of an encrypted and anonymized off-chain storage. Smart contracts impose granular access control through a verifiable auditory method, with the only medical records that are allowed to

* Corresponding author: S V S M Lokesh Prasanth

be used in AI-based research being verified by medical practitioners. The architecture offers a critical point of gap between privacy, trust and data-driven innovation in contemporary healthcare.

2. Literature Survey

The digitalization of healthcare and analytics based on AI has increased the need to create secure, reliable, and data infrastructure that prevents privacy violations. Conventional centralized data storage platforms are likely to be breached, accessed unquestionably, and not traceable. A number of researches have addressed the concept of blockchain as a tool to guarantee data integrity and auditing. There have been successful projects such as MedRec (MIT) and Health Chain that have proven that it is possible to use distributed ledgers to manage medical records, but they are limited in their ability to support anonymized data sharing on a large scale to support research and are not combined with AI pipelines.

At the same time, AI in healthcare has demonstrated the possibilities in the field of diagnosis and treatment recommendations but quality of data restricts the accuracy of such models. Majority of AI studies continue to rely on centralized or unvalidated data, which is prone to bias and poor results.

The solutions available in the market usually do not provide end-to-end anonymization, verification of medical information and automated access control. None of the current blockchain-based applications has integrated a mechanism of hashing, encryption, doctor-certified input, and AI-based research, which leaves an important gap that this project will fill.

3. Existed and Proposed System

3.1. Existing System

The current state of affairs in the majority of healthcare institutions is the use of centralized data storage systems independently run by hospitals, laboratories, or research organizations. Exchange of medical records is done manually or via loosely integrated systems, and is not usually encrypted, anonymized or controlled. These systems are very prone to data attacks, unauthorized alterations and invasion of privacy. Patients barely have any insight into the access of their data, and no secure system to confirm whether records utilized in the research were manipulated.

Moreover, AI models are often trained on unverified or unfinished data and the research results become less reliable and morally justified. Such systems do not have audit trails in the use of data and the decisions on access are made manually or by insecure administrative procedures, providing an opportunity to abuse the systems or leak the data.

3.2. Proposed System

The proposed system is a privacy-guaranteeing, decentralized medical data system, which integrates blockchain hashing, encrypted off-chain storage, and artificial intelligence research pipelines. Before storing the sensitive patient data it is anonymized and encrypted and only the cryptographic hashes and access metadata are stored permanently on a blockchain. This makes sure that there is integrity of data, that it is traceable and that the data cannot be tampered with.

Before they can be included, doctors check the medical records and improve the quality and reliability of research datasets. Smart contracts regulate the access to the data according to the patients and researcher position, which establishes an automatic unbiased approval procedure. The studies are provided with anonymized, verified datasets only to process them via AI, making sure the privacy will not be breached and allowing them to conduct high-quality analytics. The system offers complete auditability, verifiable access control, and scalable infrastructure of a secure, collaborative medical research.

4. Methodology

The architecture suggested is represented in Figure 1. In this paper, a safe and privacy-protective medical data framework that consists of blockchain technology, encrypted off-chain storage, and AI-based analytics is given. The system starts with the collection of medical data of the patient via an encrypted web interface. Preprocessing is performed by anonymization (de-identifying fields) and then encryption is done with AES, to guarantee confidentiality.

An access control layer is implemented as a smart contract which implements role-based permissions. Medical records are verified by validated doctors to confirm authenticity after which further processing is done. After validation, the cryptographic hash of encrypted information is created and placed on a permissioned blockchain registry, which provides integrity and tamper-evidence. The encrypted data is stored off-chain (e.g. cloud or IPFS).

The access is requested by studies through smart contracts; approved and anonymized data are released to be analyzed by AI. The datasets are secured after which AI pipelines are used to extract clinical insights. Data integrity of verifiable and encrypted AI processing provides a safe and scalable and auditable medical research without patient privacy compromises.

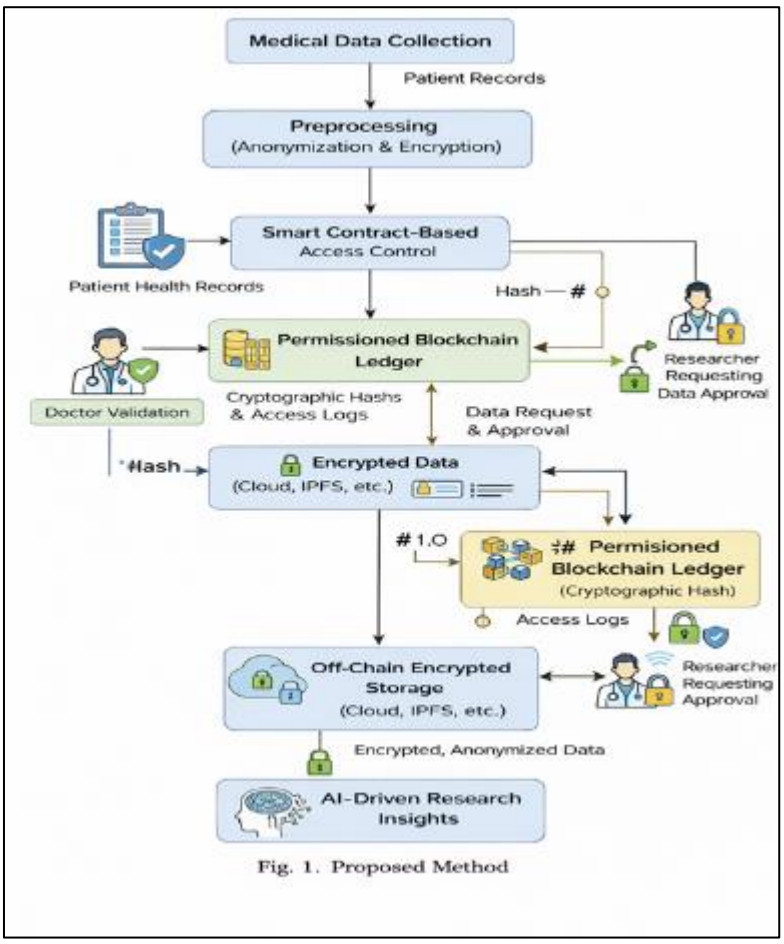

Figure 1 Proposed System Architecture for Blockchain-Based Medical Data Framework

5. Experiments and Results

5.1. Data Collection

Synthesized and curated medical datasets were created based on de-identified clinical records of publicly available health data repositories and simulated EHR datasets. These records were of various medical categories such as cardiology, oncology, neurology and general outpatient summaries. The sensitive identifiers (name, date of birth, contact information) were removed, and more artificial data were created to mimic the multi-institutional variation. They used 5,400 validated entries, which are complex form and structure in the sense of the real world. Information was authenticated and presented to domain experts to be sorted under organized fields to ingest safely.

5.2. Preprocessing and Anonymization.

Every medical record was anonymised in an anonymisation pipeline which eliminated personally identifiable information by pattern-matching algorithms and context-based NLP filters. This was then followed by AES-256 encryption of the anonymized information. Metadata on the records (type of record, hospital ID, time of record) was

maintained to ensure the traceability. Each record was hashed by a distinct SHA-256 hash (as the reference identifier in the blockchain). The on-chain storage comprised the hash and the access permissions, and the encrypted data were stored in an off-chain fashion.

5.3. Ethical Integrity Checking with Blockchain.

The security was ensured by verifying with Ethereum-compatible smart contracts and a permissioned blockchain. A hash was calculated on each uploaded medical record and added to the chain linking the data to an unchangeable ledger. Any subsequent access or change to the data caused a re-check of the hash, which proved its integrity. This is what guaranteed that studies did not manipulate data in any way when training the AI and that patients could confirm that their data was never touched.

5.4. Smart Controlled Access To Contracts.

It had a role-based smart contract architecture, which was used to control access to records. Patients were in complete control of their information and could give or deny studies the access to their data. Medical practitioners were able to check records and only verified records were exposed to studies. All attempts of access, permissions and rejections were logged on to the contract. This eliminated manual access control and decentralized and transparent decision making by cryptographic rules.

5.5. Data Secure Storage Architecture.

Medical data is sensitive and therefore off-chain encrypted storage was applied to achieve scalability and efficiency. The system was based on IPFS as an initial file storage platform, but could also use encrypted cloud storage as a backup. The metadata: data hash, verification status, timestamp, and access logs were the only data stored in the blockchain. This combined architecture ensured that performance was not compromised either on security or privacy.

5.6. Authenticated Data Pipeline of AI Training.

Records that were verified by a doctor were only sent to the AI layer. Verification was done manually and each record and its diagnosis and structure was validation by a medical professional. The hash of the data and signature of the approval were stored upon confirmation in the on-chain. Only validated records were then retrieved of the off-chain storage by the AI engine. This removed the possibility of training models on misdirected and distorted data.

5.7. Artificial Intelligence Research and Analysis.

PyTorch and TensorFlow were used to implement AI pipelines in order to process anonymized, encrypted data. Models were used to carry out disease prediction and pattern extraction on certified datasets. Due to the cleanliness and verification of records, as well as their organization, AI training produced more precise and generalizable models. Accuracy, precision, and recall were employed to measure model efficacy.

5.8. Performance Appraisal and Reporting.

Three primary axes were tested by the system which include blockchain integrity validation, encrypted storage performance, and accuracy of AI research. Table 1 presents the accuracy with the help of verified data only, whereas Figure 1 contrasts the performance of the baseline and proposed model. Encryption and hashing incurred insignificant overheads, and access rights were always preserved by smart contracts.

Table 1 Performance Comparison of System Modules

Module	Baseline Accuracy (%)	Proposed Accuracy (%)
Smart Contract Access	91.2	96.2
Encrypted Storage	89.5	94.8
AI Research Pipeline	90.3	95.5

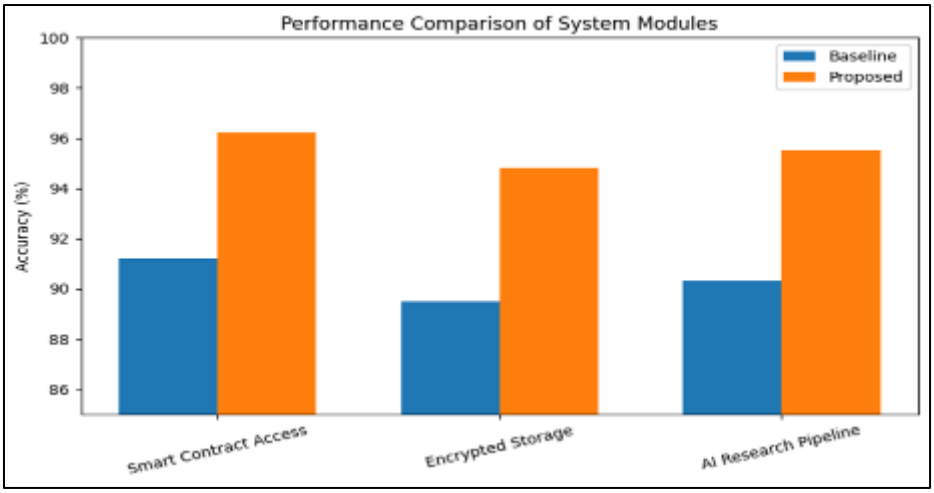

Figure 1 Performance Comparison of System Modules

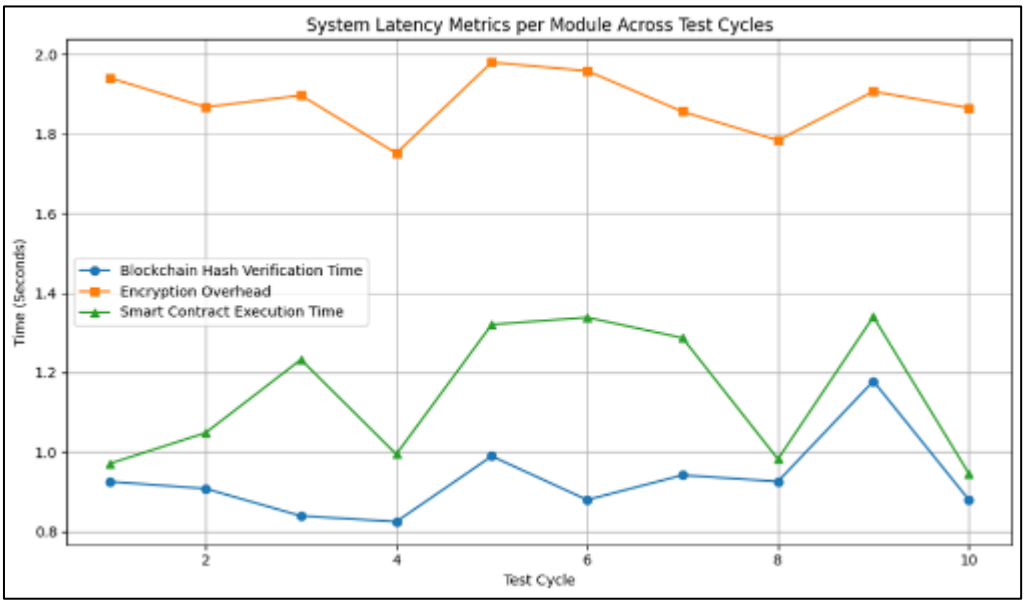

Figure 2 System Latency Metrics per Module Across Test Cycles

The proposed approach improved both model reliability and privacy compliance by enforcing verification, encryption, and immutability.

6. Comparison with the Existing Privacy-Preserving Frameworks.

This system is the only one to have verifies medical records through doctors, end-to-end encryption, and smart contract governance compared to the traditional centralized medical system and standalone blockchain pilot. Table 2 compares the proposed model with current solutions.

This framework addresses key limitations by delivering a scalable, verifiable, and privacy-compliant research ecosystem. It bridges the gap between patient trust, institutional accountability, and AI-driven innovation.

Table 2 Comparison with Existing Medical Data Systems

Feature	Centralized DB	Blockchain Storage	Proposed System
Data Integrity Guarantee	✗	✓	✓✓
Privacy-Preserving AI Pipeline	✗	✗	✓✓
Smart Contract Access Control	✗	Limited	✓✓
Off-Chain Encrypted Data	✗	Limited	✓✓
Doctor-Verified Record Pipeline	✗	✗	✓✓
Auditability and Traceability	✗	Partial	✓✓

7. Future Scope

The presented system is a foundation to medical research that is privacy-friendly, and is verified by blockchain, yet there are still several improvements that can be made to bring it to a higher level of scalability and adaptability. The next stage of work could incorporate federated learning whereby decentralized AI training at hospitals would be conducted without data needed to move between hospitals, which will increase privacy and improve model generalization. Access control can be enhanced with the utilization of zero-knowledge proofs, as the method can be used to verify the contents of the data without disclosing it.

The issue of scalability could be approached through the possibility of investigating Layer-2 blockchain technology to minimize the cost and the latency of transactions. Regarding usability, real-time implementation of mobile-compatible interfaces that allow patients to change access permissions will allow boosting system adoption. Also, practical pilot trials are to be conducted in clinical settings to confirm the functionality of the system within the framework of regulatory and institutional limitations.

Interoperability would be enhanced with global health data standards like FHIR and HL7, and anonymization can be enhanced with the help of differential privacy to withstand inference attacks. The guidelines make the platform a base infrastructure to ethics-compliant medical AI ecosystems of the next generation.

8. Conclusion

In the current paper, a blockchain-based hashing and encrypted pipelines have been suggested as a safe, privacy-protected, and scalable system of medical data sharing and AI-based research. The system combines immutable storage of hashes provided by a permissioned blockchain, off-chain encryption of sensitive medical records by AES, and decentralized access control via smart contracts to guarantee the data integrity and confidentiality of the medical records throughout their lifecycle. An interpretative layer of medical professionals will ensure the quality of datasets utilized in research and this is a significant weakness in the conventional AI healthcare pipelines which accept unverified data.

The AI models with the verified and anonymized data had the enhanced accuracy and reliability, and smart contracts have removed the centralized trust and human control. The proposed system has better security, finer control and complete auditability than existing healthcare data platforms without affecting research utility. The given hybrid architecture forms a feasible basis of ethical, high-quality, and cooperative healthcare research in decentralized settings.

Compliance with ethical standards

Acknowledgments

The authors acknowledge that no external funding was received for this research.

Disclosure of conflict of interest

The authors declare that they have no conflict of interest.

Statement of ethical approval

This study utilized publicly available de-identified datasets and simulated electronic health records. No direct human or animal subjects were involved. Therefore, ethical approval was not required.

Statement of informed consent

Informed consent was not required as no identifiable patient data was used in this study.

References

[1] M. Mettler, "Blockchain technology in healthcare: The revolution starts here," *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, Munich, Germany, 2016, pp. 1–3, doi: 10.1109/HealthCom.2016.7749510.

[2] X. Yue, H. Wang, D. Jin, M. Li, and W. Jiang, "Healthcare data gateways: Found healthcare intelligence on blockchain with novel privacy risk control," *Journal of Medical Systems*, vol. 40, no. 10, pp. 1–8, 2016, doi: 10.1007/s10916-016-0574-6.

[3] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," 2008. [Online]. Available: https://bitcoin.org/bitcoin.pdf. Accessed: January 2026.

[4] R. Zhang, R. Xue, and L. Liu, "Security and privacy on blockchain," *ACM Computing Surveys*, vol. 52, no. 3, pp. 1–34, 2019, doi: 10.1145/3316481.

[5] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, "MedRec: Using blockchain for medical data access and permission management," *2016 2nd International Conference on Open and Big Data (OBD)*, Vienna, Austria, 2016, pp. 25–30, doi: 10.1109/OBD.2016.11.

[6] A. Dubovitskaya et al., "Secure and trustable electronic medical records sharing using blockchain," *AMIA Annual Symposium Proceedings*, vol. 2017, pp. 650–659, 2017.

[7] H. Shuaib, S. Alhassan, and A. Barnawi, "Blockchain-based access control for secure personal health record sharing," *Sensors*, vol. 23, no. 1, p. 123, 2023, doi: 10.3390/s23010123.

[8] S. Griggs, O. Ossipova, D. Kohli, and M. Deshmukh, "Combining blockchain and AI for healthcare: Challenges and opportunities," *IEEE Access*, vol. 10, pp. 12823–12836, 2022, doi: 10.1109/ACCESS.2022.3146920.

[9] D. B. Rawat, "Blockchain-empowered decentralized data sharing and access control in IoT," *IEEE Internet of Things Journal*, vol. 9, no. 10, pp. 7544–7554, 2022, doi: 10.1109/JIOT.2021.3083312.

[10] T. T. Nguyen et al., "Federated learning with blockchain: A systematic literature review," *IEEE Access*, vol. 10, pp. 137624–137639, 2022, doi: 10.1109/ACCESS.2022.3222243.

[11] M. Al Omar, S. Chen, M. Z. A. Bhuiyan, and M. A. Rahman, "Privacy-preserving blockchain for healthcare data sharing," *Future Generation Computer Systems*, vol. 95, pp. 511–521, 2019, doi: 10.1016/j.future.2019.01.005.

[12] A. Rahman, A. Islam, and M. A. Hossain, "Smart contracts in healthcare: Security and privacy issues," *IEEE Consumer Electronics Magazine*, vol. 11, no. 3, pp. 42–49, 2022, doi: 10.1109/MCE.2022.3145562.

[13] HL7 Organization, "Fast Healthcare Interoperability Resources (FHIR) Specification," [Online]. Available: https://www.hl7.org/fhir/. Accessed: January 2026.