Privacy-Preserving and Secure Data Management in Medical Databases

Xinyun Peng

School of Computing and Data Science, Xiamen University Malaysia, Sepang, Malaysia

Abstract. Due to the exponential growth of clinical data and its inherent privacy sensitivity, ensuring robust privacy and security in medical database systems remains a critical challenge. This article reviews recent advances in Access Control, Data Encryption, and Blockchain technologies. Access control covers the framework design based on multi-institution attribute encryption, Elliptic Curve Cryptography (ECC) and Selective Ring Access Control (SRAC) mechanism. Data encryption discusses the encryption schemes of Fully Homomorphic Encryption (FHE), SFO-optimized Multi-Key Homomorphic Encryption (MHE-SFO) and Parallel AES (P-AES) combined with RSA. In blockchain applications, it introduces the permissioned public chain architecture based on fog computing and the management methodology of introducing sufficiency scoring (sfmd). Regarding these technologies, the article also points out the limitations of deep learning models, data encryption performance and searchability, and blockchain, suggesting future research to enhance data and search efficiency, privacy protection, trustworthy Artificial Intelligence (AI) and blockchain efficiency through techniques like semi-supervised learning, explainable AI, lightweight encryption and decentralized software. This article provides a comprehensive and in-depth review that can help readers in this field better understand the current and future directions.

1 Introduction

A database is a large organized collection of information stored in a computer system that is easy to access, manage and update. Database promotes data security, data protection, data integrity and swift data speed. Nowadays, databases are used to manage data in almost every field, and the Healthcare field is no exception. In the healthcare industry, databases play an important role in enabling centralized management and efficient access of data, improving the efficiency of medical services, and enhancing data-driven decision-making [1]. However, with the development of emerging technologies such as generative artificial intelligence, the healthcare industry currently faces many security vulnerabilities that pose significant risks to the confidentiality, integrity, and availability of healthcare data [2]. Therefore, how to strengthen privacy protection and data security has always been a hot research issue.

mec2109506@xmu.edu.my

The concept of database can be traced back to the 1960s, when it was proposed and put into practice by American engineer Bachman [3]. Databases solve the problems of scattered data, easy loss and low query efficiency when using traditional paper files or simple spreadsheets (such as Excel) to store data. Therefore, they are widely used in various industries such as business, retail, and education to manage data, optimize operations, and support decision-making [4]. In particular, the application of database technology in the medical industry has attracted widespread attention from the academic community. Samra et al. proposed the use of NoSQL technology to efficiently integrate clinical databases with traditional medical databases, and the Entity-Attribute-Value (EAV) Model for flexible data retrieval [5]. Soumma et al. proposed a Scalable Clinical Data Warehouse (NCDW) framework suitable for resource-constrained medical systems to support epidemiological analysis and public health decision-making in developing countries [6]. Epizitone et al. proposed a data-driven paradigm that aims to enhance the resilience and sustainability of healthcare applications through an integrated Health Information System (HIS) framework [7].

Among many studies, the security of medical databases is a key branch of research in this field, because they usually contain a large amount of highly sensitive privacy information, such as patient identity, diagnosis records and treatment plans. In order to strengthen privacy protection and data security in medical databases, many studies have proposed different methods and strategies in recent years. For example, blockchain technology eliminates single points of failure through a decentralized mechanism, uses data immutability and traceability to improve the security and transparency of medical data, and combines smart contracts to achieve refined permission control. Its encrypted storage and distributed architecture further ensure data integrity and effectively address the risks of information leakage and tampering in traditional medical systems [2]. Liu et al. proposed an innovative framework that combines Federated Learning (FL) and Neural Architecture Search (NAS) to protect data privacy through distributed training. Each participant (such as a hospital) only uploads model parameters instead of original data to avoid leakage of sensitive information, thus achieving the unity of high precision, high interpretability and privacy protection in medical image classification [8]. A new medical record search scheme, Privacy-Preserving Medical Record Searching Scheme (PMRSS), was proposed by Sun et al. This search method allows medical devices to encrypt medical record data and upload it to the cloud, which can effectively protect users' keywords, medical record content and access patterns and resist keyword guessing attacks and statistical attacks [9].

In recent years, significant progress has been made in enhancing the security of medical databases as numerous studies have proposed various improvement strategies, highlighting the need for a comprehensive review of the current research landscape. The rest of the article is organized as follows. Section 2 provides a comprehensive overview of existing approaches aimed at enhancing privacy protection and data security within medical databases. Section 3 critically examines the persistent challenges associated with these approaches and explores potential strategies for improvement. Finally, Section 4 synthesizes the key findings of the review and outlines directions for future research.

2 Method

2.1 Access control

2.1.1 Attribute-Based Access Control using Multiple Authorities

To enhance the protection of Personal Health Records (PHRs) stored in the cloud, Kumar et al. proposed a secure and scalable access control framework based on Multi-Authority Ciphertext-Policy Attribute-Based Encryption (MA-CP-ABE) leveraging Elliptic Curve Cryptography (ECC). The framework integrates an anonymous authentication mechanism, enabling users to authenticate without revealing their identities, thereby preserving privacy during data access. In this system, key distribution is handled by multiple independent authorities, while patients retain full control over their data by defining access policies and performing encryption. Authorized users can access and decrypt data only after passing anonymous authentication, ensuring both access control and identity protection. Furthermore, the scheme supports outsourced decryption, significantly reducing the computational overhead on user devices. This approach offers fine-grained access control, mitigates the limitations of traditional single-authority ABE schemes such as the key escrow problem and enhances the overall security, privacy, and flexibility of cloud-based PHR systems [10].

2.1.2 Selective Ring Access Control

Egala et al. proposed a model called "Fortified-Chain" based on Selective Ring based Access Control (SRAC), which is applied to the decentralized medical Internet of Things (IoMT) architecture based on blockchain, Distributed Data Storage System (DDSS) and hybrid computing, aiming to build a decentralized, flexible and reputation/context-based access control mechanism to adapt to complex IoMT medical scenarios. In the proposed system, patients define access control policies through the creation of "rings". Upon a data request, the system identifies the relevant ring: if a static permission is defined, access is granted directly; if dynamic evaluation is required, the requester's reputation score is compared to a predefined threshold. Access is allowed if the threshold is met (typically read-only); otherwise, it is denied. If no applicable rule exists, the request is forwarded to the patient for manual authorization. This mechanism balances automation with user-centric control. The SRAC model goes beyond the traditional, static access control list or role model and introduces dynamic evaluation to decide whether to grant access rights, especially when processing requests from remote or not fully trusted sources [11].

2.1.3 Steganography-Based Access Control for Encrypted EHRs

In addition to directly managing data permissions, Chinnasamy et al. proposed Hybrid Cryptographic Access Control for Electronic Health Record (HCAC-EHR), which achieves access control by securely hiding and distributing data decryption keys in an innovative way: steganography. The core of HCAC-EHR is an access control mechanism that combines hybrid encryption and steganography: the data owner first uses hybrid encryption technology to protect medical data, then uses steganography to embed the key symmetric decryption key into the carrier image, and uploads the encrypted data and the carrier image to the cloud. When the authorized user initiates a retrieval request, the cloud server returns the matching encrypted data and the corresponding carrier image. Finally, the user needs to use specific credentials (such as role access keys) obtained from the data owner to try to

extract the hidden Blowfish key from the carrier image; only when the key is successfully extracted can the user further decrypt and obtain the original medical data. HCAC-EHR combines powerful hybrid encryption technology to ensure the confidentiality of the data itself, and implements a covert, role-based key access control through steganography [12].

2.2 Data encryption

2.2.1 MedBlindTuner--Fully Homomorphic Encryption for Medical Image Fine-tuning

MedBlindTuner is a privacy protection framework proposed by Panzade et al. It is based on two core technologies: Fully Homomorphic Encryption (FHE) and Data-Efficient Image Transformer (DEiT). The core idea is to use FHE to support direct model fine-tuning training on encrypted medical data, so that medical institutions do not need to expose patient privacy data when using external computing resources. In this framework, the hospital extracts medical image features and transmits them to the cloud after encrypting them with FHE. The cloud service completes model fine-tuning and reasoning in a fully encrypted environment, and the returned results are decrypted by the hospital using a private key to obtain analysis information. MedBlindTuner combines FHE with DEiT for the first time, realizing end-to-end encryption model fine-tuning and ensuring data privacy during cloud training. Experiments on multiple public medical image datasets show that the framework can achieve accuracy close to that of plaintext training while maintaining privacy, and has good versatility [13].

2.2.2 PPEDL-MDTC--Optimal Homomorphic Encryption for Secure Medical Data Diagnosis

A novel model called PPEDL-MDTC seeks to connect Deep Learning (DL) technology with privacy-preserving encryption to provide an integrated solution for safe medical data transfer and precise illness detection. It was suggested by Alzubi and et al. The core technologies include Multiple key-based Homomorphic Encryption (MHE), Sailfish Optimization (SFO), Optimal Deep Neural Network (ODNN), etc. The sender uses SFO-optimized Multiple key-based Homomorphic Encryption (MHE-SFO) technology to encrypt medical data and transmit it through a public network; after the receiver decrypts the data, it uses cross-entropy based artificial butterfly optimization-based feature selection and inputs the results into a ODNN optimized by Chemical Reaction Optimization (CRO) for diagnostic classification. This model introduces different meta-heuristic algorithms in multiple stages such as key generation, feature selection, and model optimization, and integrates them into an integrated encryption and diagnosis system, which takes into account both data security and diagnostic accuracy [14].

2.2.3 Hybrid Encryption Scheme using Parallel-AES and RSA

Zhang et al. proposed a hybrid encryption scheme for cloud databases, combining the advantages of the improved symmetric encryption algorithm Parallel-Advanced Encryption Standard (P-AES) and the asymmetric encryption algorithm Rivest–Shamir–Adleman (RSA) to balance security and efficiency. The main process is that medical data is encrypted by P-AES, and the key K is encrypted and stored with the recipient's RSA public key. The user uses the private key to decrypt to obtain K, and then decrypts the data to restore the original text. Considering the large amount and high sensitivity of medical data

in cloud databases, symmetric encryption P-AES can process text efficiently, while asymmetric encryption algorithm RSA can manage symmetric keys with high security [15].

2.3 Blockchain

2.3.1 Fog Layer Blockchain Security Scheme

The Fog layer blockchain security scheme was presented by Ngabo et al. It refers to the deployment of a public-permissioned blockchain in the fog layer. Medical data is first collected through sensors at the edge layer and encrypted using private keys, and then the data is stored in the IPFS system. After that, a certificate for data access is created and posted as a transaction to the fog layer's blockchain. The blockchain verifies the authorization of authorized users, including doctors, who need to access the data. By altering keys or rescinding certificates, users (like patients) can always take away other people's access privileges. This scheme innovates the application of specific architectures, applying public-permissioned blockchain specifically to the fog computing layer of the Internet of Things to solve the security issues of medical data [16].

2.3.2 Blockchain Method for Medical Data Sufficiency Assessment and Management

Hovorushchenko et al. proposed a medical data management methodology based on blockchain technology, which includes two methods: medical data adequacy assessment method and medical data transaction execution method, aiming to solve the current lack of specific and effective methods for medical data management based on blockchain. Before the medical data is uploaded to the blockchain, the system evaluates whether it contains the three elements of type, time, and quality, and calculates the adequacy score (sfmd). If it is lower than the threshold, a supplementary suggestion is generated, and it is not allowed to be uploaded to the blockchain. The entry process includes data encryption, transaction generation, and writing to the blockchain after consensus verification; the acquisition process is initiated by the doctor, and the custodian sends the key after verification to complete data decryption and access. This methodology proposes a specific and quantifiable method to evaluate the integrity and quality of medical data before entering the blockchain, which is a key innovation aimed at controlling the quality of on-chain data from the source [17].

3 Discussion

3.1 Limitations and challenges

3.1.1 Potential Limitations of Data, Privacy, and Interpretability in Deep Learning-based Models

As a newly launched Optimal Homomorphic Encryption model, PPEDL-MDTC combines deep learning and artificial intelligence technologies [14], both of which have certain risks. The ODNN technology model used by PPEDL-MDTC relies on a large amount of labeled data for training, but medical data usually has problems such as high labeling costs, unbalanced samples, or limited privacy, which may affect the performance of the model in real scenarios. At the same time, deep learning models have Black Box problems. Medical

diagnosis requires transparent decision-making basis, but the complex structure of ODNN may make the results difficult to explain, affecting clinical trust.

3.1.2 Performance and Searchability Limitations of Data Encryption

Although encryption increases the security of medical databases, the technology itself makes it difficult for servers to search directly. Asymmetric encryption (such as RSA) and FHE involve complex operations (such as large exponential operations and ciphertext operations), which makes the encryption, decryption and ciphertext calculation process very time-consuming and has a high-performance overhead [13, 15]. Even with improved algorithms (such as P-AES) or hardware acceleration, there is still a significant gap compared to plaintext operations [15]. There is also a contradiction between encryption and search, because after the data is encrypted, the server cannot directly search the ciphertext for content, which seriously affects the data availability [12]. Existing searchable encryption schemes have their own shortcomings, and although hybrid encryption schemes attempt to combine the advantages of symmetric and asymmetric encryption, the specific implementation may still face problems such as large key size, long calculation time, high overhead, incompatibility with specific data types (such as multimedia), or relatively low overall efficiency/security [15].

3.1.3 Computational, Storage, and Ownership Challenges of Blockchain Architecture

As a technology that is not yet very mature, blockchain has certain shortcomings. The first is computing overhead and latency. Blockchain encryption (such as ECC digital signatures) and consensus mechanisms (such as verifying transactions) will increase the computing burden and may affect real-time performance, especially in medical emergency scenarios. The second is storage costs. The distributed ledger of blockchain requires all nodes to store a complete copy of the data, which may lead to a waste of storage resources, especially when the volume of medical data (such as images) is large [16]. Third, data ownership is controversial. Blockchain cannot automatically solve the problem of data ownership, such as the distribution of rights and interests between patients, hospitals or insurance companies.

3.2 Future prospects

In the future, these challenges and limitations are expected to be mitigated or addressed through technological advancements and improved management strategies. To address the challenges associated with deep learning models, semi-supervised learning can be considered for use. It reduces the high cost of labeled data by combining a small amount of labeled data with a large amount of unlabeled data for training, reducing the reliance on expert annotations. For example, the semi-supervised learning method proposed by Google Health significantly reduces the need for annotations in medical image analysis. In response to data privacy issues, more efficient privacy protection technologies such as federated learning can be introduced to reduce reliance on centralized data and ensure the security of sensitive data. Exploring explainable artificial intelligence may help alleviate the problem of model interpretability, and develop model output forms that can provide clinical users with decision-making basis prompts through feature importance ranking, visual analysis, and other means.

Lightweight FHE and index optimization are also future development directions. Although FHE provides strong privacy protection in theory, its high computational overhead limits its practical application. Future research may focus on developing Partially

Homomorphic Encryption (PHE) schemes and combining them with efficient index structures, such as keyword-based secure indexes, to improve search efficiency while ensuring security. In addition, for the various data types in medical data, future searchable encryption schemes need to be able to process multimedia data. For example, the combination of encrypted retrieval technology with fuzzy keyword search and image feature extraction will improve the practicality of the system in processing complex medical data.

For blockchain, in order to prevent the abuse of patients' medical privacy data, a special decentralized software can be developed in the future, which will allow users (such as patients or doctors) to open and view encrypted health documents without directly providing private keys. The software will ask the user to provide the sender's public key, and then automatically find and use the private key associated with it from the system for decryption or access [16]. This will prevent even authorized users (if their accounts are compromised) from easily obtaining and abusing the original private key, thereby enhancing security and simplifying user operations.

4 Conclusion

This article provides a comprehensive review of the current technologies for protecting the data security and privacy of medical databases. These technologies are divided into three major sections for investigation: access control, data encryption, and blockchain, and each major section is subdivided into different subsections according to the different types of technologies. Despite their benefits, these technologies present limitations, including the high data requirements and interpretability challenges of AI, the performance and searchability constraints of encryption, and the real-time constrains, storage burden, and data ownership concerns associated with blockchain. Future efforts may focus on leveraging semi-supervised learning, federated learning, and explainable AI techniques to enhance medical AI, adopting lightweight encryption with index optimization to improve retrieval efficiency, and developing searchable encryption and secure blockchain-based access mechanisms that support diverse data types.

References

1. Mahmudova, N.: 'The importance of using database management systems in hospitals', dspace.khazar.org, 2019
2. El Madhoun, N., Hammi, B.: 'Blockchain technology in the healthcare sector: overview and security analysis'. Proc. IEEE 14th Annu. Comput. Commun. Workshop Conf. (CCWC), Las Vegas, NV, USA, January 2024, pp. 0439–0446
3. Haigh, T.: 'How Charles Bachman Invented the DBMS, a Foundation of Our Digital World', Communications of the ACM, https://cacm.acm.org/opinion/how-charles-bachman-invented-the-dbms-a-foundation-of-our-digital-world/, accessed 28 April 2025
4. Deng, T.: 'The application of database systems in information management', Appl. Comput. Eng., 2024, 40, pp. 33–42
5. Samra, H.E., Li, A., Soh, B.: 'Design of a clinical database to support research purposes: Challenges and solutions', 2021
6. Soumma, S.B., Shahriar, F., Mahi, U.N., Abrar, M.H., Fahad, M.A.R., Hoque, A.S.M.L.: 'Design and Implementation of a Scalable Clinical Data Warehouse for Resource-Constrained Healthcare Systems', arXiv preprint arXiv:2502.16674, 2025

7. Epizitone, A., Moyane, S.P., Agbehadji, I.E.: 'A data-driven paradigm for a resilient and sustainable integrated health information systems for health care applications', J. Multidiscip. Healthc., 2023, pp. 4015–4025
8. Liu, X., Zhao, J., Li, J., Cao, B., Lv, Z.: 'Federated neural architecture search for medical data security', IEEE Trans. Ind. Inform., 2022, 18, (8), pp. 5628–5636
9. Sun, Y., Liu, J., Yu, K., Alazab, M., Lin, K.: 'PMRSS: Privacy-preserving medical record searching scheme for intelligent diagnosis in IoT healthcare', IEEE Trans. Ind. Inform., 2021, 18, (3), pp. 1981–1990
10. Kumar, A., Soni, A., Khare, S., Swarnkar, M.: 'Multi-authority Access Control Attribute Based Encryption with Anonymous Authentication for Maintaining Personal Health Record'. Proc. Int. Conf. Data Sci. Big Data Anal., Singapore, June 2023, pp. 447–459
11. Egala, B.S., Pradhan, A.K., Badarla, V., Mohanty, S.P.: 'Fortified-chain: a blockchain-based framework for security and privacy-assured internet of medical things with effective access control', IEEE Internet Things J., 2021, 8, (14), pp. 11717–11731
12. Chinnasamy, P., Deepalakshmi, P.: 'HCAC-EHR: Hybrid cryptographic access control for secure EHR retrieval in healthcare cloud', J. Ambient Intell. Humaniz. Comput., 2022, 13, (2), pp. 1001–1019
13. Panzade, P., Takabi, D., Cai, Z.: 'MedBlindTuner: Towards Privacy-Preserving Fine-Tuning on Biomedical Images with Transformers and Fully Homomorphic Encryption'. In: AI for Health Equity and Fairness: Leveraging AI to Address Social Determinants of Health (Springer Nature Switzerland, 2024), pp. 197–208
14. Alzubi, J.A., Alzubi, O.A., Beseiso, M., Budati, A.K., Shankar, K.: 'Optimal multiple key-based homomorphic encryption with deep neural networks to secure medical data transmission and diagnosis', Expert Syst., 2022, 39, (4), e12879
15. Zhang, F., Chen, Y., Meng, W., Wu, Q.: 'Hybrid encryption algorithms for medical data storage security in cloud database', Int. J. Database Manag. Syst. (IJDMS), 2019, 11
16. Ngabo, D., Wang, D., Iwendi, C., Anajemba, J.H., Ajao, L.A., Biamba, C.: 'Blockchain-based security mechanism for the medical data at fog computing architecture of internet of things', Electronics, 2021, 10, (17), 2110
17. Hovorushchenko, T., Moskalenko, A., Osyadlyi, V.: 'Methods of medical data management based on blockchain technologies', J. Reliab. Intell. Environ., 2023, 9, (1), pp. 5–16