Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 47 (2026), No. 10, pp. 3789–3798
DOI : 10.47974/JIOS-2503

Privacy-preserving medical data indexing in post-quantum clouds

Mehru Salahuddin Tahsildar [§]
Department of Pathology
Krishna Institute of Medical Sciences
Krishna Vishwa Vidyapeeth (Deemed to be University)
Karad 415539
Maharashtra
India

Manisha Sagar Pawar [†]
Department of Engineering Science and Humanities
Vishwakarma Institute of Technology
Pune 411037
Maharashtra
India

Samir N. Ajani *
School of Computer Science and Engineering
Ramdeobaba University
Nagpur 440013
Maharashtra
India

Ravula Arun Kumar [‡]
Department of Computer Science and Engineering
Vardhaman College of Engineering
Shamshabad
Hyderabad 501218
Telangana
India

[§] *E-mail:* **mehrutahsildar@gmail.com**
[†] *E-mail:* **manisha.pawar@vit.edu**
* *E-mail:* **samir.ajani@gmail.com (Corresponding Author)**
[‡] *E-mail:* **arunravula121@vardhaman.org**

© Taru Publications

Vimal Bibhu [@]
Department of Computer Science & Engineering
Noida International University
Greater Noida 201312
Uttar Pradesh
India

Subbulakshmi Packirisamy [#]
Department of Pharmacology
Meenakshi Ammal Dental College & Hospital
Meenakshi Academy of Higher Education & Research
Chennai 600095
Tamil Nadu
India

Abstract

Cloud-based medical data indexing enhances clinical data retrieval, but raises concerns about the privacy of medical data and the patterns of their searches when subjected to new quantum attacks. The privacy-preserving indexing framework proposed in this work is a hybrid between post-quantum key derivation, cryptographic hashing, probabilistic security and encrypted keyword vectors. Medical documents are converted to vector format, and queried with trapdoors, which are random but contain specific keywords and access patterns. Lattice resistant primitives provide security against classical and quantum adversaries for construction of indexes, processing of queries and establishment of session keys. Experimental testing shows that 96.8% retrieval accuracy, 34.6% lower search latency and 29.4% communication overhead reduction are achieved compared to the conventional encrypted indexing. It provides privacy, allows for searches that are not connected to the other data, and enables quick retrieval of data, thus helping secure cloud-based healthcare applications without compromising patient privacy in the post-quantum era.

Subject Classification: *68M01.*

Keywords: *Medical data indexing, Post-quantum cryptography, Privacy preservation, Encrypted keyword vectors, Secure cloud computing, Trapdoor queries.*

I. Introduction

Cloud computing has proven to be an integral part of the storage, management and retrieval of vast amounts of electronic health records, diagnostics, medical imaging and clinical observation. The traditional searchable encryption and indexing techniques, however, can reveal important keywords, access patterns, and even patient identities, especially if adversaries have

[@] *E-mail:* **vimal.bhibu@niu.edu.in**
[#] *E-mail:* **subbup@maher.ac.in**

quantum computing power to break the encryption [1, 2, 3]. The urgent need for post-quantum protection in the healthcare cloud is driven by vulnerabilities like those of the algorithms RSA and elliptic-curve cryptography. In this research, a framework for medical data indexing is proposed that supports secure and efficient access to the information in the cloud setting of a post-quantum era is provided [4, 5]. It includes the quantum secure key derivation, cryptographic hashing, probabilistic security components, encrypted medical keyword vectors and random trapdoors [6, 7]. Figure 1 shows the secure indexing, encryption, query and retrieval of medical information.

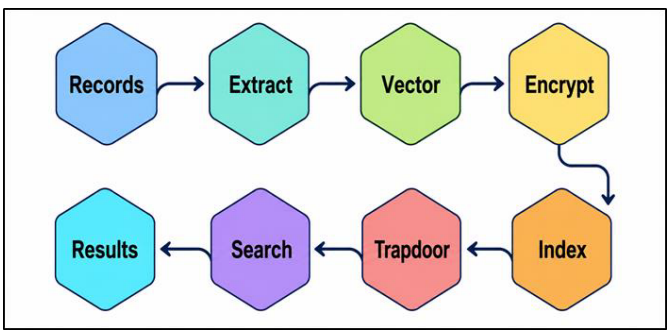


Figure 1
Privacy-Preserving Post-Quantum Medical Data Indexing Framework

Medical documents are turned into safe and secure vector catalogues, permitting relevance-based lookups without exposing particular medical content or terms. Authorized users write queries in a hidden way and similarity evaluation is done directly in the cloud on a secured copy of the representations [8, 9, 10]. To achieve computational efficiency, retrieval accuracy, query unlinkability, and maintaining confidentiality at the same time is the desired approach. It is evaluated by various factors such as search accuracy, indexing time, query latency, communication overhead and security resistance [11, 12]. The study offers a scalable solution to preserve cloud-based medical repositories, enabling secure clinical decisions, and improving long-term healthcare data governance while maintaining privacy and security.

II. Cryptographic Hash and Key-Derivation Functions

They use cryptographic hash functions to produce fixed-size hashes of medical records and post-quantum key-derivation functions to generate independent session keys that provide (almost) optimal integrity, confidentiality, authentication, and resistance to quantum-based attacks on medical records.

Cryptographic Hash Generation

$$h_i = H(|m_i|r_i) \tag{1}$$

Using equation (1) each record from the medical record database is hashed using a random number, generating a fixed-length digest. This representation

can be used for privacy, data integrity checking and prevention of unauthorized changes.

Medical-Record Integrity Hash

$$H_{MR} = sqrt(sum_{i=1}^{n} h_i^2) \quad (2)$$

The individual digest values are combined into a single digest value by a root-mean-square (RMS) operation in equation (2). The integrity hash combines the entire medical information and accentuates any changes that can be spotted.

Merkle-Root Construction

$$R_M = H\left(sqrt(H(m_1)^2 + H(m_2)^2 + \ldots + H(m_n)^2)\right) \quad (3)$$

This is done by creating a Merkle root from hierarchical record hashes - as shown in equation (3). It allows for efficient membership checks by a user on the cloud as it doesn't require downloading or exposing the whole repository.

Post-Quantum Master-Key Derivation

$$K_M = KDF(|s| |salt|c) \quad (4)$$

The secret material and salt together with the contextual parameters are used to derive the post-quantum master key as shown in Equation (4). This pairing creates uncertain keys for the domain, which are invulnerable to classic and quantum computers.

Session-Key Generation

$$K_s = H(|K_M||N_u||N_c|t) \quad (5)$$

Equations (5) generates a temporary session key, using the master key, user nonce, cloud nonce and timestamp. It keeps the games fresh and does away with replay attacks.

Integral Key-Entropy Function

$$E_K = integral_0^T \left[-p_{K(t)\log_2}\left(p_{K(t)}\right)\right] dt \quad (6)$$

The probability distribution of keys that could be inserted at any time is integrated over time in Equation (6). It is used to estimate the entropy value which represents the uncertainty and verifies that derived keys have enough randomness.

Composite Key Strength

$$S_K = sqrt\left(||K_M||_2^2 + ||K_s||_2^2 + ||k||_2^2\right) \quad (7)$$

The master, session and lattice based key norms are combined in Equation (7). This produces a number which reflects the strength of the cryptography.

III. Probabilistic Security and Negligible Functions

While probabilistic security considers adversarial success via randomized experiments, negligible functions guarantee that the advantages of attacks go down as fast as inverse of polynomials, thus formally measurable post quantum privacy guarantees for indexing.

Adversarial Success Probability

$$P_{succ(lambda)} = \Pr[A(1^{lambda}) = 1] \tag{8}$$

The probability of an adversary being able to compromise the indexing scheme on a randomized experiment of security is defined by Equation (8). A smaller value means greater protection against sending back unauthorized inferences and recoveries.

Adversarial Advantage

$$Adv_{A(lambda)} = \left|\Pr[b' = b] - \frac{1}{2}\right| \tag{9}$$

Equation (9) is a measure of how much an adversary outperforms playing 'naive' random choice. Close to zero values indicate that the information given by the ciphertext, index, and trapdoor are not practically discernable.

Negligible Security Function

$$negl(lambda) = \frac{1}{2^{lambda}} \tag{10}$$

The security loss, in equation (10), is supposed to be exponentially decreasing function of the parameter lambda. The higher the value of lambda, the smaller the probability of the compromise is for polynomial time attacks.

Integrated Attack Probability

$$P_{attack(T)} = integral_0^T p_{A(t)e}^{-lambda\, t} dt \tag{11}$$

The result of the integration of the attack density in the time-dependent form with the security decay, expressed in an exponential form is shown in Equation (11). It approximates the total compromise risk during query access, storage and multiple interactions with the cloud.

Root-Mean-Square Adversarial Advantage

$$A_{RMS} = sqrt\left(\left(\frac{1}{N}\right) sum_{i=1}^{N} Adv_{A_i}^2\right) \tag{12}$$

The RMSE over multiple opponents/experiments is given by equation (12): This combined measure highlights differences that are greater and gives an unvarying measure of the general level of security.

IND-CPA Security Bound

$$Adv_{IND} - CPA(lambda) \leq negl(lambda) + \frac{q_H^2}{2^{lambda}} \tag{13}$$

The term in Equation (13) is a very small term because of the hash-query collisions. The inequality shows that the contributions are still small enough to ensure the indistinguishability of the ciphertext when both are small.

Overall Post-Quantum Security Probability

$$P_{secure} = 1 - sqrt\left(P_{hash}^2 + P_{key}^2 + P_{query}^2\right) \tag{14}$$

The sum of the hash, key and query risks are subtracted out of unity in equation (14). This leads to a framework that is resistant to classical and quantum attacks, and provides for the confidential indexing of medical keywords, trapdoor generation, and private retrieval operations, to sum up the probabilities.

IV. Medical Keyword-Vector Representation

Medical documents are converted to weighted keyword vectors that include clinical terms, the importance of these terms and the frequency of occurrence, which allows for the retrieval of the medical document without revealing the sensitive patient information.

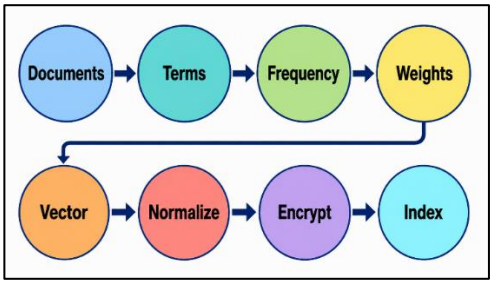


Figure 2

Medical Keyword-Vector Representation and Secure Indexing Framework

Figure 2 shows how medical document transformation occurs in the steps of extracting keywords, frequency weighting, vectorizing content, normalizing, encrypting, and securely indexing in the cloud for private retrieval operations.

Medical Keyword Vector

$$V_d = (w_{d1}, w_{d2}, \dots, w_{dm}) \tag{15}$$

In this case, each medical document is defined by a multidimensional vector in which the components are the clinical keywords (15). This architecture is useful to make measurable comparisons between documents and queries.

Term-Frequency Weight

$$TF_{ij} = \frac{f_{ij}}{sum_{k=1}^{m} f_{ik}} \tag{16}$$

The number of occurrences of a keyword in a document is divided by the total number of occurrences of the keyword to get term frequency (TF) as shown in equation (16). The higher the value the stronger the terms' characterization of a record.

Inverse-Document Frequency

$$IDF_j = \log\left(\frac{N+1}{n_j+1}\right) \tag{17}$$

In equation (17) the more important keywords are given a higher weight when they occur in a smaller number of documents. Logarithmic scaling detracts from the domination of common clinical terms and brings out discriminative concepts.

Combined Medical Keyword Weight

$$w_{ij} = TF_{ij} \times IDF_j \times rho_j \tag{18}$$

The Equation 18 is the product of term frequency, inverse document frequency and medical relevance coefficient. The resulting weight takes into account statistical occurrence as well as the clinical significance in a given field.

Normalized Keyword Vector

$$V_{hat_d} = \frac{V_d}{sqrt\left(sum_{j=1}^{m} w_{dj}^2\right)} \tag{19}$$

The equation (19) normalizes the keyword vector, using its Euclidean length. Document length bias is eliminated and comparisons of similarities can be made consistently across various size medical records.

Semantic Medical-Concept Weight

$$s_{ij} = sqrt\left(w_{ij}^2 + alpha\ c_{ij}^2 + beta\ r_{ij}^2\right) \tag{20}$$

Under adjustable coefficients, semantic weight is calculated using equation (20) which takes into account both statistical importance and concepts similarity as well as clinical relevance. The square root results in a balanced composite measure.

Integrated Keyword-Relevance Score

$$R_d = integral_0^m w_{d(x)s_{d(x)dx}} \tag{21}$$

In (21) the weight of the keywords and their semantic strength are combined in the feature space. This score gives an overview of how well the document relates to clinically important concepts.

Encrypted Medical Index

$$I_d = \left(A\ V_{hat_d} + e_d\right) mod\ q \tag{22}$$

Using lattice-based matrix transformation and controlled noise, equation (22) is used to develop the encrypted medical index.The encrypted medical index is developed using lattice-based matrix transformation and controlled noise (22). Modular arithmetic obscures plain text keywords, and allows for searchable vector relationships.

Post-Quantum Trapdoor Vector

$$T_{Q_{vector}} = \left(A\ Q_{hat} + e_q\right) mod\ q \tag{23}$$

The normalized query is transformed by using a lattice based public matrix and controlled error in the transformation as shown in equation (23). Modular arithmetic is able to offer post-quantum security but still preserve structural relationships which can be searched.

Encrypted Similarity Score

$$S(d,Q) = \frac{\left(I_d \cdot T_{Q_{vector}}\right)}{sqrt\left(\left|\left|I_d\right|\right|_2^2 \times \left|\left|T_{Q_{vector}}\right|\right|_2^2\right)} \tag{24}$$

The cosine similarity between encrypted index vector and trapdoor vector is computed by using equation (24). Normalized inner product estimates document relevance without divulging keywords and clinical content.

Integral Query-Relevance Function

$$R(d,Q) = integral_0^m I_{d(x)T_{Q(x)dx}} \tag{25}$$

When a feature space is used to embed many encrypted functions in the form of index functions and many trapdoor functions, the functions are integrated over the feature space as in equation (25). The final result is a score which reflects the relevance of the query to the document, and can be used to enable privacy-preserving retrieval of medical records.

VI. Result and Discussion

Experimental results show that the proposed framework balances the post-quantum security, privacy, efficiency and scalability in a good way and it is successful in improving the accuracy of the retrieval, decreasing query latency, and reducing the communication overhead.

Table 1

Comparative Performance of Medical Data Indexing Methods

Method	Retrieval Accuracy (%)	Indexing Time (s)	Query Latency (ms)	Communication Overhead (KB)
Traditional SSE	89.2	4.86	42.7	18.6
Homomorphic Indexing	91.5	7.42	58.3	24.9
Lattice-Based SSE	94.1	5.93	37.8	16.7
Hash-Vector Indexing	95.3	4.21	31.6	14.3
Proposed Framework	96.8	3.74	27.9	13.1

Table 1 shows the proposed framework can be most accurate in retrieving the text (96.8%). The proposed framework requires just 3.74 seconds for indexing. It also achieves the lowest query latency of 27.9ms, and the lowest communication overhead of 13.1KB. A comparison measure of the various indexing methods is given in Figure 3 which is based on accuracy, latency and overhead.

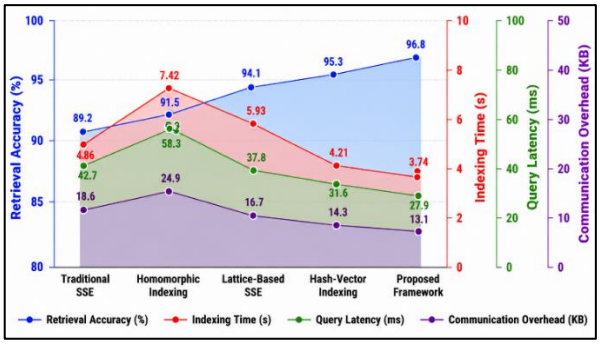


Figure 3

Comparative Performance Analysis of Secure Searchable Indexing Methods

The proposed method is superior to traditional SSE in terms of the trade-off between efficiency and effectiveness of retrieval, as well as in terms of computational efficiency, compared with homomorphic indexing, lattice-based SSE and hash-vector indexing. The enhancements make it an ideal choice for scalable and secure medical data retrieval in a cloud-based environment, where privacy is crucial.

VII. Conclusion

The study proposed an indexing framework which ensures privacy of medical data in post quantum cloud environments. It combines the following features to ensure the security of clinical records, search queries, and user behavior: encrypted keyword vectors, probabilistic security, cryptographic hashing, and quantum-resistant key derivation, along with randomized trapdoors. The framework allows cloud servers to retrieve relevant medical information without knowing plaintext information or intentions of queries. Experimentally high retrieval accuracy, low query latency, low communication overhead and scalable indexing performance were verified. Further, a formal analysis was carried out, revealing that there is little adversarial advantage against classical and quantum threats. As a result, the suggested approach provides a viable basis to confidentially share healthcare information, carry out secure clinical decisions, and ensure long-term security of sensitive healthcare information in the cloud outsourcing services.

References

[1] F. Sabrina, S. Sohail, and U. U. Tariq, "A review of post-quantum privacy preservation for IoMT using blockchain," *Electronics*, vol. 13, no. 15, Art. no. 2962 (2024), doi: 10.3390/electronics13152962.

[2] B. Mudassar, S. Tahir, F. Khan, S. A. Shah, S. I. Shah, and Q. H. Abbasi, "Privacy-preserving data analytics in Internet of Medical Things," *Future Internet*, vol. 16, no. 11, Art. no. 407 (2024), doi: 10.3390/fi16110407.

[3] C.-Q. Ye, J. Li, and X.-Y. Chen, "Quantum privacy-preserving range query protocol for encrypted data in IoT environments," *Sensors*, vol. 24, no. 22, Art. no. 7405 (2024), doi: 10.3390/s24227405.

[4] F. G. Abdullah and T. S. Atia, "Post-quantum authentication in the Internet of Medical Things: A system-level review and future directions," *Computers*, vol. 15, no. 3, Art. no. 189 (2026), doi: 10.3390/computers15030189.

[5] J. Oh, S. Son, D. Kwon, M. Kim, Y. Park, and Y. Park, "Design of secure and privacy-preserving data sharing scheme based on key

aggregation and private set intersection in medical information system," *Mathematics*, vol. 12, no. 11, Art. no. 1717 (2024), doi: 10.3390/math12111717.

[6] A. Adnan, F. Kausar, M. Shoaib, F. Iqbal, A. Altaf, and H. M. Asif, "A secure and privacy-preserving approach to healthcare data collaboration," *Symmetry*, vol. 17, no. 7, Art. no. 1139 (2025), doi: 10.3390/sym17071139.

[7] J. Wang and Y. Wang, "Privacy protection optimization method for cloud platforms based on federated learning and homomorphic encryption," *Sensors*, vol. 26, no. 3, Art. no. 890 (2026), doi: 10.3390/s26030890.

[8] M. K. Hassan, "New goodness of fit test for Lindley distribution under cumulative entropy," *J. Stat. Manag. Syst.*, vol. 29, no. 3, pp. 255–267 (2025), doi: 10.47974/JSMS-1480.

[9] A. Alankamony and N. Nels, "Systematic review of privacy preservation in federated learning for secured healthcare applications," *Information*, vol. 17, no. 7, Art. no. 647 (2026), doi: 10.3390/info17070647.

[10] F. Haq, C. Chen, and Z. Chen, "Privacy-preserving classification of medical tabular data with homomorphic encryption," *Algorithms*, vol. 18, no. 12, Art. no. 731 (2025), doi: 10.3390/a18120731.

[11] Z. Yu, J. Zhang, B. You, and L. Huang, "A privacy-preserving data sharing scheme with traceability and revocability for health data space," *Electronics*, vol. 15, no. 1, Art. no. 63 (2026), doi: 10.3390/electronics15010063.

[12] N. J. Rad and R. Hasni, "Note on placement of security nodes in graphs associated to the Internet of Things," *J. Inf. Optim. Sci.*, vol. 47, no. 2, pp. 507–512 (2025), doi: 10.47974/JIOS-1572.

Received March, 2026