

Machine Learning-Driven Query Optimization for Secure and Efficient Blockchain Electronic Medical Record Management

Mission Franklin¹, Alalibo Ralph Fiberesima²

¹Department of Computer Engineering, Rivers State University, Port Harcourt, Rivers State, Nigeria

²Department of Computer Science, Federal Polytechnic of Oil and Gas, Bonny, Rivers State, Nigeria

DOI: <https://doi.org/10.51584/IJRIAS.2026.11090045>

Received: 16 September 2026; Accepted: 21 September 2026; Published: 08 October 2026

ABSTRACT

The increasing volume of electronic medical records (EMRs) has created a growing need for healthcare information systems that can provide secure, reliable, and efficient access to patient data. Although blockchain offers strong security, immutability, transparency, and controlled data sharing, the increasing size and complexity of blockchain records can affect query performance and data retrieval. This study proposes a Machine Learning-Driven Query Optimization Framework for secure and efficient blockchain-based EMR management. The framework integrates machine learning, permissioned blockchain, smart contracts, and security mechanisms to optimize query execution while protecting sensitive medical information. Machine learning is employed to analyse query patterns and support the selection of efficient execution strategies, thereby reducing processing overhead and improving system responsiveness. The proposed framework was evaluated by experts using functional, security, privacy, reliability, scalability, usability, and query-efficiency criteria. The results showed positive expert assessments across all criteria, with Query Efficiency recording the highest mean score (4.28), followed by Functionality (4.23) and Security (4.16). These findings suggest that integrating machine learning with blockchain can improve the efficiency of EMR data retrieval without undermining essential security and access-control requirements. The framework provides a foundation for intelligent and scalable healthcare information management and can support secure information exchange among authorized healthcare stakeholders. Further empirical testing with real-world healthcare workloads is recommended to validate its performance and scalability.

Keywords: Blockchain, Electronic Medical Records, Machine Learning, Query Optimization, Healthcare Data Security, Smart Contracts.

INTRODUCTION

The healthcare industry has experienced a rapid digital transformation over the past decade, leading to the widespread adoption of Electronic Medical Records (EMRs) to improve the quality, accessibility, and continuity of patient care. EMRs provide a digital platform for storing, managing, and retrieving patients' clinical information, enabling healthcare professionals to make timely and informed decisions while reducing the inefficiencies associated with paper-based record systems. The increasing volume of healthcare data generated from hospitals, diagnostic laboratories, pharmacies, wearable devices, and telemedicine platforms has further heightened the demand for secure, scalable, and efficient data management systems (Kruse et al., 2018; World Health Organization, 2021).

In spite of these advantages, conventional centralized EMR systems continue to face significant challenges related to security, privacy, interoperability, and data availability. Centralized databases are susceptible to cyberattacks, unauthorized access, insider threats, and single points of failure, all of which may compromise sensitive patient information. Furthermore, fragmented healthcare information systems often hinder the seamless exchange of medical records across healthcare providers, thereby affecting the continuity and quality of patient care (Kruse et al., 2018; McGhin et al., 2019).

Blockchain technology has emerged as a promising solution for addressing these challenges because of its

decentralized architecture, immutability, transparency, and cryptographic security. Blockchain maintains a distributed ledger in which every transaction is validated through consensus mechanisms and permanently recorded, making unauthorized modifications extremely difficult. These characteristics enhance data integrity, accountability, and traceability while enabling trusted data sharing among authorized healthcare stakeholders. Therefore, blockchain has attracted considerable attention as a secure platform for electronic medical record management and healthcare information exchange (Azaria et al., 2016; Kuo et al., 2017).

Although blockchain can improve the security, integrity, and controlled sharing of healthcare information, its use as a data-management infrastructure introduces challenges for efficient data storage and retrieval. As blockchain ledgers accumulate transactions, healthcare applications may require increasingly complex queries over current and historical records, including joins, aggregations, and temporal queries. Such query requirements can be difficult to process efficiently because blockchain data structures and query interfaces differ from those of conventional database systems and provide more limited support for complex queries and historical data access (Przytarski et al., 2022). These limitations create a need for specialized query-processing and optimization techniques. In conventional database environments, adaptive query processing has been proposed to address situations in which complex queries, limited statistics, or changing execution conditions make static optimize-then-execute strategies less effective (Deshpande et al., 2007).

These limitations can increase query latency, reduce throughput, and affect the responsiveness of healthcare applications that require real-time access to patient information, particularly during emergency medical situations (Dinh et al., 2018; Xu et al., 2021).

Machine learning (ML) has become an effective approach for improving database performance by enabling intelligent and adaptive query optimization. Rather than relying solely on static cost models and predefined optimization rules, machine learning algorithms learn from historical query workloads, execution statistics, and data access patterns to predict efficient query execution plans. This capability allows database systems to adapt to changing workloads, improve resource allocation, reduce execution time, and enhance overall system performance. Advances in learned query optimization demonstrate that ML can outperform traditional optimization strategies under dynamic and large-scale data environments (Marcus & Papaemmanouil, 2019; Zhou et al., 2022).

Integrating machine learning with blockchain offers a potential approach to improving the security and query-processing efficiency of electronic medical record systems. Blockchain provides tamper resistance, traceability, and controlled access, while machine learning can learn from query characteristics and execution data to select efficient query plans (Marcus et al., 2019; Vaidya et al., 2022). This combination enables healthcare organizations to manage growing volumes of medical records securely while maintaining the fast data retrieval required for effective clinical decision-making (Agbo et al., 2019).

In addition, modern healthcare systems must comply with stringent regulatory and privacy requirements while supporting secure information sharing among hospitals, clinics, laboratories, insurance providers, and patients. Intelligent query optimization techniques can improve scalability and responsiveness without compromising blockchain's security guarantees. Machine learning models such as Random Forest, Gradient Boosting, Extreme Gradient Boosting (XGBoost), and Reinforcement Learning have demonstrated considerable potential for learning workload characteristics and adapting query execution strategies to changing operational conditions, thereby improving the efficiency of large-scale distributed data management systems (Marcus & Papaemmanouil, 2019; Zhou et al., 2022).

Against this background, this study proposes a Machine Learning-Driven Query Optimization Framework for Secure and Efficient Blockchain Electronic Medical Record Management. The proposed framework seeks to integrate machine learning algorithms with blockchain technology to improve query processing performance while preserving the confidentiality, integrity, availability, and privacy of electronic medical records. The framework will be evaluated using performance indicators such as query execution time, response latency, throughput, resource utilization, prediction accuracy, and security effectiveness. The findings are expected to contribute to the development of intelligent, scalable, and secure healthcare information systems capable of supporting next-generation digital healthcare services.

1.1 Statement of the Problem

The increasing adoption of Electronic Medical Records (EMRs) has significantly improved healthcare delivery by enabling digital storage and sharing of patient information. However, conventional EMR systems remain vulnerable to cyberattacks, unauthorized access, data breaches, and interoperability challenges, which threaten the confidentiality, integrity, and availability of sensitive patient data (Kruse et al., 2018). Blockchain technology has been proposed as a secure alternative because it provides decentralized storage, immutable records, and transparent audit trails that enhance trust and data integrity (Kuo et al., 2017; Agbo et al., 2019).

Despite these security advantages, blockchain-based EMR systems often experience poor query performance as the volume of medical records and transactions increases. The distributed nature of blockchain, together with consensus mechanisms and immutable ledger structures, can lead to higher query execution time, increased latency, and greater computational overhead when retrieving patient records (Dinh et al., 2018). These performance limitations may delay timely access to critical medical information, particularly in emergency healthcare situations where rapid clinical decision-making is essential.

Traditional query optimization techniques are largely rule-based or cost-based and are not sufficiently adaptive to the dynamic workloads and complex data structures of blockchain environments. Recent advances in machine learning have demonstrated its capability to improve database query optimization by learning from historical workloads and predicting efficient query execution plans (Marcus & Papaemmanouil, 2019; Zhou et al., 2022). However, existing studies have primarily examined blockchain security or machine learning-based query optimization independently, with limited research integrating both approaches for blockchain-enabled electronic medical record management.

Thus, there is a need for an intelligent framework that combines machine learning-driven query optimization with blockchain technology to improve query efficiency while preserving the security and privacy of electronic medical records. Addressing this gap will contribute to the development of secure, scalable, and high-performance healthcare information systems capable of supporting timely and reliable access to patient data.

1.2 Aim and objectives of the Study

The aim of this study is to develop a machine learning-driven query optimization framework for secure and efficient blockchain electronic medical record management.

The specific objectives are to:

1. Examine the challenges of query processing in blockchain electronic medical record systems;
2. Develop a machine learning-driven query optimization framework for secure and efficient blockchain electronic medical record management;
3. Evaluate the performance of the proposed framework using query processing and security metrics.
4. Compare the proposed framework with conventional blockchain query processing approaches.

1.3 Significance and Scope of the Study

This study proposes a machine learning-driven query optimization framework for secure and efficient blockchain electronic medical record management. It contributes to healthcare informatics by improving query processing performance while enhancing the security and privacy of electronic medical records. The study focuses on the development and evaluation of the proposed framework within a permissioned blockchain environment using performance metrics such as query execution time, response latency, throughput, resource utilization, prediction accuracy, and security effectiveness. The findings will provide valuable insights for healthcare providers, researchers, and system developers in designing intelligent, secure, and scalable electronic medical record systems.

LITERATURE REVIEW

2.1 Conceptual Review

- **Electronic Medical Records (EMRs):** Electronic Medical Records (EMRs) are digital versions of patients' medical records that contain information such as medical history, diagnoses, medications, laboratory results, and treatment plans. They enable efficient storage, retrieval, and sharing of patient information, thereby improving healthcare delivery and continuity of care (HealthIT.gov, 2023; World Health Organization, 2021). The adoption of EMRs has evolved from simple digital record-keeping to integrated healthcare systems that support interoperability and emerging technologies such as artificial intelligence and blockchain. EMRs improve healthcare quality by reducing medical errors, enhancing information accessibility, and increasing operational efficiency. However, they remain vulnerable to cybersecurity threats, data breaches, interoperability challenges, and high implementation costs (Kruse et al., 2018; McGhin et al., 2019).
- **Blockchain Technology in Healthcare:** Blockchain is a decentralized distributed ledger technology that securely records transactions in an immutable and transparent manner without requiring a central authority (Nakamoto, 2008). Its key characteristics include decentralization, transparency, immutability, cryptographic security, and consensus-based validation, which ensure data integrity and trust (Casino et al., 2019). Blockchain networks are categorized into public, private, and consortium blockchains. In healthcare, private and consortium blockchains are preferred because they provide controlled access, enhanced privacy, and regulatory compliance (Kuo et al., 2017). A blockchain architecture comprises distributed nodes, transactions, blocks, consensus mechanisms, and smart contracts that collectively secure and manage healthcare data. Blockchain has numerous healthcare applications, including electronic medical record management, secure health information exchange, patient identity management, pharmaceutical supply chain tracking, and clinical data sharing. These applications improve data security, privacy, and interoperability across healthcare systems (Agbo et al., 2019).
- **Blockchain-Based Electronic Medical Record Management:** Blockchain-based Electronic Medical Record (EMR) management integrates blockchain technology with healthcare information systems to provide secure, decentralized, and tamper-resistant storage and sharing of patient records. Blockchain-enabled EMRs improve data integrity, interoperability, and patient control over medical information by maintaining immutable records and audit trails (Azaria et al., 2016; Kuo et al., 2017). Smart contracts are self-executing programs deployed on the blockchain that automate access control, patient consent, and data-sharing processes. They ensure that only authorized users can access medical records based on predefined rules, thereby improving efficiency and reducing administrative overhead (Azaria et al., 2016). The adoption of blockchain in EMR management offers benefits such as enhanced security, transparency, data integrity, and interoperability. However, challenges including scalability, query performance, storage overhead, regulatory compliance, and integration with legacy healthcare systems remain significant (Agbo et al., 2019). Furthermore, protecting patient privacy requires robust encryption, authentication, and access control mechanisms to ensure compliance with healthcare data protection regulations (McGhin et al., 2019).
- **Query Processing in Blockchain Systems:** Query processing is the process of retrieving and manipulating data from a database in response to user requests. In blockchain systems, query processing involves searching and accessing data stored across a distributed ledger while maintaining data integrity and consistency (Gan et al., 2023). Blockchain data are organized into sequentially linked blocks containing validated transactions secured through cryptographic hash functions. This immutable structure enhances security but makes data retrieval more complex than in traditional relational databases (Nakamoto, 2008). Query execution in blockchain systems typically involves locating relevant blocks, validating transactions, and retrieving records through smart contracts or blockchain application programming interfaces (BAPIs). Unlike conventional databases, blockchain queries often require additional validation steps, increasing computational overhead (Dinh et al., 2018). As blockchain datasets grow, query processing faces challenges such as high execution latency, limited scalability, storage overhead, and increased resource consumption. These limitations highlight the need for intelligent query optimization techniques, including machine learning, to improve the efficiency of blockchain-based applications such as electronic medical record management (Dinh et al., 2018; Zhou et al., 2022).

- **Query Optimization Techniques:** Query optimization is the process of selecting the most efficient execution plan to retrieve data while minimizing execution time and resource consumption. Traditional query optimization relies on predefined heuristics and database statistics to determine efficient execution plans (Graefe, 1993). Two widely used approaches are cost-based and rule-based optimization. Cost-based optimization evaluates multiple execution plans using estimated costs such as CPU, memory, and I/O usage, whereas rule-based optimization applies predefined rules and heuristics to select query plans (Graefe, 1993). Although effective in conventional databases, these approaches are less suitable for dynamic blockchain environments with changing workloads. Adaptive query optimization addresses this limitation by continuously adjusting execution plans based on runtime conditions and workload characteristics, thereby improving performance in dynamic systems (Deshpande et al., 2007). In addition, indexing and caching techniques enhance query performance by reducing data access time and minimizing repeated computations, making them essential components of modern database optimization (Silberschatz et al., 2020).
- **Machine Learning for Query Optimization:** Machine learning (ML) is a branch of artificial intelligence that enables systems to learn from data and improve decision-making without explicit programming. In database systems, ML is increasingly used to predict efficient query execution plans, estimate query costs, and optimize resource utilization (Jordan & Mitchell, 2015). Supervised learning uses labelled historical query data to train models that predict query execution time or the best execution plan, while reinforcement learning enables an agent to learn optimal query optimization strategies through continuous interaction with the database environment (Marcus & Papaemmanouil, 2019). Learned query optimization replaces or enhances traditional rule-based and cost-based optimizers by using ML models that adapt to changing workloads and data distributions. Common algorithms used for query optimization include Decision Trees, Random Forest, XGBoost, Deep Neural Networks (DNNs), and Deep Reinforcement Learning, which have demonstrated improved query performance in modern database systems (Marcus & Papaemmanouil, 2019; Zhou et al., 2022).
- **Security and Privacy in Blockchain Medical Records:** Security and privacy are critical requirements for blockchain-based electronic medical record (EMR) systems because they protect sensitive patient information from unauthorized access and cyber threats. Confidentiality ensures that medical records are accessible only to authorized users, while integrity guarantees that records cannot be altered without authorization. Availability ensures that authorized users can access medical records whenever needed for healthcare delivery (Kuo et al., 2017). Blockchain strengthens security through authentication, authorization, and access control, ensuring that only verified users can access or modify patient records. In addition, encryption protects medical data during storage and transmission, while privacy-preserving techniques such as cryptographic hashing, digital signatures, and secure key management help maintain patient privacy and regulatory compliance (Agbo et al., 2019; McGhin et al., 2019). Together, these mechanisms provide a secure and trustworthy environment for managing electronic medical records on blockchain platforms.

2.2 Empirical Review

1. Azaria et al. (2016) developed MedRec, a blockchain-based Electronic Medical Record (EMR) framework that improved data security, patient control, and secure information sharing through smart contracts. However, the framework did not address query processing efficiency or intelligent query optimization.
2. Kuo et al. (2017) reviewed blockchain applications in healthcare and found that the technology enhances security, transparency, and interoperability. Despite these benefits, challenges such as scalability, transaction latency, and computational overhead remain significant.
3. Dinh et al. (2018) proposed BLOCKBENCH, a benchmarking framework for evaluating private blockchain platforms. Their findings showed that blockchain systems provide strong security but suffer from lower query and transaction performance than traditional databases, highlighting the need for performance optimization.
4. Agbo et al. (2019) conducted a systematic review of blockchain in healthcare and concluded that while blockchain improves privacy, security, and interoperability, issues such as scalability, storage overhead, and system performance remain unresolved, requiring intelligent optimization techniques.

5. Marcus and Papaemmanouil (2019) applied deep reinforcement learning to database query optimization and demonstrated improved query execution compared to traditional methods. However, their work focused on relational databases and did not consider blockchain-based Electronic Medical Record systems.
6. Chamola et al. (2022) proposed an AI-assisted blockchain framework integrated with IPFS for secure Electronic Medical Record (EMR) management. The framework improved data security, privacy, and accessibility but did not address machine learning-driven query optimization.
7. Hassija et al. (2022) developed a blockchain and machine learning framework for secure medical consultations. The study enhanced healthcare efficiency, reduced costs, and improved security, but focused on consultation services rather than blockchain query optimization.
8. Gan et al. (2023) proposed an encrypted blockchain-based medical data search method with access control to improve secure and efficient retrieval of electronic medical records. The study enhanced search efficiency, privacy, and data security but did not apply machine learning techniques for query optimization in blockchain-based EMR systems.
9. Palaiokrassas et al. (2024) conducted a systematic mapping study on machine learning applications in blockchain systems. The review highlighted the growing use of machine learning for blockchain analytics but did not develop or evaluate a framework for query optimization in blockchain-based Electronic Medical Record systems.
10. Zhou et al. (2023) proposed MSTDB, a scalable semantic blockchain database that uses a hybrid on-chain/off-chain storage architecture and Merkle Semantic Trie indexing to improve query processing. Experimental results showed significant improvements in query efficiency and scalability. However, the study focused on blockchain database architecture and indexing rather than applying machine learning techniques for query optimization in Electronic Medical Record (EMR) systems.

The reviewed studies demonstrate that blockchain technology has significantly improved the security, privacy, and interoperability of Electronic Medical Record (EMR) systems, while machine learning has enhanced query optimization in conventional databases. However, limited research has integrated machine learning-driven query optimization into blockchain-based EMR systems. This gap motivates the present study, which proposes a machine learning-driven framework to improve query processing efficiency, scalability, and secure management of blockchain-based electronic medical records.

METHODOLOGY

3.1 Research Design: The study adopts a Design Science Research (DSR) approach based on Peffers et al. (2007) to design, implement, demonstrate, and evaluate a machine-learning-driven query optimization framework for secure blockchain-based electronic medical record (EMR) management. The artefact integrates machine learning, PostgreSQL, and Hyperledger Fabric to improve query performance while maintaining data security.

3.2 Proposed Framework: Figure 1 presents the proposed Machine Learning-Driven Query Optimization Framework for Secure and Efficient Blockchain Electronic Medical Record (EMR) Management. The framework comprises six integrated layers: EMR Repository, Permissioned Blockchain Network, Smart Contract Layers, Machine Learning Query Optimizer, Query Processing Engine, and Security and Privacy Module. The EMR Repository stores medical records, the Permissioned Blockchain Network provides controlled and tamper-resistant transactions, and Smart Contracts enforce access rules. The Machine Learning Query Optimizer selects efficient execution plans, while the Query Processing Engine executes optimized queries. The Security and Privacy Module provides authentication, authorization, encryption, access control, and privacy protection.

The key contribution is the integration of adaptive machine-learning-based query optimization with secure permissioned blockchain EMR management. Unlike approaches that focus mainly on blockchain security and data integrity, the framework simultaneously addresses query performance, security, and privacy, enabling efficient retrieval of medical records without compromising their protection.

A query is generated from the EMR data, and its characteristics are extracted as input features for the Random

Forest and XGBoost models, which predict the optimal execution plan based on the observed workload. The selected plan is subjected to smart-contract authorization and executed through PostgreSQL and Hyperledger Fabric, with AES-256 encryption, SHA-256 hashing, and RBAC supporting data security. Performance is evaluated using execution time, latency, throughput, resource utilization, query optimization accuracy, and security/privacy performance against the conventional PostgreSQL query optimizer.

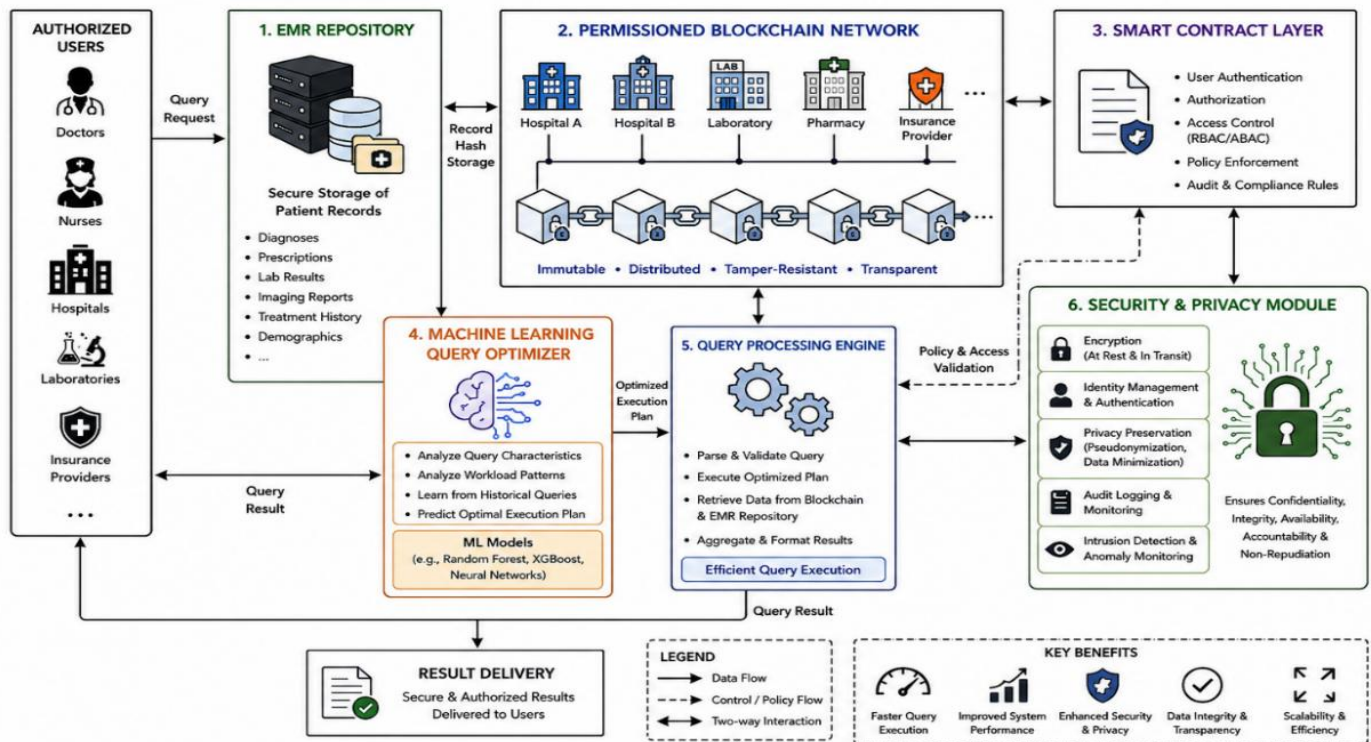


Figure 1: Proposed Framework

3.3 Framework Workflow

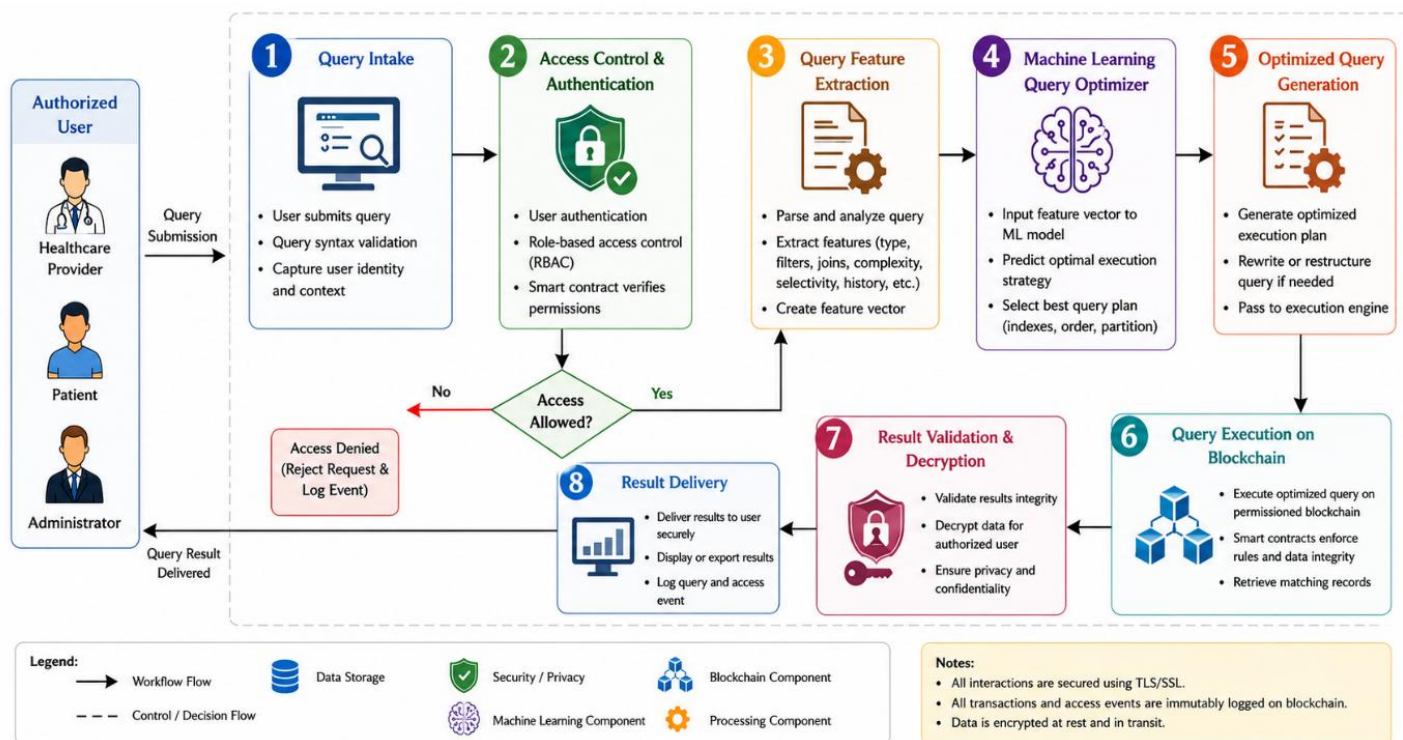


Figure 2: Operational workflow

The operational workflow shown in Figure 2 begins when an authorised healthcare user submits an EMR

query. The query is authenticated through the smart-contract and RBAC mechanisms, and its characteristics are extracted for the Machine Learning Query Optimizer. The Random Forest and XGBoost models analyse the query features and predict an optimal execution plan based on the observed workload. The selected plan is then executed through PostgreSQL and Hyperledger Fabric, subject to the applicable access-control and security mechanisms. AES-256 encryption and SHA-256 hashing support data confidentiality and integrity, while access activities are recorded for audit purposes. The requested records are subsequently retrieved and securely returned to the authorised user. The resulting execution is evaluated using execution time, latency, throughput, resource utilization, query optimization accuracy, and security/privacy performance.

3.4 Machine-Learning-Based Query Optimization: The study employs Random Forest and XGBoost as supervised learning models. Input features include query type, number of tables, joins, predicates, estimated records, query complexity, data volume, transaction workload, and relevant system-resource characteristics. The target variable is the empirically determined optimal execution plan, selected from candidate plans based on the lowest measured execution cost. The 500 selected MIMIC-IV EMR records are divided into 80% training (400 records) and 20% testing (100 records).

Query optimization accuracy is defined as:
$$\text{Accuracy} = \frac{N_{\text{correct plan predictions}}}{N_{\text{test queries}}} \times 100 \dots\dots\dots(1)$$

Model performance is evaluated using accuracy, precision, recall, F1-score, and prediction error where applicable.

3.5 Blockchain and Security Implementation: The framework uses Hyperledger Fabric 3.0 with four peer nodes representing participating healthcare entities. Java-based chaincode implements smart-contract functions and access-control rules. AES-256 encryption, SHA-256 hashing, and Role-Based Access Control (RBAC) are used to protect EMR confidentiality, integrity, and authorized access.

3.6 Experimental Environment: The framework is implemented using Python 3.12, Scikit-learn, and PostgreSQL on Ubuntu 22.04 LTS, using an Intel Core i7 12th-generation processor, 32 GB RAM, and 1 TB NVMe SSD.

3.7 Baseline and Evaluation: The proposed ML optimizer is compared with PostgreSQL's conventional query optimizer under the same workload and experimental conditions. Evaluation considers query execution time, latency, throughput, resource utilization, query optimization accuracy, and security/privacy performance.

Execution-time improvement is calculated as:
$$\text{Improvement (\%)} = \frac{T_b - T_p}{T_b} \times 100 \dots\dots\dots(2)$$

where T_b represents baseline execution time and T_p represents execution time under the proposed framework.

3.8 Experimental Procedure

The experiment used 500 selected MIMIC-IV EMR records, divided into 80% training and 20% testing subsets. Query workloads were generated from the respective subsets, relevant query features were extracted, and candidate execution plans were executed to establish the empirically optimal plans. Random Forest and XGBoost were trained using the training workload, while the test workload was used to evaluate their execution-plan predictions against the established optimal plans. The selected plans were executed within the PostgreSQL–Hyperledger Fabric environment and compared with the conventional PostgreSQL optimizer using execution time, latency, throughput, resource utilization, and query optimization accuracy. Security and privacy were assessed using the implemented AES-256 encryption, SHA-256 hashing, and role-based access control mechanisms. Finally, the framework was evaluated by experts using a five-point Likert-scale instrument covering functionality, security, privacy, usability, and practical applicability.

3.9 Model and Framework Validation

The machine learning models were validated using the test workload by comparing their predicted execution

plans with the empirically determined optimal plans. Model performance was assessed using query optimization accuracy and relevant classification metrics. Separately, the framework was evaluated by 8 purposively selected experts with relevant academic qualifications and at least five years of professional or research experience in healthcare information systems, databases, machine learning, blockchain, cybersecurity, or related fields. A researcher-developed 24-item structured questionnaire was used to evaluate eight dimensions: functionality, security, privacy, scalability, query efficiency, reliability, usability, and overall performance. Each dimension contained three items rated on a five-point Likert scale, ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). Mean scores and standard deviations were used to summarize the responses. The expert evaluation was treated as expert validation of the framework's perceived quality and acceptability rather than as a statistically generalizable survey of healthcare users, and it was not used to determine machine-learning prediction accuracy.

RESULTS AND ANALYSIS

4.1 ML Model Performance: The performance of the Random Forest and XGBoost models was evaluated using accuracy, precision, recall, and F1-score. As shown in Table 4.1, both models achieved strong predictive performance, with XGBoost producing higher values across all four metrics.

Table 4.1. ML Model Performance

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Random Forest	93.2	92.8	92.5	92.6
XGBoost	95.8	95.4	95.6	95.5

XGBoost achieved the highest classification performance, with an accuracy of 95.8%, precision of 95.4%, recall of 95.6%, and F1-score of 95.5%. This indicates that XGBoost provided stronger predictive performance for the execution-plan classification task in the experimental workload.

4.2 Query Performance Comparison: The proposed framework was compared with the conventional PostgreSQL optimizer using execution time, query latency, throughput, CPU utilization, and memory utilization. The results are presented in Table 4.2.

Table 4.2. Query Performance Comparison

Metric	PostgreSQL Baseline	Proposed Framework	Improvement (%)
Execution time (ms)	62.3	36.2	41.9
Query latency (ms)	68.7	41.5	39.6
Throughput (queries/s)	16.1	27.6	71.4
CPU utilization (%)	78.4	61.7	21.3
Memory utilization (%)	72.6	58.9	18.9

The proposed framework reduced mean execution time from 62.3 ms to 36.2 ms, representing a 41.9% improvement. Query latency also decreased by 39.6%, while throughput increased from 16.1 to 27.6 queries/s, representing a 71.4% increase. CPU and memory utilization were reduced by 21.3% and 18.9%, respectively. These results indicate that ML-driven execution-plan selection improved query-processing efficiency relative to the conventional PostgreSQL optimizer.

4.3 Query Optimization Accuracy

Query optimization accuracy as shown in Table 4.3, was evaluated by comparing the execution plan selected by the framework with the empirically determined optimal plan for each test query. The XGBoost model correctly selected the empirically optimal execution plan for 94 of the 100 test queries, resulting in an end-to-end query optimization accuracy of 94.0%.

Table 4.3. Query Optimization Accuracy

Measure	Result
Test queries	100
Correctly predicted optimal plans	94
Incorrect predictions	06
Query optimization accuracy	94.0%

4.4 Security and Privacy Evaluation

The security and privacy mechanisms incorporated into the framework were evaluated based on the implemented encryption, hashing, access-control, smart-contract authorization, and blockchain audit mechanisms as shown in Table 4.4.

Table 4.4. Security and Privacy Evaluation

Criterion	Mechanism	Result
Confidentiality	AES-256	Protected
Data integrity	SHA-256	Verified
Access control	RBAC	Enforced
Smart-contract authorization	Java chaincode	Enforced
Auditability	Blockchain transaction records	Maintained

The framework employed AES-256 to protect sensitive EMR data, while SHA-256 supported data-integrity verification. RBAC restricted access according to authorized roles, and Java-based chaincode enforced smart-contract authorization. Blockchain transaction records also provided an auditable record of transactions. Together, these mechanisms provided the security and access-control foundation for the proposed EMR management framework.

4.5 Expert Evaluation of the Framework

The proposed framework was evaluated by 8 experts, as shown in Table 4.5, using a five-point Likert-scale instrument covering functionality, security, privacy, scalability, query efficiency, reliability, usability, and overall performance.

Table 4.5. Expert Evaluation Results

Criterion	Mean	SD
Functionality	4.23	0.44
Security	4.16	0.32
Privacy	3.81	0.35
Scalability	3.96	0.42
Query Efficiency	4.28	0.40
Reliability	3.75	0.41
Usability	3.91	0.46
Overall Performance	4.01	0.40

The results show generally positive assessments, with mean scores ranging from 3.75 to 4.28. Query efficiency received the highest rating ($M = 4.28$, $SD = 0.40$), followed by functionality ($M = 4.23$, $SD = 0.44$) and security ($M = 4.16$, $SD = 0.32$). Reliability recorded the lowest mean ($M = 3.75$, $SD = 0.41$), while overall performance was rated 4.01 ($SD = 0.40$). The 8 experts perceived the framework as functional, secure, efficient, and usable, while the comparatively lower ratings for reliability and privacy indicate areas for further refinement. These results reflect perceived framework quality and acceptability rather than machine-learning prediction accuracy.

DISCUSSION

The findings demonstrate that the proposed framework improved query processing for blockchain-based Electronic Medical Record (EMR) management. XGBoost achieved 95.8% classification accuracy, with precision, recall, and F1-score of 95.4%, 95.6%, and 95.5%, respectively, indicating strong performance in predicting execution plans. However, the 94.0% query optimization accuracy shows that the model correctly selected the empirically optimal execution plan for 94 of the 100 test queries. The difference between these measures highlights the distinction between model classification performance and actual end-to-end optimal-plan selection.

The performance comparison further shows that the proposed framework reduced execution time from 62.3 ms to 36.2 ms and latency from 68.7 ms to 41.5 ms, while increasing throughput from 16.1 to 27.6 queries/s. CPU and memory utilization were also reduced. These results indicate that workload-aware execution-plan selection can improve query-processing efficiency compared with the conventional PostgreSQL optimizer.

The security mechanisms, comprising AES-256 encryption, SHA-256 hashing, role-based access control, smart contracts, and Hyperledger Fabric, provide the framework with mechanisms for confidentiality, integrity, controlled access, and transaction traceability. The expert evaluation supports these technical findings, with high mean ratings for Query Efficiency (4.28), Functionality (4.23), and Security (4.16). The overall performance rating of 4.01 also indicates positive expert perception of the framework. Lower ratings for reliability (3.75) and privacy (3.81) identify areas requiring further refinement.

These findings, therefore suggest that integrating machine learning-based query optimization with permissioned blockchain and security mechanisms can improve both the processing efficiency and controlled management of EMR queries. The results also show that expert perception, ML classification performance, and end-to-end query optimization are distinct dimensions of evaluation and should be interpreted separately.

CONCLUSION

This study proposed a Machine Learning-Driven Query Optimization Framework for secure and efficient blockchain-based Electronic Medical Record (EMR) management. The framework integrates machine learning, permissioned blockchain, smart contracts, and security mechanisms to improve query processing while supporting the confidentiality, integrity, and controlled access of medical records. The experimental results showed that XGBoost achieved 95.8% classification accuracy, while correctly selecting the empirically optimal execution plan for 94 of 100 test queries, resulting in 94.0% end-to-end query optimization accuracy. Compared with the conventional PostgreSQL optimizer, the proposed framework also achieved lower execution time and latency, higher throughput, and reduced resource utilization. Expert evaluation further produced positive assessments, with high ratings for query efficiency, functionality, and security and an overall performance mean of 4.01. These findings indicate that integrating machine learning-driven query optimization with permissioned blockchain can improve EMR query-processing efficiency while maintaining security and controlled access. The framework provides a practical foundation for further refinement and validation using larger datasets, diverse workloads, and real-world healthcare environments.

RECOMMENDATIONS

Healthcare institutions should consider adopting machine learning-driven query optimization and permissioned blockchain to improve EMR query efficiency, security, and controlled access. Smart contracts, encryption, hashing, and RBAC should be integrated to strengthen data protection, while reliability and privacy should receive further attention.

FUTURE WORK

Future research should evaluate the framework using larger datasets, diverse workloads, and real-world healthcare environments. Further studies should explore advanced ML models, XAI, and federated learning to improve optimization accuracy, transparency, privacy, and scalability.

REFERENCES

1. Agbo, C. C., Mahmoud, Q. H., & Eklund, J. M. (2019). Blockchain technology in healthcare: A systematic review. *Healthcare*, 7(2), Article 56. <https://doi.org/10.3390/healthcare7020056>
2. Azaria, A., Ekblaw, A., Vieira, T., & Lippman, A. (2016). MedRec: Using blockchain for medical data access and permission management. In *Proceedings of the 2016 IEEE 2nd International Conference on Open and Big Data (OBD)* (pp. 25–30). IEEE. <https://doi.org/10.1109/OBD.2016.11>
3. Casino, F., Dasaklis, T. K., & Patsakis, C. (2019). A systematic literature review of blockchain-based applications: Current status, classification and open issues. *Telematics and Informatics*, 36, 55–81. <https://doi.org/10.1016/j.tele.2018.11.006>
4. Chamola, V., Goyal, A., Sharma, P., Hassija, V., Binh, H. T. T., & Saxena, V. (2022). Artificial intelligence-assisted blockchain-based framework for smart and secure electronic medical record management. *Neural Computing and Applications*, 34(24), 22959–22969. <https://doi.org/10.1007/s00521-022-07087-7>
5. Deshpande, A., Ives, Z. G., & Raman, V. (2007). Adaptive query processing. *Foundations and Trends® in Databases*, 1(1), 1–140. <https://doi.org/10.1561/19000000001>
6. Dinh, T. T. A., Wang, J., Chen, G., Liu, R., Ooi, B. C., & Tan, K.-L. (2018). BLOCKBENCH: Benchmarking blockchain systems. In *Proceedings of the 2017 ACM International Conference on Management of Data* (pp. 1085–1100). Association for Computing Machinery. <https://doi.org/10.1145/3035918.3064033>
7. Gan, C., Yang, H., Zhu, Q., Zhang, Y., & Saini, A. (2023). An encrypted medical blockchain data search method with access control mechanism. *Information Processing & Management*, 60(6), 103499. <https://doi.org/10.1016/j.ipm.2023.103499>
8. Graefe, G. (1993). Query evaluation techniques for large databases. *ACM Computing Surveys*, 25(2), 73–169. <https://doi.org/10.1145/152610.152611>
9. Hassija, V., Ratnakumar, R., Chamola, V., Agarwal, S., Mehra, A., Kanhere, S. S., & Binh, H. T. T. (2022). A machine learning and blockchain-based secure and cost-effective framework for minor medical consultations. *Sustainable Computing: Informatics and Systems*, 35, 100651. <https://doi.org/10.1016/j.suscom.2021.100651>
10. Office of the National Coordinator for Health Information Technology. (2023). *What is an electronic health record (EHR)?* <https://www.healthit.gov/faq/what-electronic-health-record-ehr>
11. Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260. <https://doi.org/10.1126/science.aaa8415>
12. Kruse, C. S., Kristof, C., Jones, B., Mitchell, E., & Martinez, A. (2018). Barriers to electronic health record adoption: A systematic literature review. *Journal of Medical Systems*, 42(12), Article 246. <https://doi.org/10.1007/s10916-018-1078-6>
13. Kuo, T.-T., Kim, H.-E., & Ohno-Machado, L. (2017). Blockchain distributed ledger technologies for biomedical and healthcare applications. *Journal of the American Medical Informatics Association*, 24(6), 1211–1220. <https://doi.org/10.1093/jamia/ocx068>
14. Marcus, R., & Papaemmanouil, O. (2019). Deep reinforcement learning for join order enumeration. In *Proceedings of the First International Workshop on Exploiting Artificial Intelligence Techniques for Data Management* (pp. 1–4). Association for Computing Machinery. <https://doi.org/10.1145/3329785.3329929>
15. Marcus, R., Negi, P., Mao, H., Zhang, C., Alizadeh, M., Kraska, T., Papaemmanouil, O., & Tatbul, N. (2019). Neo: A learned query optimizer. *Proceedings of the VLDB Endowment*, 12(11), 1705–1718. <https://doi.org/10.14778/3342263.3342644>
16. McGhin, T., Choo, K.-K. R., Liu, C. Z., & He, D. (2019). Blockchain in healthcare applications: Research challenges and opportunities. *Journal of Network and Computer Applications*, 135, 62–75. <https://doi.org/10.1016/j.jnca.2019.02.027>
17. Nakamoto, S. (2008). *Bitcoin: A peer-to-peer electronic cash system*. <https://bitcoin.org/bitcoin.pdf>
17. Palaiokrassas, G., Bouraga, S., & Tassioulas, L. (2024). Machine learning on blockchain data: A systematic mapping study. *arXiv*. <https://doi.org/10.48550/arXiv.2403.17081>
18. Przytarski, D., Stach, C., Gritti, C., & Mitschang, B. (2022). Query processing in blockchain systems: Current state and future challenges. *Future Internet*, 14(1), 1. <https://doi.org/10.3390/fi14010001>

19. Silberschatz, A., Korth, H. F., & Sudarshan, S. (2020). *Database system concepts* (7th ed.). McGraw-Hill.
20. Vaidya, K., Dutt, A., Narasayya, V., & Chaudhuri, S. (2022). Leveraging query logs and machine learning for parametric query optimization. *Proceedings of the VLDB Endowment*, 15(3), 401–413. <https://doi.org/10.14778/3494124.3494126>
21. World Health Organization. (2021). *Global strategy on digital health 2020–2025*. World Health Organization. <https://www.who.int/publications/i/item/9789240020924>
22. Xu, X., Weber, I., & Staples, M. (2021). *Architecture for blockchain applications*. Springer.
23. Zhou, Z., Marcus, R., Mao, Y., & Papaemmanouil, O. (2022). AI meets database systems: AI-driven query optimization and database tuning. *Proceedings of the VLDB Endowment*, 15(12), 3666–3669.
24. Zhou, E., Hong, Z., Xiao, Y., Zhao, D., Pei, Q., Guo, S., & Akerkar, R. (2023). MSTDB: A hybrid storage-empowered scalable semantic blockchain database. *IEEE Transactions on Knowledge and Data Engineering*, 35(8), 8228–8244. <https://doi.org/10.1109/TKDE.2022.3220522>

APPENDIX A

Expert Evaluation Questionnaire

Title:

Expert Evaluation Questionnaire for the Machine Learning-Driven Query Optimization Framework for Secure and Efficient Blockchain-Based EMR Management

Instructions:

Please indicate your level of agreement with each statement by ticking (✓) one option for each item.

Likert Scale:

SD = Strongly Disagree (1) | D = Disagree (2) | N = Neutral (3) | A = Agree (4) | SA = Strongly Agree (5)

No.	Evaluation Item	SD (1)	D (2)	N (3)	A (4)	SA (5)
	Functionality					
1	The framework performs its intended EMR query-processing functions effectively.	☐	☐	☐	☐	☐
2	The machine learning component effectively supports execution-plan selection.	☐	☐	☐	☐	☐
3	The framework integrates its major components effectively for EMR management.	☐	☐	☐	☐	☐
	Security					
4	The framework provides effective protection against unauthorized access to EMR data.	☐	☐	☐	☐	☐
5	The security mechanisms adequately protect the confidentiality and integrity of EMR data.	☐	☐	☐	☐	☐
6	The blockchain and smart-contract mechanisms provide effective access control and transaction security.	☐	☐	☐	☐	☐
	Privacy					
7	The framework provides adequate protection of patient information during data processing.	☐	☐	☐	☐	☐
8	The framework adequately limits access to EMR data based on authorized user roles.	☐	☐	☐	☐	☐
9	The framework provides appropriate mechanisms for protecting sensitive patient information.	☐	☐	☐	☐	☐
	Scalability					
10	The framework can accommodate increasing numbers of EMR	☐	☐	☐	☐	☐

	queries and users.					
11	The framework can support increasing volumes of healthcare data without substantial performance degradation.	☐	☐	☐	☐	☐
12	The framework can be extended to support additional healthcare institutions or blockchain nodes.	☐	☐	☐	☐	☐
	Query Efficiency					
13	The framework improves the efficiency of EMR query processing.	☐	☐	☐	☐	☐
14	The machine learning-based optimizer effectively supports the selection of efficient execution plans.	☐	☐	☐	☐	☐
15	The framework provides timely retrieval of requested EMR information.	☐	☐	☐	☐	☐
	Reliability					
16	The framework provides consistent query-processing performance.	☐	☐	☐	☐	☐
17	The framework can reliably process authorized EMR queries without unnecessary failures.	☐	☐	☐	☐	☐
18	The framework maintains dependable operation during varying query workloads.	☐	☐	☐	☐	☐
	Usability					
19	The framework is easy for authorized healthcare users to use.	☐	☐	☐	☐	☐
20	The framework provides a clear and understandable query-processing process.	☐	☐	☐	☐	☐
21	The framework can be used effectively without extensive technical support.	☐	☐	☐	☐	☐
	Overall Performance					
22	The framework provides an effective overall solution for blockchain-based EMR management.	☐	☐	☐	☐	☐
23	The integration of machine learning and blockchain provides meaningful performance benefits.	☐	☐	☐	☐	☐
24	The framework is suitable for practical application in healthcare information management.	☐	☐	☐	☐	☐

Respondent Information

Area of Expertise: _____

Highest Qualification: _____

Years of Relevant Experience: _____

Professional/Research Area: _____

Signature (Optional): _____

Date: _____