HealthChain: A Hybrid Blockchain for Scalable and Secure Healthcare Data Management

Ahmad Mousa Altamimi
Princess Sumaya University for Technology, Amman, Jordan
a.altamimi@psut.edu.jo (corresponding author)

Lamia Al-Kershi
Princess Sumaya University for Technology, Amman, Jordan
lam20228026@std.psut.edu.jo

Qusay Abdo
Princess Sumaya University for Technology, Amman, Jordan
qus20200963@std.psut.edu.jo

Received: 3 February 2026 | Revised: 21 February 2026 and 13 March 2026 | Accepted: 15 March 2026

 DOI: https://doi.org/10.48084/etasr.17941

ABSTRACT

Blockchain technology has emerged as a promising foundation for secure, transparent, and patient-centric healthcare data management; however, the scalability, latency, and data storage limitations of conventional blockchains hinder their adoption in data-intensive healthcare environments involving Electronic Health Records (EHRs), medical imaging, and continuous sensor data streams. To address these limitations, sidechain architectures have been proposed as a viable solution, enabling off-chain processing and storage while maintaining cryptographic anchoring to a secure main blockchain. In this context, the paper introduces HealthChain, a formally specified sidechain-based architecture for scalable, privacy-preserving healthcare data management. Unlike prior work that remains largely conceptual, HealthChain provides an explicit definition of system components, cross-chain interaction mechanisms, data management policies, and governance responsibilities. Furthermore, the proposed architecture separates governance and auditability functions, maintained on a permissioned main blockchain, from high-throughput healthcare data processing, delegated to domain-specific sidechains. This design enables efficient handling of large-scale healthcare data while preserving security and compliance requirements. The proposed model is evaluated using blockchain-relevant performance metrics, including transaction throughput, latency, storage overhead, and cross-chain communication cost. Comparative analysis demonstrates that HealthChain significantly improves scalability and efficiency over a traditional main-chain-only approach, while preserving decentralization, data integrity, and regulatory compliance. Overall, the results highlight the potential of sidechain-based architectures to enable practical, large-scale deployment of blockchain technologies in healthcare systems.

Keywords-blockchain; sidechains; healthcare data management; scalability; privacy; interoperability

I. INTRODUCTION

The healthcare sector is undergoing rapid digital transformation, driven by the growth of data generated from Electronic Health Records (EHRs), medical imaging, genomics, wearable devices, and clinical trials [1]. These technologies produce large volumes of heterogeneous and highly sensitive data, creating significant challenges in data management, secure sharing, and interoperability across healthcare providers and patients [2]. Meanwhile, traditional healthcare Information Technology (IT) infrastructures are predominantly centralized and proprietary, limiting seamless data exchange and increasing risks related to data breaches and privacy violations [3].

Blockchain technology offers an alternative paradigm by enabling decentralized data integrity, tamper resistance, and transparent auditability without reliance on a single trusted intermediary. Moreover, through smart contracts, blockchain systems can enforce automated workflows and data-sharing policies, making them suitable for applications such as secure EHR exchange and pharmaceutical supply chain management [4]. However, direct deployment of healthcare applications on conventional blockchain platforms remains impractical. This is because both public and permissioned blockchains (e.g., Ethereum) suffer from inherent limitations in transaction throughput, latency, and on-chain storage capacity, rendering them unsuitable for the scale and complexity of healthcare data [5]. Specifically, the main challenge of building directly on top

of a public blockchain is the scalability issue that arises when handling large volumes of transactions.

To mitigate these limitations, several off-chain scaling techniques have been proposed, most notably sharding and sidechains [7]. Sharding partitions the blockchain network into smaller subsets ("shards") that process transactions in parallel, thereby improving throughput and scalability [8]. However, sharding requires careful consideration of security, data consistency, and compatibility issues, making it complex to implement [9]. In contrast, sidechains provide independent blockchain environments that are cryptographically linked to a main chain through two-way pegs, enabling secure transfer of assets and state commitments between chains [10]. In a typical two-way peg mechanism, assets transferred from the main chain are locked, and corresponding representations are minted on the sidechain, and vice versa during reverse transfer [11]. This mechanism enables interoperability, flexibility, and customizable execution environments, allowing sidechains to be optimized for specific application domains. This architectural innovation supports off-chain transaction processing, enhancing scalability, and enables transactions faster and at a lower cost without requiring access to the main blockchain [12].

The application of sidechains in healthcare domains such as Personal Health Record (PHR) management, telemedicine platforms, and supply-chain traceability has been explored [13, 14]. While these works demonstrate the potential of sidechain-based solutions, many remain conceptual, lack formal system specification, or do not provide reproducible performance evaluation at the blockchain level. Moreover, prior studies often fail to clearly delineate governance, storage, and computation responsibilities across blockchain layers.

To address these gaps, this paper advances the state of the art in sidechain-based healthcare blockchain systems through four key contributions. First, it proposes a formally defined hybrid blockchain architecture that separates governance, identity management, and auditability functions on a permissioned consortium main chain, while delegating high-throughput data processing to domain-specific healthcare sidechains. Second, it introduces a healthcare-oriented cross-chain protocol, including a reproducible operational workflow for secure data ingestion, policy anchoring, and integrity verification. Third, it develops a storage complexity model tailored to large-scale healthcare datasets, including EHRs and medical imaging, supported by quantitative storage overhead analysis. Finally, it presents a reproducible experimental evaluation, including system deployment details, scalability benchmarking, and comparison against an empirically implemented main-chain-only baseline.

To validate the proposed approach, the system has been implemented using Hyperledger Fabric as the healthcare sidechain platform, enabling secure and low-latency transaction processing, while a Quorum-based Ethereum network serves as the main blockchain for anchoring cryptographic state commitments generated by the sidechain.

II. SIDECHAIN

Sidechain architecture extends traditional blockchain design by introducing independent chains that are cryptographically linked to a main blockchain through a two-way peg, as illustrated in Figure 1. A key advantage of sidechains lies in their ability to operate under customized configurations, including independent consensus mechanisms, block parameters, and access control models. This flexibility enables higher throughput and lower latency compared to the main chain, making sidechains suitable for data-intensive applications. Additionally, sidechains offload computation and storage from the main blockchain while preserving security guarantees through periodic cryptographic anchoring. Sidechains can also facilitate interoperability across blockchain networks through smart contracts and interchain communication protocols. These mechanisms enable the creation of cross-chain bridges, supporting secure data exchange, transaction execution, and coordinated operations across multiple blockchain environments [15]. Such protocols allow chains to trigger cross-chain events, share state information, and execute complex workflows spanning multiple networks.

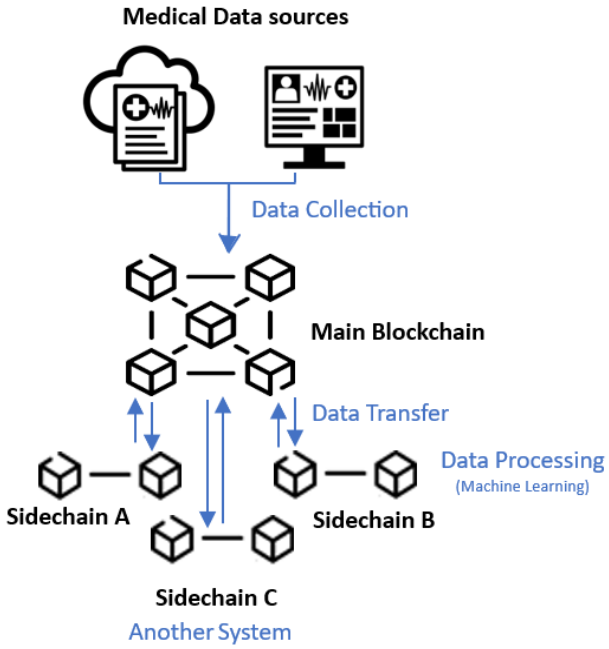

Fig. 1. The sidechain architecture.

In healthcare contexts, sidechains enable domain-specific optimization, such as privacy-aware data access, large data handling, and regulatory compliance, without overburdening the main blockchain. This allows developers to build decentralized applications that can leverage the security and trust of the main blockchain while still having their own separate blockchain. When designed well, a sidechain can operate smoothly without compromising the performance of the primary network, which means its architecture, consensus method, and other components must be carefully tuned [16].

In fact, blockchain adoption in healthcare has been widely studied, with applications ranging from EHR sharing and clinical trials to pharmaceutical supply chains. These studies demonstrate blockchain's potential to ensure data integrity, provenance, and patient-centric access control; yet, they also highlight persistent challenges related to scalability and privacy [17]. To address these limitations, the use of sidechains in healthcare systems has been investigated [10], [18].

Sidechain-based healthcare solutions have been proposed for PHR management, telemedicine, and decentralized identity systems. For example, authors in [18] explored sidechain integration in PHR systems, demonstrating improvements in security, privacy, and interoperability, while also identifying challenges related to data quality, governance, and user adoption. Similarly, authors in [10, 16, 19] examined blockchain-based healthcare data exchange frameworks and reported enhanced interoperability and security, alongside unresolved issues in regulatory compliance, scalability, and governance. Additional studies focusing on EHR systems [20] and telemedicine platforms [12] further confirmed that, while sidechains offer architectural advantages, challenges such as data governance, system scalability, and user acceptance are significant barriers.

Beyond academic research, several real-world implementations demonstrate the applicability of sidechain-enabled healthcare systems. Platforms such as IBM Blockchain, Medicalchain, and Patientory leverage blockchain technologies to enable secure and patient-centric health data management, while solutions like Chronicled (MediLedger) apply blockchain architectures to pharmaceutical supply chain traceability [21].

More recent efforts have explored the integration of blockchain with Artificial Intelligence (AI), where sidechains are used to support scalable and privacy-preserving data pipelines for predictive analytics and personalized medicine [22]. While these hybrid approaches enable advanced healthcare services, they often conflate data analytics performance with blockchain system performance, making it difficult to isolate and evaluate the contribution of the underlying blockchain architecture.

Overall, existing literature indicates that sidechains can offer significant benefits in terms of scalability, privacy, and interoperability in healthcare systems; however, several open challenges remain. In particular, the lack of standardized cross-chain communication protocols limits seamless interoperability across heterogeneous blockchain environments. Furthermore, the diversity of healthcare regulatory frameworks all over the world necessitates adaptable mechanisms for data sovereignty and compliance enforcement [23]. Additional concerns include fault tolerance, privacy preservation, data provenance, and system-level governance, which continue to be active areas of research [24].

III. HEALTHCHAIN'S METHODOLOGY

This study starts by conducting an intensive literature review to explore the applications and advancements of sidechain blockchain technology in healthcare. The collected studies were subjected to thematic analysis to identify prevailing trends, limitations, research gaps, and innovation.

Based on these findings, the HealthChain framework is designed as a sidechain-enabled architecture and evaluated along four principal dimensions: scalability, privacy preservation, interoperability, and deployment feasibility. Additionally, the study incorporates a focused analysis of AI integration within blockchain-based healthcare systems, along with associated security and trust considerations.

A. HealthChain Architecture

HealthChain adopts a two-layer blockchain architecture, as demonstrated in Figure 2, consisting of a permissioned consortium main blockchain and a healthcare-optimized sidechain:

- The main blockchain (Layer 1) is responsible for identity management, access control enforcement, audit logging, and cryptographic anchoring of sidechain states. It operates under a permissioned consortium model to ensure governance, accountability, and regulatory compliance.
- The healthcare sidechains (Layer 2) are domain-specific blockchain environments optimized for high-throughput processing and large-scale data reference management, supporting domains such as EHRs, medical imaging, and Internet of Medical Things (IoMT) systems.

Additionally, a federated two-way peg mechanism is employed to enable secure interaction between layers. This mechanism is implemented using a combination of multi-signature authorization, Hash Time-Locked Contracts (HTLCs), nonce-based replay protection, and Merkle-proof verification. A fully trustless peg is intentionally not adopted due to healthcare regulatory frameworks requiring identifiable governance and institutional accountability.

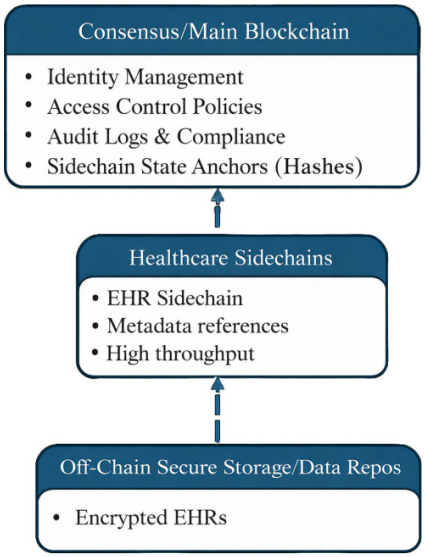

Fig. 2. HealthChain layered architecture.

The architectural separation ensures that governance and compliance functions remain centralized and strongly

consistent, while data-intensive operations are executed on scalable sidechains. Meanwhile, large healthcare datasets are stored off-chain and referenced through encrypted pointers anchored to the main blockchain, minimizing storage overhead while preserving verifiability.

The design is motivated by non-functional system constraints, as a single-layer blockchain cannot simultaneously satisfy the requirements of strong consistency, high throughput, low latency, and regulatory control. Thus, the two-layer blockchain is deliberately separated to address different requirements that are difficult to satisfy simultaneously on a single ledger. Specifically, the main chain prioritizes immutability and trust guarantees, whereas the sidechains prioritize performance and flexibility.

To ensure robustness, the system incorporates several consistency and fault-tolerance mechanisms, including: i) periodic anchoring that stores sidechain block hashes on the main chain, creating an immutable checkpoint, ii) Merkle-proof verification that ensures that any sidechain state corresponds to the anchored root, and iii) rollback protection that leverages main-chain finality to prevent long-range attacks, while failure-recovery procedures validate sidechain state against the latest anchor.

B. Data Storage and Capacity Model

HealthChain adopts a reference-based storage model to minimize on-chain data overhead and enable faster data preprocessing. Instead of storing raw healthcare data, the system stores cryptographic hashes and encrypted references, while the actual data reside in off-chain distributed storage systems. This design choice directly impacts transaction throughput and latency by minimizing block size growth and state replication costs. The main blockchain stores only policy-level metadata and periodic state anchors, ensuring that governance operations remain lightweight and scalable. Consequently, the overall system achieves improved performance compared to monolithic blockchain designs that attempt to store or manage large data objects directly on-chain. Figure 3 shows the data storage model. Let:

- $M = \{m_1, m_2, \dots, m_k\}$ denote main-chain validator nodes.
- $S = \{s_1, s_2, \dots, s_n\}$ sidechain validator nodes.
- D denote healthcare data.
- $h = H(D)$ denote the cryptographic hash of data.
- P denote access-control policy.
- $A = Hash(Block\_sidechain)$ denote sidechain state anchor.

Each transaction is defined as:

$$T = (h, P, timestamp, signature) \quad (1)$$

Main-chain storage complexity is expressed as $O(|H(D)|)$, while sidechain storage scales are expressed as $O(|D|)$. This design enables efficient support for large EHR files and medical imaging datasets without violating blockchain storage constraints. Furthermore, this supports the following security properties:

- Integrity: Hash anchoring enables tamper detection.
- Consistency: Sidechain block hashes are anchored periodically.
- Authorization: Smart-contract policy enforcement.

As future formal verification, it is planned to express HealthChain in Temporal Logic of Actions (TLA+) for state-machine verification and ProVerif for cryptographic protocol analysis.

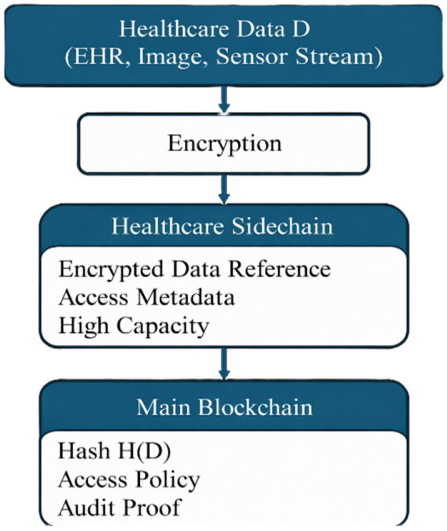

Fig. 3. HealthChain data storage.

C. Operational Workflow

Algorithm 1 formalizes the end-to-end operational workflow of the proposed HealthChain architecture for secure, scalable, and auditable healthcare data management.

Algorithm 1: HealthChain Data Management Workflow

```
Input: Patient data D, Access policy P
Output: Anchored healthcare record
1: Store D in sidechain S
2: Compute h ← Hash(D)
3: Submit (h, P) to main chain
4: Validate via consortium consensus
5: Anchor sidechain block hash to main chain
6: Enable authorized access via smart contract
```

In HealthChain, raw healthcare data are first encrypted using symmetric cryptography, ensuring that no sensitive plaintext is exposed on-chain. The system stores only secure references (pointers) to the encrypted data within the sidechain. In Step 2, a cryptographic hash $h = H(D)$ is computed, serving as a compact and immutable fingerprint of the data, enabling future integrity verification without requiring on-chain storage

of large healthcare files. This design also supports General Data Protection Regulation (GDPR) compliance, including the right to erasure, since only hashes are stored on-chain while actual data can be deleted off-chain through data removal and key revocation.

In Step 3, the tuple (h, P) is submitted to the consortium's main blockchain, where P represents the access control policies (e.g., roles, identities, or consent conditions) that may access or decrypt the underlying healthcare data. In Step 4, submitted transactions are validated using a permissioned consensus protocol (e.g., Raft), providing low latency and deterministic finality, which are essential in healthcare systems. In Step 5, the corresponding sidechain block hash is anchored to the main blockchain, establishing a cryptographic linkage that prevents unauthorized modification of sidechain records. Finally, in Step 6, access to encrypted healthcare data is granted through smart contracts, which enforce policy P, generate audit logs, and ensure that all data access events are traceable and non-repudiable.

Additionally, once the data are transferred to the sidechain, machine learning algorithms or other data processing tools would be used to analyze the data and identify patterns that may indicate a potential health issue. Processed outputs (e.g., predictions or summaries) may then be stored on the sidechain or selectively anchored to the main chain, enabling advanced healthcare analytics while preserving system scalability and auditability.

D. Trust Model and Threat Model

HealthChain operates under a permissioned consortium trust model, where validators are independent healthcare stakeholders (e.g., hospitals, regulators, insurers). Membership is governed through a Public Key Infrastructure (PKI)-based identity management system, ensuring authorized access.

In the adversary model, this work considers several potential threats: i) insider attackers within participating institutions, ii) network-level attackers attempting to disrupt or intercept communications, and iii) the possibility of sidechain compromise.

To mitigate these risks, HealthChain employs quorum-based consensus, distributes validator authority across multiple institutions, and anchors audit records to a main chain. Cross-chain anchoring further enhances integrity by preventing undetected tampering with transaction history.

E. Evaluation Criteria

To rigorously evaluate the effectiveness of the proposed HealthChain architecture, blockchain-specific performance metrics are employed. These metrics are selected to directly reflect the system's core design objectives, namely scalability, efficiency, and suitability for healthcare data management.

- Transaction throughput is the number of confirmed transactions N per unit time:

$$\text{Throughput} = \frac{N}{T}(\text{tx/s}) \quad (2)$$

where tx refers to a transaction. This metric captures the effectiveness of the sidechain architecture in supporting high transaction volumes relative to the baseline.

- Latency measures the time required for a transaction to progress from submission to final confirmation. For a transaction i, latency is defined as:

$$L_i = t_i^{\text{confirm}} - t_i^{\text{submit}} \quad (3)$$

where t_i^{submit} denotes the submission time and t_i^{confirm} denotes the time at which the transaction becomes irreversible. The average latency over all transactions is calculated by:

$$\bar{L} = \frac{1}{N}\sum_{i=1}^{N} L_i \quad (4)$$

This metric evaluates the responsiveness of the system and the performance impact of consensus and cross-chain anchoring.

- Storage overhead measures the efficiency of on-chain storage relative to the original healthcare data size, where S_{on} denote the total amount of data stored on-chain, and S_{raw} denote the size of the original healthcare data:

$$\text{Storage Overhead} = \frac{S_{\text{on}}}{S_{\text{raw}}} \quad (5)$$

A lower ratio indicates more efficient storage utilization.

Secondly, to evaluate the sidechain proposed approach, a set of machine learning classifiers was applied to the dataset. The models used were a Decision Tree classifier [25], Random Forest classifier [26], CatBoost classifier [27], XGBoost classifier [28], and an Ensemble Voting classifier [29], which combined their predictions. These models were chosen for their proven effectiveness in classification tasks and their ability to capture both simple and complex data patterns.

The performance of the models was assessed using widely adopted evaluation metrics, including accuracy, precision, recall, and F1-score. These metrics are calculated using [30]:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (6)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (7)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (8)$$

$$\text{F1} - \text{score} = 2\frac{Precision \cdot Recall}{Precision+Recall} \quad (9)$$

where:

- True Positive (TP) denotes correctly predicted positive instances.
- False Positive (FP) denotes incorrectly predicted positive instances.
- True Negative (TN) denotes correctly predicted negative instances.
- False Negative (FN) denotes incorrectly predicted negative instances.

These metrics provide a comprehensive view of the models' predictive capabilities, particularly in imbalanced datasets where accuracy alone is insufficient. The evaluation was conducted on a separate test set to ensure an unbiased estimate of model performance.

IV. EXPERIMENTAL RESULTS

A. Dataset Description

The experimental evaluation was conducted using the publicly available Medical Student Dataset [31]. The dataset consists of 180,000 records with 13 attributes capturing demographic, anthropometric, physiological, and lifestyle information. During preprocessing, records with null values across all features were excluded, reducing the dataset by 6,489 records. The final distribution was 88,933 non-diabetes versus 84,578 diabetes cases, of which 80% was used for training and 20% for testing. The key attributes of each case included:

- Student ID (unique identifier).
- Age (in years).
- Gender (male/female).
- Height (in cm).
- Weight (in kg).
- Blood Type (categorical: A, B, AB, O).
- Body Mass Index (BMI) (calculated from height and weight).
- Temperature (body temperature, °C).
- Heart Rate (beats per min).
- Blood Pressure (mmHg).
- Cholesterol (mg/dL).
- Diabetes (binary: Yes/No).
- Smoking (binary: Yes/No).

B. Experimental Setup

To evaluate the performance of the proposed HealthChain architecture, a prototype system was implemented using a hybrid blockchain environment consisting of a permissioned healthcare sidechain and a consortium-based main blockchain. The experimental setup was designed to reflect realistic healthcare data management requirements while enabling controlled performance measurements.

The healthcare sidechain was implemented using Hyperledger Fabric, selected for its permissioned access model, low-latency transaction processing, and fine-grained data confidentiality mechanisms. Fabric smart contracts (chaincode) were deployed to manage patient identifiers, transaction verification, synchronization across contracts, and block creation. The experimental deployment consisted of four interconnected smart contracts, reflecting a multi-department healthcare environment.

The main blockchain served as an anchoring layer and was conceptually modeled using a permissioned Ethereum-based network. The main chain was responsible only for storing cryptographic hashes and state commitments generated by the healthcare sidechain, thereby ensuring data integrity and auditability without storing raw medical data on-chain.

The deployment environment consisted of six Hyperledger Fabric peers, three orderer nodes, and four Quorum nodes, running on Intel i7 systems with 16 GB Random Access Memory (RAM) and the Ubuntu 22.04 operating system. External validity is a limitation of the current study, and future work will incorporate anonymized clinical datasets to improve real-world applicability.

C. Comparative Evaluation

A comparative evaluation is first performed between HealthChain and the conventional main chain design. The proposed model operates through multiple sequential stages, each contributing to the total processing time.

Initially, a Universally Unique Identifier (UUID) is generated for each patient upon admission to ensure privacy. The generation and duplication check require approximately 0.0000005 s, which is effectively negligible. After patient data entry, the block is finalized and broadcast to all contracts in the system. In the current configuration (four contracts), the average time required for block closure and dissemination is 0.0325 s. Upon receipt, several processes are executed, including transaction verification, synchronization, and block creation. Verification of whether a transaction has been tampered with or not requires 0.1248 s, synchronization across contracts requires 0.4357 s, block creation (processing two transactions) requires 0.0124 s, and synchronization of a tampered block requires 1.5218 s.

1) Latency

For the main-chain-only system, global consensus and sequential block validation introduce significant delays, leading to an average latency of about 10 s. In contrast, HealthChain reduces confirmation time by offloading validation and synchronization to sidechains. Using measured verification, synchronization, block creation, and anchoring times, the average transaction latency of HealthChain is:

$$LHC = 0.0325 + 0.1248 + 0.4357 + 0.0124 + 1.5218$$

$$LHC = 2.13\ s$$

2) Throughput

In the main-chain-only architecture, all transactions compete for a single consensus process, resulting in a low throughput of approximately 0.1 tx/s. In contrast, HealthChain distributes transaction processing across multiple contracts and sidechains, achieving a throughput of 0.47 tx/s, with the potential for linear scaling as additional contracts are introduced. Using the full transaction lifecycle time per transaction:

$$Throughput_{HC} = \frac{1}{2.13} = 0.47\ tx/s$$

The observed throughput is relatively low due to several design and implementation constraints, including execution in a prototype environment with single-threaded processing, overhead from cross-chain anchoring, the inclusion of encryption and hashing operations, small block sizes, and the absence of batching optimizations. In contrast, many Hyperledger Fabric benchmark studies rely on synthetic workloads, omit cross-chain anchoring, and minimize cryptographic overhead. Consequently, the reported results more accurately reflect the practical performance limitations of real-world healthcare workflows.

3) Storage Overhead

In the main-chain-only design, complete medical records are stored on-chain, yielding a storage overhead of 1.0. On the other hand, HealthChain stores only lightweight metadata (e.g., UUIDs and cryptographic hashes) on-chain, while raw healthcare data are maintained off-chain. For the calculation of the storage overhead, this study considered a representative healthcare record size of 1 MB, which falls within the range commonly reported for EHRs containing structured clinical information and associated metadata [32]. This results in a storage overhead of 2.44×10^{-4}, demonstrating substantial storage efficiency.

$$Storage\ Overhead_{HC} = \frac{256}{1{,}048{,}576} \approx 2.44 \times 10^{-4}$$

4) Scalability

In the main-chain-only architecture, throughput quickly saturates due to reliance on a single consensus mechanism, exhibiting constant scalability O(1). In contrast, HealthChain enables parallel transaction execution across multiple contracts and sidechains, allowing throughput to increase proportionally with system expansion. Consequently, HealthChain demonstrates linear scalability O(n), making it suitable for large-scale healthcare environments.

Additional scalability experiments were conducted by incrementally increasing the number of sidechains. With one sidechain, throughput was 0.47 tx/s; with two sidechains, throughput increased to 0.89 tx/s; and with three sidechains, throughput reached 1.32 tx/s, while latency remained below 2.3 s across all configurations (Table I). These results indicate near-linear scalability proportional to the number of sidechains, supporting an O(n) scaling trend within the evaluated setup.

TABLE I. SCALABILITY SUMMARY

Number of sidechains	Throughput (tx/s)	Average latency (s)
1	0.47	2.13
2	0.89	2.21
3	1.32	2.28

The overall comparison of the baseline against the proposed HealthChain is presented in Table II.

5) Classification Models Evaluation

Table III summarizes the classification performance metrics, where Class 1 corresponds to diabetes cases. Although the Random Forest classifier achieved the highest overall accuracy (0.72) and precision (0.76), it exhibited a low recall (0.25), indicating limited sensitivity for detecting positive diabetes cases. However, the comparatively high accuracy of Random Forest is largely attributable to its strong performance on the majority class, whereas the lower recall reflects reduced sensitivity toward minority positive cases. In contrast, the Decision Tree and Voting classifiers achieved a more balanced trade-off between precision and recall, resulting in higher F1-scores for the positive class.

Table IV illustrates the confusion matrix for all the classifiers.

TABLE II. COMPARATIVE EVALUATION

Metric	Main-chain only	HealthChain
Throughput (tx/s)	0.1	0.47
Latency (s)	10.0	2.13
Storage overhead	1.0	2.44×10^{-4}
Scalability	O(1)	O(n)

TABLE III. PERFORMANCE COMPARISON OF CLASSIFICATION MODELS

Metric	CatBoost	Voting	Decision tree	Random forest	XGBoost
Accuracy	0.66	0.63	0.62	0.72	0.65
F1 (Class 1)	0.16	0.45	0.47	0.38	0.01
Precision (Class 1)	0.57	0.46	0.45	0.76	0.26
Recall (Class 1)	0.09	0.45	0.49	0.25	0.01
Macro F1	0.47	0.59	0.59	0.60	0.40

TABLE IV. CLASSIFIER CONFUSION MATRIX

Classifier	TN	FP	FN	TP
CatBoost	1,969	75	969	99
Voting	1,477	567	587	481
Decision tree	1,397	647	540	528
Random forest	1,960	84	799	269
XGBoost	2,024	20	1,061	7

V. DISCUSSION

The performance differences between the main-chain-only and the HealthChain architectures arise directly from the architectural design choices of the two systems. In a main-chain-only blockchain, all healthcare transactions, access requests, and audit events must be processed and validated by a single global consensus mechanism. This creates a bottleneck that limits transaction throughput and increases confirmation latency as the system load grows. Furthermore, because healthcare data references and metadata are stored on-chain, storage requirements grow rapidly with data volume, resulting in very high storage overhead. In contrast, HealthChain distributes workload across multiple sidechains, each operating its consensus process. This parallelization allows transaction throughput to scale as additional sidechains are introduced, while the main blockchain remains responsible only for governance, access control, and anchoring.

As a result, throughput increases and latency decreases due to reduced contention and faster finality on sidechains. Moreover, storage overhead is significantly reduced in

HealthChain because only compact cryptographic hashes, metadata, and access policies are recorded on the main blockchain. Additionally, large healthcare data are encrypted and stored, preventing blockchain state growth from becoming a scalability constraint. Consequently, the system exhibits near-linear scalability with respect to the number of sidechains, while a main-chain-only architecture remains inherently limited by its single-chain design.

The results also offer valuable insights into the suitability of various machine learning models for classifying health-related data within the proposed HealthChain model. The superior accuracy of the Random Forest classifier (0.72) can be attributed to its ensemble nature, which reduces variance by combining multiple decision trees and capturing non-linear feature interactions. This robustness is especially valuable when working with heterogeneous data such as demographic, physiological, and lifestyle attributes. However, the comparatively lower F1-score (Class 1) (0.38) indicates that Random Forest, while strong in overall prediction, does not fully resolve the imbalance between classes. In contrast, the Decision Tree achieved a higher F1-score (0.47) for the positive class. This suggests that combining multiple models improved the system's ability to balance precision and recall, thereby reducing the bias toward the majority class. Such a balance is crucial, where FN can have severe consequences.

Nevertheless, several limitations should be acknowledged. First, the dataset used represented a synthetic or generalized sample of medical students, which may not fully capture the complexity of real-world clinical populations. Moreover, while the experimental results demonstrated feasibility, large-scale validation on real-world healthcare data integrated into blockchain environments is necessary before practical deployment [33].

Regarding the operational costs and adoption barriers, estimated deployment costs include maintaining approximately three to five blockchain nodes per participating hospital, with projected cloud infrastructure expenses ranging from $500 to $1,500 per month per institution, depending on workload and availability requirements. Key adoption challenges include governance among participating healthcare organizations and obtaining regulatory approval to ensure compliance with healthcare data protection and interoperability standards.

VI. CONCLUSION

This paper presents HealthChain, a formally specified sidechain-based architecture for scalable and privacy-preserving healthcare data management. Unlike prior conceptual approaches, HealthChain provides explicit system definitions, operational workflows, and performance evaluation.

The results demonstrate that HealthChain achieves higher throughput and lower latency by offloading healthcare transactions and data references to sidechains while maintaining secure anchoring on the main blockchain. Moreover, the results demonstrated the feasibility of deploying machine learning classifiers within the HealthChain model. The Random Forest classifier achieved the highest predictive accuracy (0.72), confirming its robustness on heterogeneous health data. In contrast, the Decision Tree and Voting classifiers provided a more balanced trade-off between precision and recall, achieving higher F1-scores for the positive class. These findings emphasize that the model choice should be guided not only by accuracy but also by balanced performance, particularly in healthcare scenarios where FNs carry significant risks.

By integrating such classifiers into the blockchain-enabled sidechains, HealthChain can support scalable and secure predictive analytics alongside decentralized data governance. The dual capability of efficient computation with trustworthy interoperability offers a pathway toward more resilient and intelligent health information systems. Future research should extend this work by validating the model on real-world clinical data to ensure generalizability and practical adoption.

DECLARATION OF COMPETING INTERESTS

Not applicable to this work.

ACKNOWLEDGMENT

Not applicable to this work.

DATA AVAILABILITY

The dataset used is provided in [31].

REFERENCES

[1] F. M. AbdelSalam, "Blockchain Revolutionizing Healthcare Industry: A Systematic Review of Blockchain Technology Benefits and Threats," *Perspectives in Health Information Management*, vol. 20, no. 3, Sep. 2023, Art. no. 1b.

[2] M. S. B. Kasyapa and C. Vanmathi, "Blockchain integration in healthcare: a comprehensive investigation of use cases, performance issues, and mitigation strategies," *Frontiers in Digital Health*, vol. 6, Apr. 2024, Art. no. 1359858, https://doi.org/10.3389/fdgth.2024.1359858.

[3] S. Felemban *et al.*, "Current application of blockchain technology in healthcare and its potential roles in Urology," *BJU International*, vol. 136, pp. S5-S17, Oct. 2025, https://doi.org/10.1111/bju.16757.

[4] H. Taherdoost, "Privacy and Security of Blockchain in Healthcare: Applications, Challenges, and Future Perspectives," *Sci*, vol. 5, no. 4, Oct. 2023, Art. no. 41, https://doi.org/10.3390/sci5040041.

[5] V. Sitharamulu, G. Sucharitha, S. Nandan Mohanty, S. Janbhasha, and D. Kothandaraman, "A private Ethereum blockchain for organ donation and transplantation based on intelligent smart contracts," *Egyptian Informatics Journal*, vol. 28, Dec. 2024, Art. no. 100542, https://doi.org/10.1016/j.eij.2024.100542.

[6] J. Werth, M. Berenjestanaki, H. Barzegar, N. El Ioini, and C. Pahl, "A Review of Blockchain Platforms Based on the Scalability, Security and Decentralization Trilemma:," in *Proceedings of the 25th International Conference on Enterprise Information Systems*, 2023, pp. 146–155, https://doi.org/10.5220/0011837200003467.

[7] M. K. Pawar, P. Patil, and P. S. Hiremath, "A Study on Blockchain Scalability," in *ICT Systems and Sustainability*, vol. 1270, M. Tuba, S. Akashe, and A. Joshi, Eds. Singapore: Springer Singapore, 2021, pp. 307–316.

[8] F. Hashim, K. Shuaib, and N. Zaki, "Sharding for Scalable Blockchain Networks," *SN Computer Science*, vol. 4, no. 1, Oct. 2022, Art. no. 2, https://doi.org/10.1007/s42979-022-01435-z.

[9] Y. Liu *et al.*, "Building blocks of sharding blockchain systems: Concepts, approaches, and open problems," *Computer Science Review*, vol. 46, Nov. 2022, Art. no. 100513, https://doi.org/10.1016/j.cosrev.2022.100513.

[10] A. Singh, K. Click, R. M. Parizi, Q. Zhang, A. Dehghantanha, and K.-K. R. Choo, "Sidechain technologies in blockchain networks: An examination and state-of-the-art review," *Journal of Network and Computer Applications*, vol. 149, Jan. 2020, Art. no. 102471, https://doi.org/10.1016/j.jnca.2019.102471.

[11] R. Deepa and M. S. Arya, "Blockchain-Sidechain Based Data Storage for Reimaging Electronic Health Record via Optimized Interplanetary File System," in *Information and Communication Technology for Competitive Strategies (ICTCS 2020)*, vol. 191, A. Joshi, M. Mahmud, R. G. Ragel, and N. V. Thakur, Eds. Singapore: Springer Singapore, 2022, pp. 1097–1110.

[12] A. S. Yadav, N. Singh, and D. S. Kushwaha, "Sidechain: storage land registry data using blockchain improve performance of search records," *Cluster Computing*, vol. 25, no. 2, pp. 1475–1495, Apr. 2022, https://doi.org/10.1007/s10586-022-03535-0.

[13] O. Kuznetsov, P. Sernani, L. Romeo, E. Frontoni, and A. Mancini, "On the Integration of Artificial Intelligence and Blockchain Technology: A Perspective About Security," *IEEE Access*, vol. 12, pp. 3881–3897, 2024, https://doi.org/10.1109/ACCESS.2023.3349019.

[14] D. Bhumichai, C. Smiliotopoulos, R. Benton, G. Kambourakis, and D. Damopoulos, "The Convergence of Artificial Intelligence and Blockchain: The State of Play and the Road Ahead," *Information*, vol. 15, no. 5, May 2024, Art. no. 268, https://doi.org/10.3390/info15050268.

[15] Monika and R. Bhatia, "Interoperability Solutions for Blockchain," in *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, Oct. 2020, pp. 381–385, https://doi.org/10.1109/ICSTCEE49637.2020.9277054.

[16] I. Yaqoob, K. Salah, R. Jayaraman, and Y. Al-Hammadi, "Blockchain for healthcare data management: opportunities, challenges, and future recommendations," *Neural Computing and Applications*, vol. 34, no. 14, pp. 11475–11490, July 2022, https://doi.org/10.1007/s00521-020-05519-w.

[17] N. Atzei, M. Bartoletti, and T. Cimoli, "A Survey of Attacks on Ethereum Smart Contracts (SoK)," in *Principles of Security and Trust*, vol. 10204, M. Maffei and M. Ryan, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2017, pp. 164–186.

[18] K. Divya. and M. Mohan, "Sidechain: A Scalable Blockchain," in *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, May 2022, pp. 1337–1342, https://doi.org/10.1109/ICAAIC53929.2022.9793041.

[19] M. Li, H. Tang, A. R. Hussein, and X. Wang, "A Sidechain-Based Decentralized Authentication Scheme via Optimized Two-Way Peg Protocol for Smart Community," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 282–292, 2020, https://doi.org/10.1109/OJCOMS.2020.2972742.

[20] H. Wang and R. Zhou, "The Application of Blockchain to Electronic Health Record Systems: A Review," in *2021 International Conference on Information Technology and Biomedical Engineering (ICITBE)*, Dec. 2021, pp. 397–401, https://doi.org/10.1109/ICITBE54178.2021.00092.

[21] H. S. Adams. "NTT & Olympus: World's First Cloud Endoscopy System." Healthcare Digital. [Online]. Available: https://healthcare-digital.com/technology-and-ai/ntt-olympus-worlds-first-cloud-endoscopy-system.

[22] B. Chavali, S. K. Khatri, and S. A. Hossain, "AI and Blockchain Integration," in *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, June 2020, pp. 548–552, https://doi.org/10.1109/ICRITO48877.2020.9197847.

[23] S. S. Mohammed Abdul, "Navigating Blockchain's Twin Challenges: Scalability and Regulatory Compliance," *Blockchains*, vol. 2, no. 3, pp. 265–298, July 2024, https://doi.org/10.3390/blockchains2030013.

[24] A. Qambar, K. Shuaib, and M. Gergely, "Governing Blockchains in the Healthcare Ecosystem," in *Blockchain for Biomedical Research and Healthcare*, P. Kumar and A. Kumari, Eds. Singapore: Springer Nature Singapore, 2024, pp. 145–170.

[25] J. R. Quinlan, "Learning decision tree classifiers," *ACM Computing Surveys*, vol. 28, no. 1, pp. 71–72, Mar. 1996, https://doi.org/10.1145/234313.234346.

[26] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, https://doi.org/10.1023/A:1010933404324.

[27] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, Montréal, Canada, 2018, pp. 6639–6649.

[28] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California, USA, Aug. 2016, pp. 785–794, https://doi.org/10.1145/2939672.2939785.

[29] T. G. Dietterich, "Ensemble Methods in Machine Learning," in *Multiple Classifier Systems*, vol. 1857, Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 1–15.

[30] D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," 2020, https://doi.org/10.48550/ARXIV.2010.16061.

[31] *Medical Students Dataset*. (2023), S. Salem. [Online]. Available: https://www.kaggle.com/datasets/slmsshk/medical-students-dataset/data.

[32] J. C. Mandel, D. A. Kreda, K. D. Mandl, I. S. Kohane, and R. B. Ramoni, "SMART on FHIR: a standards-based, interoperable apps platform for electronic health records," *Journal of the American Medical Informatics Association*, vol. 23, no. 5, pp. 899–908, Sept. 2016, https://doi.org/10.1093/jamia/ocv189.

[33] A. S. Alfakeeh, "A Blockchain-Enabled IoT Framework for Secure Attack Detection and Advanced Feature Selection in Smart Healthcare," *Engineering, Technology & Applied Science Research*, vol. 15, no. 5, pp. 28219–28223, Oct. 2025, https://doi.org/10.48084/etasr.13349.

AUTHORS PROFILE

Ahmad Altamimi received the Ph.D. degree in Computer Science from Concordia University, Montreal, QC, Canada, in 2014. His major field of study is software engineering with applications in healthcare systems. He has received numerous awards and fellowships throughout his academic career. He has served as a Program Committee and Steering Committee Member for several international conferences. He has authored or coauthored more than 55 refereed publications, including journal and conference papers, book chapters, and books. His recent research interests include Bioinformatics, Blockchain, Healthcare, and Software Engineering. His research interests span healthcare, machine learning, privacy protection, and cybersecurity. His recent work focuses on developing advanced security models based on Blockchain technology to support interoperable healthcare systems.

Lamia T. Al-Kershi received the Doctor of Dental Medicine (D.D.M) degree from the University of Science and Technology, Amman, Jordan, and the M.Sc degree in Health Information Technology from Princes Sumaya University for Technology, Amman, Jordan. She is currently a Data Analyst and Epidemiological Reporting Consultant with the Eastern Mediterranean Public Health Network (EMPHNET), Amman, Jordan. Her work focuses on public health surveillance, data science applications, and digital health systems. She has contributed to regional research collaborations such as the Meningitis and Septicemia Mapping Network (MenMap) and presented findings at international scientific conferences, including the European Meningococcal and Haemophilus Disease Society Congress (EMGM).

Qusay Abdo received a bachelor's degree from Princess Sumaya University for Technology, Amman, Jordan. He is currently a Data Analyst at the Arab Bank Group, Amman, Jordan. His work focuses on public health, data science applications, and digital health systems. He has contributed to many research collaborations and presented findings at international scientific conferences.