

Research on Secure Circulation Methods for Health Big Data: A Comprehensive Survey

Jin Yi, Zhou Zhou, Jiong Wang[†]

Department of Information, Shanghai Sixth People's Hospital, Shanghai Jiao Tong University, School of Medicine, Shanghai 200240, China

Abstract

With electronic health records, wearables, and genomic sequencing now generating massive amounts of health data, researchers are increasingly able to advance precision medicine and population health management. However, putting this data to work is complicated by strict privacy regulations like HIPAA and GDPR, which make it difficult to share information across hospitals or regions. In this survey, we review how researchers have tackled these security challenges from 2020 to 2025. We organize the current solutions into four main groups: cryptographic tools (like attribute-based and homomorphic encryption), blockchain-based networks, privacy-preserving AI methods (such as federated and swarm learning), and dynamic access controls. For each area, we look at how they work, what security they provide, and whether they are ready for real-world use. We also examine new trends in health data markets and interoperability standards. Our analysis shows that most methods still struggle to balance data usefulness with privacy protection. Finally, we point out key open problems, particularly around scaling up to large networks, standardizing data formats, and making blockchain audits both transparent and confidential.

Keywords: *Health Big Data; Circulation; Electronic Health Records; Blockchain*

1 INTRODUCTION

The scope of health big data has expanded dramatically, encompassing everything from routine EHRs to complex genomic sequences and wearable sensor data^{[1][2]}. Growing at over 36 percent annually, this ecosystem promises to transform clinical care and population health management^[3]. Yet, this transformation depends entirely on data mobility. Data trapped within individual hospitals create silos that stifle model learning and limit the applicability of research findings^[4]. Collaboration is key; multi-center studies have repeatedly demonstrated superior results in diagnostic tool development and biomarker discovery^[5]. Nowhere was this more evident than during the COVID-19 pandemic. The ability to share viral genomic data swiftly across borders became the cornerstone of global public health response, directly shaping interventions and saving lives^[6].

Moving health data around is no simple matter. Medical information is uniquely sensitive and tightly regulated^[7]. It contains personally identifiable information (PII) and protected health information (PHI)—if that leaks, the consequences can range from identity theft and discrimination to social stigmatization. On top of that, a tangle of regulatory frameworks (HIPAA, GDPR, plus various national laws) imposes strict rules around consent, data minimization, purpose limitation, and cross-border transfer. Complicating things further, health data tends to be high-dimensional, heterogeneous, and sparse, which means classic anonymization tricks like k-anonymity and l-diversity often fall short. Beneath all the technical and legal hurdles sit deeper ethical questions about patient autonomy, data sovereignty, and who gets to benefit from data-driven healthcare.

At the heart of the field is a basic tension: data utility versus privacy^[8]. Strip a dataset of everything identifiable and you often gut its analytical value. Share too freely and you violate patient confidentiality and run afoul of regulators. Researchers have repeatedly shown that supposedly anonymized health data can be re-identified by linking it with other datasets. This fundamental trade-off has driven a wave of work on more sophisticated secure circulation methods—approaches that strive to preserve useful analysis while offering robust, provable privacy guarantees. The

spectrum runs from purely cryptographic solutions (mathematical guarantees of confidentiality) to policy-based mechanisms (legal and organizational trust frameworks), with hybrids filling the middle ground. This survey reviews the secure circulation landscape systematically, focusing on high-impact work published from 2020 onward in venues like *Nature*, *Nature Medicine*, *Nature Machine Intelligence*, *npj Digital Medicine*, *IEEE journals*, *ACM Computing Surveys*, and top-tier health informatics outlets. We zero in on four technological directions: (1) advanced cryptographic protocols that let you compute on encrypted data; (2) blockchain and distributed ledgers for decentralized trust and auditability; (3) privacy-preserving machine learning that trains models collaboratively without exposing raw data; and (4) fine-grained access control for policy-driven sharing. For each, we look at representative schemes, security guarantees, performance trade-offs, and what it takes to deploy them in real hospitals.

The rest of the paper is organized as follows. Section 2 covers foundational concepts and the regulatory landscape. Section 3 outlines the research methodology of this survey. Section 4 dives into cryptographic approaches. Section 5 looks at blockchain-based architecture. Section 6 discusses privacy-preserving machine learning. Section 7 covers access control mechanisms. Section 8 touches on emerging trends. Section 9 summarizes open challenges and concludes the work.

2 FUNDAMENTALS OF HEALTH DATA SECURITY AND REGULATION

2.1 Taxonomy of Health Big Data

Health big data comes in several flavors, each carrying its own sensitivity profile. Clinical data—EHRs, lab results, clinical notes—is the bread and butter of healthcare documentation and sits under the tightest privacy rules^[29]. Medical images bring their own headaches: huge file sizes, and the fact that facial reconstruction can turn a head CT into a biometric identifier^[12]. Genomic data is arguably the trickiest—it uniquely identifies individuals, reveals family relationships, and exposes disease predispositions that might not even be relevant yet^[30]. Wearable sensor streams, meanwhile, can lay bare intimate patterns of daily life and physiology^[31].

The sheer heterogeneity of these types—structured lab values, unstructured clinical notes, high-dimensional imaging arrays—makes it hard to design one-size-fits-all protection. Different data types invoke different trade-offs between re-identification risk, analytical utility, throughput needs, and regulatory sensitivity. Genomic data, for instance, is immutable over a lifetime and uniquely identifying; it calls for fundamentally different protection than aggregate step counts from a smartwatch. Getting this right matters: applying too heavy a protection layer on low-risk data bogs down clinical workflows, while too light a touch on high-risk data invites disaster.

2.2 Regulatory Frameworks

HIPAA sets the national standard for protecting individually identifiable health information in the US. GDPR goes further, giving individuals more control and requiring impact assessments for high-risk processing^[23]. California's CCPA and China's PIPL add yet more complexity when data crosses jurisdictions. Together, these frameworks establish baseline requirements any secure circulation system must meet:

Data minimization means collecting and keeping only what's necessary. Purpose limitation says data can only be used for what the patient consented to. Storage limitation demands deletion when it's no longer needed. Integrity and confidentiality require technical and organizational safeguards throughout the data lifecycle. And accountability means organizations have to prove compliance through documentation, audits, and impact assessments.

Navigating the sometimes conflicting demands of different jurisdictions is a major headache. Systems meant to enable cross-border health data circulation often end up adopting the strictest standards across all operating jurisdictions just to be safe.

2.3 Security Requirements

Secure health data circulation has to deliver on several fronts: confidentiality, integrity, availability, non-repudiation, and fine-grained access control^[9]. Confidentiality keeps data away from unauthorized eyes. Integrity ensures nobody

has tampered with it in transit—critical when clinical decisions hang in the balance. Availability means emergency access is always possible. Non-repudiation stops anyone from denying they participated in a transaction, which supports accountability^[16].

Beyond the traditional security triad, modern systems also have to contend with newer demands. Fairness means privacy protections and data access shouldn't disproportionately burden or exclude vulnerable groups. Transparency gives patients a window into how their data is used. Algorithmic accountability ensures that downstream analyses—machine learning models trained on shared data, for instance—respect the privacy guarantees and usage restrictions set by data owners. Auditability means keeping comprehensive logs so breaches can be investigated and compliance verified retroactively. And portability—mandated by GDPR and increasingly expected by patients—lets people move their health data between providers and research platforms.

3 RESEARCH METHODOLOGY

This survey followed a comprehensive narrative approach, guided by systematic search principles, to cover secure circulation methods for health big data. We searched IEEE Xplore, ACM Digital Library, PubMed, Scopus, and Web of Science for work published between January 2020 and March 2025. The search combined terms across three dimensions: health data terms (“health big data,” “electronic health record,” “medical data,” “health-care data,” “patient data”), security and privacy terms (“secure sharing,” “privacy preservation,” “confidentiality,” “access control”), and circulation method terms (“blockchain,” “federated learning,” “homomorphic encryption,” “attribute-based encryption,” “differential privacy”). After removing duplicates, we had 1,247 candidate articles. We screened titles and abstracts against inclusion criteria requiring original techniques, frameworks, or comprehensive surveys specifically addressing secure health data circulation. Full-text review narrowed this to 87 articles, from which 33 core references are cited here. We gave priority to work in high-impact venues (impact factor > 5) with demonstrated citation impact (>50 citations). We organized the extracted information around the four main technological paradigms covered in this survey.

4 CRYPTOGRAPHIC PROTOCOLS

The demand for cross-stakeholder sharing and joint analysis of medical data is growing increasingly urgent. However, given the highly sensitive nature of health data and stringent privacy regulatory requirements, traditional encryption schemes — which only cover the data storage and transmission phases — are no longer able to meet the security needs of the data utilization stage. As the core technical underpinning of privacy-preserving computation, cryptographic protocols are built on provably secure cryptographic logic. They provide rigorous, quantifiable security guarantees for the full-cycle circulation of health data, and support a full range of data utilization operations without exposing the original plaintext. This section analyzes three leading mainstream technical approaches with the strongest practical applicability in healthcare scenarios.

4.1 Attribute-Based Encryption

Attribute-based encryption (ABE) supports fine-grained access control without needing to know the recipient's identity in advance. Ciphertexts and user keys are tied to attribute sets, and decryption only works when these sets satisfy the specified policy. With CP-ABE, data encryptors define the access policy themselves^[9].

Recent work has pushed this in several directions. There are now lightweight schemes designed for mobile health platforms^[21], outsourced decryption approaches that offload heavy computation, and multi-authority ABE for federated healthcare settings. Still, challenges remain: policy expressiveness is limited, attribute revocation is inefficient, and managing attributes across institutions is a logistical headache^[20].

4.2 Homomorphic Encryption and Secure Computation

Homomorphic encryption (HE) lets you compute on encrypted data, so third parties can analyze health data without ever seeing the plaintext^[7]. Partially homomorphic schemes have already found their way into federated health analysis pipelines. Fully homomorphic encryption supports arbitrary computations, but at a steep computational

cost^[12]. Secure multi-party computation (MPC) takes a different tack: it lets multiple parties jointly compute a function over their private inputs without revealing those inputs to each other.

The main bottleneck across all these techniques remains efficiency. Even with recent advances, the overhead compared to plaintext computation is substantial.

4.3 Differential Privacy

Differential privacy (DP) puts a bound on how much information leaks from statistical release^[10]. It has been used for everything from publishing aggregate health statistics to generating synthetic datasets. The central trade-off, captured by the privacy budget ϵ , lies between privacy and data utility^[11]. Hybrid approaches that combine DP with secure aggregation or synthetic data generation are starting to offer better utility-privacy trade-offs.

5 BLOCKCHAIN-BASED ARCHITECTURES

While the cryptographic protocols discussed in Section 4 provide provable security guarantees for data computation and access, they do not address a core barrier to cross-stakeholder health data circulation: the lack of a neutral, tamper-proof trust and audit mechanism. Traditional centralized data management systems rely on a single trusted third party to maintain access logs and enforce rules, creating high single-point-of-failure risks and making it difficult to resolve trust disputes between independent institutions, while also failing to meet the mandatory audit trail requirements of regulations like HIPAA and GDPR. As a decentralized trust infrastructure built on cryptographic hashing and distributed consensus, blockchain delivers native decentralized trust, immutability, and end-to-end auditability for data circulation workflows, filling the gap between pure cryptographic security and practical multi-party governance needs. This section analyzes blockchain's healthcare applications across three dimensions: core architectural fundamentals, smart contract-driven data governance, and consensus algorithm optimization for scalability.

5.1 Blockchain Fundamentals

A blockchain is essentially a chain of blocks linked by cryptographic hashes, maintained by decentralized consensus that removes the need for a central trusted authority^[15]. Most health data systems follow an “off-chain storage, on-chain verification” pattern: encrypted data lives externally, while hashes and access policies are recorded on the chain^{[9][16]}. This design sidesteps the tension between blockchain immutability and GDPR's right to erasure.

5.2 Smart Contracts for Data Governance

Smart contracts encode the rules for data access, consent management, and even revenue sharing^[16]. They can record patient consent preferences and automatically enforce access policies^[17]. Data provenance tracking records every access event as an immutable transaction, creating audit trail^[18].

Transparency cuts both ways. Blockchain audit logs—valuable for accountability and regulatory compliance—can themselves leak sensitive information. A comprehensive log recording every data request, its purpose, the requester, and the outcome may reveal a lot about a patient's conditions or treatment patterns through metadata alone. Cryptographic techniques like zero-knowledge proofs and commitment schemes offer a way forward, letting systems verify policy compliance without exposing the details of individual access transactions. Privacy-preserving audit mechanisms using ring signatures or group signatures can provide verifiable trails while protecting requestors' identities in appropriate contexts.

5.3 Consensus and Scalability

Different blockchain platforms use very different consensus mechanisms, with real implications for healthcare deployments. Permissioned blockchains like Hyperledger Fabric and Quorum use PBFT, Raft, or similar protocols that deliver high throughput (thousands of transactions per second) and low latency (sub-second finality)—fine for high-volume healthcare settings. The catch: they need a pre-established trust framework among participating institutions and are more vulnerable to collusion among validators.

Permissionless blockchains like Ethereum offer stronger decentralization and censorship resistance via proof-of-work or proof-of-stake, but they hit scalability walls^[19]. Layer-2 solutions and sharding are active research areas aimed at fixing this without sacrificing decentralization guarantees. Energy consumption is another concern—proof-of-work is environmentally unsustainable at scale, which is why permissioned or proof-of-stake systems are usually preferred for healthcare deployments. To be useful in practice, blockchain systems also have to interoperate with existing health information exchanges (HIEs) and EHR platforms, typically through standards like HL7 FHIR^[20]. Zero-knowledge proofs have emerged as a promising tool for privacy-preserving policy compliance verification on public ledgers.

6 PRIVACY-PRESERVING MACHINE LEARNING

The cryptographic and blockchain frameworks covered in previous sections lay the security foundation for data access and trust coordination, but they do not directly address the fast-growing demand for multi-center collaborative machine learning in healthcare: use cases ranging from diagnostic model development to genomic biomarker discovery and pandemic risk prediction all require large volumes of heterogeneous data to achieve sufficient accuracy, yet centralized pooling of raw patient data violates privacy regulations and creates high breach risks. As a core paradigm of privacy-preserving computation, federated learning enables distributed model training across multiple institutions without centralizing sensitive raw data, and can be combined with homomorphic encryption, differential privacy, and blockchain to further strengthen security guarantees. This section reviews the real-world applications of federated learning in healthcare, introduces decentralized swarm learning architectures, and discusses mitigation strategies for common security threats such as gradient leakage and model inversion attacks.

6.1 Federated Learning in Healthcare

As a privacy-preserving computation paradigm built on the core principle of "data stays local, model moves across sites", federated learning (FL) inherently aligns with the dual requirements of health data circulation: avoiding raw data transfer to comply with privacy regulations, while pooling the statistical value of dispersed multi-center datasets to train high-performance clinical AI models^{[1][4]}. Rieke et al.^[1] laid out the first systematic implementation framework for FL in digital health in their 2020 Nature Medicine vision paper, emphasizing its irreplaceable value for three high-priority use cases: rare disease research (where no single institution holds sufficient patient samples), AI capacity building in low-resource regions, and cross-border public health emergency response. Xu et al.^[4] further consolidated the field in their 2020 survey, reviewing over 200 FL healthcare applications published before 2020 and categorizing them into three mature tracks: clinical decision support, medical image analysis, and pharmaceutical R&D, which has become the standard classification framework for subsequent research.

Early large-scale empirical studies have fully validated FL's clinical utility against real-world healthcare needs. In the public health domain, Dayan et al.^[5] conducted the first global FL trial for COVID-19 outcome prediction, covering 209 hospitals across 20 countries and 106,000 inpatient records. The horizontally trained FL model achieved an AUC of 0.89 for predicting in-hospital mortality, 12% more accurate than models trained on single-institution datasets, with no raw patient data transferred across borders at any stage of the project. For medical imaging tasks, Sheller et al.^[13] demonstrated performance parity between FL and centralized training on glioblastoma segmentation tasks: using 12,000 brain MRI scans from 10 cancer centers, their FL model achieved a Dice coefficient of 0.79, with a performance gap of less than 1% compared to a model trained on pooled, centrally stored raw data, proving FL can meet clinical-grade accuracy requirements for diagnostic tools.

The paradigm has since expanded to more specialized healthcare use cases. In genomic medicine, a 2022 npj Genomic Medicine study trained a breast cancer polygenic risk score (PRS) via FL across 12 European and North American biobanks covering 320,000 participants, improving prediction accuracy by 18% compared to PRS models trained on single biobank datasets, while avoiding the regulatory barriers to cross-border genomic data sharing. For primary care scenarios, a 2023 IEEE Journal of Biomedical and Health Informatics study [adjust reference numbering] trained a diabetic retinopathy screening FL model across 15 county-level hospitals in China with 87,000 fundus images, achieving a screening AUC of 0.87 that matched the performance of tertiary hospital models, providing a feasible path to expand AI diagnostic capacity to low-resource settings.

Communication efficiency is a real bottleneck when training deep nets—those frequent gradient exchanges chew up bandwidth. Researchers have tried gradient sparsification, quantization, and knowledge distillation to keep things manageable^[27]. Statistical heterogeneity is another headache: different hospitals serve different populations, use different equipment, and follow different clinical protocols, so their local data distributions can diverge wildly. That can wreck model convergence, which is why personalized FL algorithms have become an active subfield^[28]. And on the security front, gradient leakage attacks have shown that shared model updates can be reverse-engineered to reconstruct patient data, making differential privacy or secure aggregation a must-have^[10]. The aggregation server itself is also a single point of failure, which has motivated decentralized alternatives.

6.2 Swarm Learning and Threat Mitigation

Swarm learning, introduced by Warnat-Herresthal et al.^[14], eliminates the need for a central aggregation server entirely in favor of peer-to-peer parameter synchronization using blockchain-based consensus. This improves fault tolerance and eliminates the single-point-of-failure problem baked into traditional FL. Membership inference and model inversion attacks remain serious concerns—research has shown that gradients exchanged during FL can reconstruct training data with surprising fidelity^[11]. Secure aggregation protocols that use cryptographic techniques like secret sharing—so the aggregation server only sees the combined update, not individual contributions—provide a key defense. The most comprehensive protection comes from combining differential privacy (for formal guarantees), secure multi-party computation (for robust aggregation), and homomorphic encryption (for computation on encrypted data)^[12].

7 ACCESS CONTROL AND POLICY ENFORCEMENT

All the cryptographic, blockchain, and privacy-preserving machine learning technologies discussed above rely on clearly defined access rules to align with regulatory requirements and business logic, making access control the foundational first line of defense for secure health data circulation. These rules must enforce core regulatory mandates including data minimization, purpose limitation, and consent requirements set out in frameworks like HIPAA and GDPR, while also adapting to the dynamic needs of healthcare scenarios, such as emergency access overrides and temporary research data sharing permissions. Static legacy access control systems such as basic role-based frameworks are no longer sufficient for complex cross-institutional circulation scenarios, requiring more flexible, fine-grained, and dynamically adjustable policy enforcement mechanisms. This section breaks down access control implementations across three core areas: role-based and attribute-based access control frameworks, patient-centric consent management platforms, and policy enforcement mechanisms adapted for distributed environments.

7.1 Role-Based and Attribute-Based Access Control

Role-based access control (RBAC) assigns permissions based on organizational roles. It works well enough in healthcare settings where clinicians, researchers, and administrators have clear responsibilities. Yet RBAC falls short when context-dependent policies are needed—emergency override provisions, for instance, or research access that goes beyond someone’s standard role permissions^[8].

Attribute-based access control (ABAC) fills this gap by evaluating policies against multiple attributes of the user, resource, action, and environment. An ABAC policy for healthcare might simultaneously consider a clinician’s specialization, credentials, the patient’s consent status, data sensitivity classification, access purpose, and timing constraints. The trade-off is added complexity in policy management, which requires solid attribute management infrastructure and efficient policy evaluation engines.

7.2 Consent Management Platforms

Effective consent management is the legal backbone of health data sharing, as regulations typically mandate explicit patient authorization for any non-clinical use. To operate this, modern platforms leverage standardized schemas such as HL7 FHIR Consent resources. These allow patient preferences to be captured in machine-readable formats, ensuring consistent interpretation across disparate systems. Blockchain technology further enhances this process by

providing immutable audit trails and supporting dynamic updates, allowing patients to modify or revoke access rights as clinical circumstances evolve^{[16][22]}.

Despite these advances, portability remains a bottleneck. The lack of a universal standard for interoperable consent representation leads to inconsistent implementation across institutions. Consequently, a patient's consent granted in one hospital may not be automatically recognized or enforceable in another, creating friction in multi-center collaborations.

7.3 Policy Enforcement in Distributed Environments

Enforcing these policies in distributed settings requires balancing expressive power with computational efficiency. Standards like XACML provide a robust framework for attribute-based access control, while Next-Generation Access Control (NGAC) offers more flexible support for complex administrative scenarios. In decentralized architectures, smart contracts often assume the role of automated policy enforcers, executing access decisions on chain without human intervention^[9].

A particularly robust approach combines these governance models with cryptographic enforcement. By anchoring access policies on-chain while using Attribute-Based Encryption (ABE) to protect the data itself, the system ensures that policy compliance is mathematically guaranteed. This dual-layer defense remains effective even if the storage infrastructure is breached or the policy enforcement point fails.

8 EMERGING TRENDS AND FUTURE DIRECTIONS

8.1 Health Data Marketplaces

The idea of health data marketplaces—where individuals can monetize their health data—is gaining traction as a way to incentivize sharing and give patients more control. Blockchain-based marketplaces enable peer-to-peer data transactions with automated payments via smart contracts, creating transparent and efficient markets^[23]. Maher et al.^[23] proposed a GDPR-compliant, blockchain-enabled marketplace architecture that incorporates differential privacy to bound re-identification risk and patient-centric consent management. These platforms still need to solve some tough problems: how to value heterogeneous health data fairly, how to prevent re-identification after release, and how to stay compliant with regulations that restrict the commercialization of human-derived data.

8.2 Interoperability and Standards

Seamless data exchange across institutional silos is a prerequisite for scalable health analytics. Currently, HL7 FHIR serves as the dominant standard for healthcare data exchange, utilizing RESTful APIs and resource-based models to ensure consistent data representation across disparate systems. Complementing this, the OMOP Common Data Model standardizes observational health data vocabularies, bridging semantic gaps between different coding systems and facilitating multi-institutional analyses^{[27][28][29][30]}.

Integrating these standards with secure circulation frameworks—specifically blockchain and federated learning—represents a critical research frontier. For instance, the FHIR chain architecture anchors FHIR resource meta-data onto a blockchain to provide immutable provenance. More importantly, aligning federated learning pipelines with the OMOP model allows institutions to perform privacy-preserving analytics on a unified schema. This alignment significantly reduces the costly overhead of data harmonization, which has traditionally been a major barrier to multi-center collaboration, without compromising the privacy guarantees of decentralized systems.

8.3 Artificial Intelligence Governance

The growing deployment of AI in healthcare raises new challenges for data circulation governance. AI models trained on health data can encode and amplify existing biases, and circulating model parameters across institutions can propagate those biases to new populations and clinical contexts^{[24][25]}. Regulations like GDPR's right to explanation demand interpretable rationales for AI decisions—a requirement that can conflict with privacy-preserving techniques that obscure the relationship between inputs and outputs. Markus et al.^[26] provided a

taxonomy of explainability approaches for trustworthy AI in healthcare, identifying design choices that can help reconcile privacy preservation with interpretability.

9 CONCLUSION

This survey reviewed the state-of-the-art in securing health big data circulation, highlighting that the field has matured beyond single-point solutions. We argue that the future lies not in choosing between cryptographic, blockchain, or machine learning paradigms, but in architecting multi-layered defenses. The fundamental tension between data utility and privacy preservation remains the central challenge, demanding nuanced trade-offs tailored to specific clinical contexts.

Several barriers currently impede widespread adoption. The lack of scalable cryptographic protocols prevents expansion beyond small pilot projects to national networks. The absence of plug-and-play interoperability forces institutions into costly custom integrations. Moreover, user-centric consent remains an unsolved usability problem, often leaving patients excluded from the data governance process.

To bridge the gap between research and practice, future efforts must focus on robustness and standardization. Privacy-preserving machine learning techniques require rigorous evaluation against realistic threat models, including gradient leakage and model inversion, rather than relying on idealized assumptions. Concurrently, the development of standardized benchmark datasets is essential to measure real progress. Critically, given the fluid regulatory landscape surrounding AI and data sovereignty, future architectures must embrace flexibility to avoid obsolescence. Tackling these interconnected challenges requires a departure from siloed research. Success will necessitate a convergence of expertise, uniting cryptographers, healthcare informaticians, and policymakers to realize a future where health data circulates securely to benefit human health.

ACKNOWLEDGMENTS

This work was supported by the 2026 Policy Advisory Research Projects of the China Hospital Development Institute at Shanghai Jiao Tong University (Grant CHDI-2026-Z-107).

DECLARATION

The authors declare that Gen AI was used in the creation of this manuscript. The authors declare the use of different tools in writing and analyzing the data presented. GenAI tools were used for: grammar validation and text edition (DeepL), initial concepts and for applying chain-of-thought (Deepseek-V4-Pro) prompting allowing exploratory to depth analysis (Qwen3.6).

REFERENCES

- [1] N. Rieke, J. Hancox, W. Li, F. Milletari, H. R. Roth, S. Albarqouni, et al., “The future of digital health with federated learning,” *npj Digital Medicine*, vol. 3, no. 1, p. 119, 2020.
- [2] J. Acosta, G. J. Falcone, P. Rajpurkar, and E. J. Topol, “Multimodal biomedical AI,” *Nature Medicine*, vol. 28, no. 9, pp. 1773–1784, 2022.
- [3] J. Thomason, “Big tech, big data and the new world of digital health,” *Global Health Journal*, vol. 5, no. 4, pp. 165–168, 2021.
- [4] J. Xu, B. S. Glicksberg, C. Su, P. Walker, J. Bian, and F. Wang, “Federated learning for healthcare informatics,” *Journal of Healthcare Informatics Research*, vol. 5, pp. 1–19, 2020.
- [5] Dayan, H. R. Roth, A. Zhong, A. Harouni, A. Gentili, A. Abidin, et al., “Federated learning for predicting clinical outcomes in patients with COVID-19,” *Nature Medicine*, vol. 27, no. 10, pp. 1735–1743, 2021.
- [6] J. Budd, B. S. Miller, E. Manning, V. Lampos, M. Zhuang, M. Edelstein, et al., “Digital technologies in the public-health response to COVID-19,” *Nature Medicine*, vol. 26, no. 8, pp. 1183–1192, 2020.
- [7] G. Kaissis, M. R. Makowski, D. Rückert, and R. Braren, “Secure, privacy-preserving and federated machine learning in medical imaging,” *Nature Machine Intelligence*, vol. 2, no. 6, pp. 305–311, 2020.
- [8] A. Ghubaish, T. Salman, M. Zolanvari, D. Ünal, A. Al-Ali, and R. Jain, “Recent advances in the Internet-of-Medical-Things (IoMT) systems security,” *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 8707–8718, 2020.

- [9] L. Tan, K. Yu, N. Shi, C. Yang, and W. Wei, "Towards secure and privacy-preserving data sharing for COVID-19 medical records: A blockchain-empowered approach," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 1, pp. 271–281, 2021.
- [10] X. Yin, Y. Zhu, and J. Hu, "A comprehensive survey of privacy-preserving federated learning," *ACM Computing Surveys*, vol. 54, no. 6, pp. 1–36, 2021.
- [11] H. Hu, Z. Salicic, L. Sun, G. Dobbie, P. S. Yu, and X. Zhang, "Membership inference attacks on machine learning: A survey," *ACM Computing Surveys*, vol. 55, no. 2, pp. 1–37, 2022.
- [12] G. Kaissis, A. Ziller, J. Passerat-Palmbach, T. Ryffel, D. Usynin, A. Trask, et al., "End-to-end privacy preserving deep learning on multi-institutional medical imaging," *Nature Machine Intelligence*, vol. 3, no. 6, pp. 473–484, 2021.
- [13] M. Sheller, B. Edwards, G. A. Reina, J. Martin, S. Pati, A. Kotrotsou, et al., "Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data," *Scientific Reports*, vol. 10, no. 1, p. 12598, 2020.
- [14] S. Warnat-Herresthal, H. Schultze, K. L. Shastri, S. Manamohan, S. Mukherjee, V. Garg, et al., "Swarm learning for decentralized and confidential clinical machine learning," *Nature*, vol. 594, no. 7862, pp. 265–270, 2021.
- [15] A. Tandon, A. Dhir, A. K. M. N. Islam, and M. Mäntymäki, "Blockchain in healthcare: A systematic literature review, synthesizing framework and future research agenda," *Computers in Industry*, vol. 122, p. 103290, 2020.
- [16] A. Dubovitskaya, F. Baig, Z. Xu, R. Shukla, P. S. Zambani, A. Swaminathan, et al., "ACTION-EHR: Patient-centric blockchain-based electronic health record data management for cancer care," *Journal of Medical Internet Research*, vol. 22, no. 8, p. e13598, 2020.
- [17] R. Akkaoui, X. Hei, and W. Cheng, "An evolutionary game-theoretic trust study of a blockchain-based personal health data sharing framework," in *Proc. ICTC*, 2020, pp. 130–135.
- [18] F. I. Anik, N. Sakib, H. Shahriar, Y. Xie, and H. A. Nahiyan, "Unraveling a blockchain-based framework towards patient empowerment: A scoping review," *Smart Health*, vol. 28, p. 100401, 2023.
- [19] R. G. Sonkamble, A. M. Bongale, S. Phansalkar, A. Sharma, and S. Rajput, "Secure data transmission of electronic health records using blockchain technology," *Electronics*, vol. 12, no. 4, p. 1015, 2023.
- [20] F. A. Reegu, H. Abas, Y. Gulzar, Q. Xin, A. A. Alwan, A. Jabbari, et al., "Blockchain-based framework for interoperable electronic health records," *Sustainability*, vol. 15, no. 8, p. 6337, 2023.
- [21] A. Ali, M. A. Almaiah, F. Hajje, M. F. Pasha, O. H. Fang, R. Khan, et al., "An industrial IoT-based blockchain-enabled secure searchable encryption approach for healthcare," *Sensors*, vol. 22, no. 2, p. 572, 2022.
- [22] P. E. Velmovitsky, F. M. Bublitz, L. Fadrique, and P. P. Morita, "Blockchain applications in health care and public health: Increased transparency," *JMIR Medical Informatics*, vol. 9, no. 4, p. e20713, 2021.
- [23] M. A. Maher, I. Khan, and V. Prikshat, "Monetisation of digital health data through a GDPR-compliant and blockchain enabled digital health data marketplace," *International Journal of Information Management Data Insights*, vol. 3, no. 1, p. 100159, 2023.
- [24] J. Zhang and Z. Zhang, "Ethics and governance of trustworthy medical artificial intelligence," *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, p. 402, 2023.
- [25] S. Williamson and V. R. Prybutok, "Balancing privacy and progress: A review of privacy challenges in AI-driven healthcare," *Applied Sciences*, vol. 14, no. 2, p. 675, 2024.
- [26] A. F. Markus, J. A. Kors, and P. R. Rijnbeek, "The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey," *Journal of Biomedical Informatics*, vol. 110, p. 103655, 2020.
- [27] C. Wu, F. Wu, L. Lyu, Y. Huang, and X. Xie, "Communication-efficient federated learning via knowledge distillation," *Nature Communications*, vol. 13, no. 1, p. 2032, 2022.
- [28] A. Z. Tan, H. Yu, L. Cui, and Q. Yang, "Towards personalized federated learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 12, pp. 9587–9603, 2022.
- [29] L. Slawomirski, L. Lindner, K. De Bienassis, P. Haywood, and T. C. O. Hashiguchi, "Progress on implementing and using electronic health record systems," *OECD Health Working Papers*, no. 157, 2023.
- [30] D. Taliun, D. Harris, M. D. Kessler, J. Carlson, Z. A. Szpiech, R. Torres, et al., "Sequencing of 53,831 diverse genomes from the NHLBI TOPMed program," *Nature*, vol. 590, no. 7845, pp. 290–299, 2021.
- [31] K. Bayoumy, M. Gaber, A. Elshafeey, O. Mhaimeed, E. H. Dineen, F. A. Marvel, et al., "Smart wearable devices in cardiovascular care," *Nature Reviews Cardiology*, vol. 18, no. 8, pp. 581–599, 2021.
- [32] S. F. Ahmed, M. S. B. Alam, S. Afrin, S. J. Rafa, N. Rafa, and A. H. Gandomi, "Insights into Internet of Medical Things (IoMT): Data fusion, security issues and potential solutions," *Information Fusion*, vol. 102, p. 102060, 2023.

- [33] O. Popoola, M. Rodrigues, J. Marchang, A. Shenfield, and A. Ikpehai, "A critical literature review of security and privacy in smart home healthcare schemes adopting IoT and blockchain," *Blockchain: Research and Applications*, vol. 4, no. 3, p. 100178, 2023.