Blockchain Application on Big Data Security

Esraa Elgamal
Information System
Benha University
Cairo, Egypt
esraa.Kamal@fci.bu.edu.eg

Walaa Medhat
Benha University, Nile University
Cairo, Egypt
wmedhat@nu.edu.eg

Mohamed Abd Elfatah
Benha University
Cairo, Egypt
mohamed.abdo@fci.bu.edu.eg

Nashwa Abdelbaki
Nile University
Cairo, Egypt
nabdelbaki@nu.edu.eg

***Abstract*— In recent years, advances in technology in several industries have resulted in massive data collections on the web. It raises worries about large data security and protection. The advent of Blockchain technology has caused a revolution in the security field for different applications. The distributed ledger is stored on each Blockchain node, which enhances security and data transparency. On the Blockchain network, illegal users are not authorized to undertake any fault transactions. In this article, we will discuss how Blockchain may be employed to secure the big data. We explain the problems that the Blockchain faced with big data and its solutions. We summarize recent works of Blockchain with big data and the present issues and trends. We demonstrate that Blockchain technology is still in its initial phases of validation and there are no large-scale application scenarios available, particularly in the big data sector. Finally, we narrow our study to the Healthcare industry and offer the following research directions for its primary issues.**

Keywords—Blockchain, Big Data, Healthcare, IOT, Agriculture, Electronic Voting, Social Network.

I. INTRODUCTION

Because of the fast expansion of the Internet and the popularity of communication, the concept of big data became more widely known and used. These data come from a variety of sources, including internet transactions, social networks, health records, science data sensors, mobile phones, and associated applications. For many enterprises, big data adoption comes down to one question: how can we leverage big data's value while successfully managing big data security risks? [1].

Theft of sensitive or personal information kept online, DDoS (Distributed Denial of Service) attacks, or ransomware are all examples of big data security issues. Furthermore, attacks on an organization's large data storage might result in significant financial consequences such as damages, legal expenses, and penalties or punishments. Incoming data is a major source of big data security best practices. Data in transit might be intercepted or damaged, and data in storage could be stolen or held hostage while sitting on cloud or on-premises servers. Finally, the output data, which may appear insignificant, may provide an entry point for hackers or other harmful parties. [1]. Blockchain technology use is one of the current methods presented to overcome huge data security challenges. The Blockchain technology offers several advantages that distinguish it and make it suitable for integration with electronic systems in various fields. This is what we will explain in this research study, to what extent the Blockchain has solved several problems in different fields, and what is expected from it. As the Healthcare industry faces numerous challenges, we concentrate on it in our study. We identify these issues and provide a summary of the most recent research. We suggest and recommend the researchers for the expected next step of Blockchain technology-based Healthcare systems.

The reminder of this paper is divided into five sections. Section II provides an overview about Blockchain. Section III covers earlier efforts linked to the use of Blockchain technology in big data systems, as well as Blockchain applications in several domain areas. Section IV discusses the importance of using Blockchain technology in the healthcare system and refers to the most current study in this field, which addresses the following research topic. Section V summarizes the article and discusses the future work.

II. BLOCKCHAIN BACKGROUND

Satoshi Nakamoto invented Blockchain technology when he introduced bitcoin, the first decentralized cryptocurrency. It is a sort of peer-to-peer storage that is distributed. Each Blockchain node contains an immutable copy of the same data [2]. It offers a secure, shared digital ledger for data storage in a private or public peer-to-peer network and use consensus techniques to ensure data consistency across nodes. Blockchain networks are classified into three types: public Blockchain, private Blockchain, and consortium Blockchain. TABLE I. shows the comparison between them.

TABLE I. BLOCKCHAIN NETWORK TYPES

Property	***Public Blockchain***	***Consortium Blockchain***	***Private Blockchain***
Consensus decision	All miners in the network	Determinned nodes	The responsible authority
Consensus Process	Permissionless	permissioned	permissioned
Centralization	No	Partial	Yes
Immutability	not tampered	tampered	tampered
Permission to Read	Any one	Public or Restricted nodes	Public or Restricted nodes
Efficiency (use of resources)	Low	High	High
Advantages	- Independence -Transparency -Trust	-Access control -Scalability -Security	-Access control - high performance
Disadvantages	-Performance -Scalability -security	- Transparency	-Trust -Auditability

979-8-3503-0030-7/23/$31.00 ©2023 IEEE

Use Cases	- Asset ownership -Supply chain	-Banking -Research -Supply chain	

A. *The structure and process of Blockchain*

Blockchain is made up of blocks. As seen in Fig. 1, each block is linked to the blocks adjacent to it.

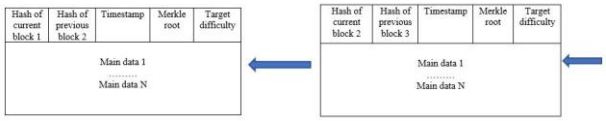

Fig. 1. Blockchain Architecture

The most important concepts in Blockchain are node, transaction, block, chain, miners, and consensus [3]. Fig. 2 depicts how a new block is added to the Blockchain network. An authorized participant sends a transaction, then the transaction is represented by the block. This block is distributed to all network nodes. An authorized node validates the transaction and uploads the new block to the Blockchain network, and this update is disseminated across the network.

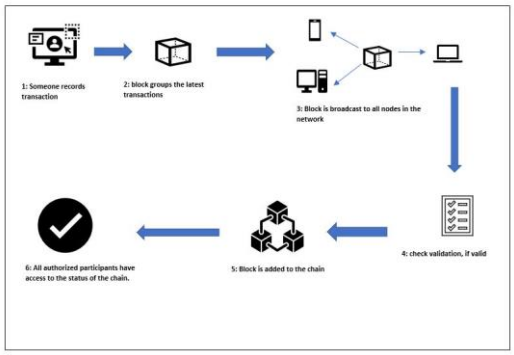

Fig. 2. Blockchain Process

B. *Consensus Algorithm*

To keep this ledger in a consistent state, all nodes in the Blockchain system should agree on a similar content-updating mechanism. Without majority permission, blocks should not be considered as part of the Blockchain. This is known as a consensus method, and it involves the creation of blocks that are then added to the current ledger for future usage. The consensus algorithm needs particular computation resources and manages the number of blocks to get activated by the node [4]. There are different consensus algorithms, each of them has advantages and disadvantages [5]. We can classify these algorithms according to the centralization that may lead to 51% attack. 51% attack means one of the miners has all the control or power of the Blockchain that moves Blockchain system from decentralization to centralization. Therefore, examples of the algorithms that can be exposed to centralization are: Proof of Work (PoW), Proof of Stake (PoS), Delegated Proof of Stake (DPoS), Delegated Byzantine Fault Tolerance (DBFT),s and Proof of Capacity (PoC). There are other algorithms that don't suffer from centralization problems such as Practical Byzantine Fault Tolerance (PBFT), Simplified Byzantine Fault Tolerance (SBFT), Proof of Elapsed Time (PoET), which is very decentralized, Proof of Authority, and Directed Acyclic Graphs (DAG).

TABLE II. shows the most popular consensus algorithm and the comparison between them in brief.

TABLE II. CONSENSUS ALGORITHMS

Consensus Algorithm	***Blockchain Platform***	***Smart contract***	***Pros***	***Cons***
PoW	Bitcoin	No	- Better Security -Less opportunity for 51% attack	-Centralization of Miners -greater energy consumption
PoS	NXT	Yes	-More decentralized - efficient for energy consumption	Nothing-at-stake problem
DPoS	Lisk	No	-Scalable, more secure -efficient for energy consumption	-Double spending attack - partially centralized.
DBFT	Lisk	No	-Scalable - more secure - efficient for energy consumption	-Double spending attack - partially centralized.
PoC	Burstcoin	Yes	-Cheap - efficient distributed.	Decentralizatio n issue
PBFT	Hyperledg er Fabric	Yes	-Reduction in energy	-Sybil Attack
SBFT	Chain	No	-Better validation Security	-unsuitable for public Blockchain
PoET	Hyperledg er Sawtooth	Yes	-Cheap	-unsuitable for public Blockchain
PoA	Decred	Yes	-Less probability of 51% attack	-Greater energy consumption.
DAG	IOTA	No	-Low cost for network scalability	-unsuitable for Smart Contracts.

C. *Blockchain Features*

The strength of Blockchain technology is due to its features that add value to all Blockchain-based electronic systems [6]. Its security, interoperability, data sharing, and data access features are critical needs for any information system, particularly in the big data industry. As shown in Fig. 3, Blockchain technology can meet all these objectives and more.

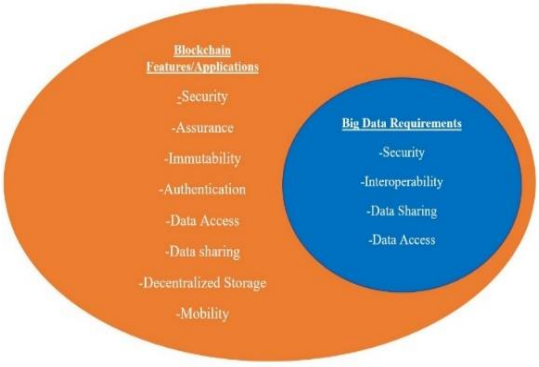

Fig. 3. Blockchain Features and Big Data Requirements

III. LITERATURE REVIEW

Blockchain technology has proven to be useful in a multitude of fields, particularly information security. The majority of Blockchain applications were employed to safeguard data in big data sector. However, owing to the scalability of Blockchain systems, various issues have been discovered when merging big data with Blockchain. Even in cloud Blockchain is used to achieve the security, privacy, reputation systems and transaction management as authors used in [7]. Blockchain's limited storage capacity causes it to be inconsistent with large amounts of data and affects its speed and performance. In TABLE III. we present the general problems that faces Blockchain application in big data fields.

TABLE III. BLOCKCHAIN PROBLEMS WITH BIG DATA

Reference	*Problem*	*Solution*
Mystiko—Blockchain Meets Big Data. [1]	There are several challenges to merging Blockchain with big data: 1) High transaction throughput is not supported by public Blockchain. 2) It is not scalable in terms of big data storage and administration. 3) There is no keyword search or retrieval capabilities.	"Mystiko" a novel Blockchain storage system based on the Cassandra distributed database, is proposed that can store massive volumes of data. "Mystiko" has the following characteristics: 1-High transaction throughput. 2-Extreme scalability. 3-Extremely high availability. 4-Full-text search capabilities.
Design of Personnel Big Data Management System Based on Blockchain [8]	Combining big data and Blockchain technology to solve a security problem to necessitates to addressing the Blockchain's limited storage capacity.	Data in blocks is separated into two categories: the Blockchain system stores core data, whereas the central database stores non-essential data.
Chain-based big data access control infrastructure. [9]	Big data storage generates certain unwanted overheads.	Propose a sovereign Blockchain based off-chain that offers a virtual container for parties to transact.

There are still ideas being presented and discussed to address these issues. We investigate if there is a strong trend to overcome the challenges that Blockchain applications face in large data systems. As a result, we may conclude that Blockchain technology is still in its development stages the application and research. Currently, no large-scale application scenario exists, particularly with big data. In the next section, we will look at many Blockchain application proposals in various sectors. We make it simpler for academics to comprehend how Blockchain may be used in each area and what steps must be taken next to complete the progress that Blockchain can achieve in each sector.

A. Blockchain Application on Big Data Security

1) Educational Field

Blockchain applications have many obstacles due to limited data storage capacity and potential security risks. Creating individual science credit data, a Taobao platform for intelligent education, a degree certificate system, and a new ecosystem of free educational resources are the four primary application modes of blockchain in the education sector[10].

2) Electronic Voting Field

Electronic voting has not been particularly effective in general, owing to security and privacy vulnerabilities discovered over time. In [11], the authors suggested a consortium Blockchain architecture using a hash algorithm and discussed the polling process's effectiveness, hashing techniques, utility block generation, sealing, data gathering, and result declaration. One of the electronic voting problems is centralization in voting schemes that are based on the blind signature FOO, mix-net, and homomorphic encryption technology. Because voting systems are recorded, controlled, computed, and verified centrally, these issues make electronic voting inefficient and untrustworthy. For large-scale voting, the authors in [12] employing Blockchain technology, ring signature and homomorphic ElGamal encryption.

3) Agriculture Field

Some issues have been discovered in agriculture systems, such as limited traceability and difficulties in government monitoring, making customers' trust in traceability outcomes harder. In [13][14], the authors presented a system based on Hyperledger for agricultural product traceability that securely stores traceability data. In addition to [14] the suggested system covers the whole agricultural products industrial chain and enabled customers to query the original source of agricultural product traceability.

4) Social Network Field

Blockchain technology is employed in all forms of social networks, for example, to secure users' data privacy and identify fake data. To address the Blockchain-limited size problem, in [15] the authors suggested a methodology for encrypting sensitive user data on a distributed Blockchain while sending non-sensitive data through the primary system. To transfer data, Vehicular Social Networks (VSNs) employ Ciphertext-policy attribute-based encryption (CP-ABE). Traditional CP-ABE methods store and grant access policy over the cloud, which loses credibility as a result of centralization. In [16], the authors described a secure and verifiable one-to-many data-sharing scheme based on Blockchain technology that records access regulations, allows user self-certification, and ensures cloud non-repudiation. When a vehicle user does not wish to exchange data in VSNs, this system also permits data revocation.

5) Healthcare Field

Blockchain technology's application in Healthcare systems is becoming increasingly important. This is due to the necessity of patient data privacy and data record exchange between hospitals and providers. To achieve privacy, the authors of [17] developed a platform that is totally controlled by the patient. This platform achieves personal data authorization and privacy. They built their platform using smart contracts and cryptographic mechanisms. According to [18], Blockchain is being utilized in the Healthcare sector to store and safeguard big data with multiple stakeholders. They created a framework that integrates IoT, Machine Learning and Blockchain for this aim. IoT was utilized to restore data from the wearable

devices of patients. Blockchain with Proof of Stake (PoS) was used to ensure the pseudo-anonymity of the patient's identity while still providing verified and trustworthy data. To detect anomalies and predict certain scenarios, machine learning was applied.

As highlighted in [19], the Estonian government has undertaken a Blockchain-based effort to protect Electronic Health Records, which are accessible to 95%-99% of Estonian patients. The UAE has announced the creation of a Blockchain platform for healthcare data storage. Swiss hospitals have also deployed a Blockchain-based system to track medical items safely and effectively.

Although Blockchain technology is appropriate for a wide range of IoT applications, it does have significant restrictions. In Healthcare systems, the incorporation of Blockchain technology with IoT has faced several challenges, including the high processing power of the PoW consensus method, poor scalability, and significant transaction confirmation latency. As a result, in [20], the authors introduced a transactional GHOSTDAG protocol for remote patient monitoring. GHOSTDAG is promoted as a solution to Healthcare security challenges that does not compromise scalability.

To explore the literature related to Blockchain, we located fifty studies on Blockchain covering survey studies and application studies in different domains. In our study, we focus on applications studies and summarize the most important studies in different domains. From these studies, the Healthcare field obtained or won the largest share of research interests for Blockchain applications to solve different issues that Healthcare systems suffer from. Then comes the field of IoT. Fig. 4 shows the percentages of different themes in application areas for Blockchain.

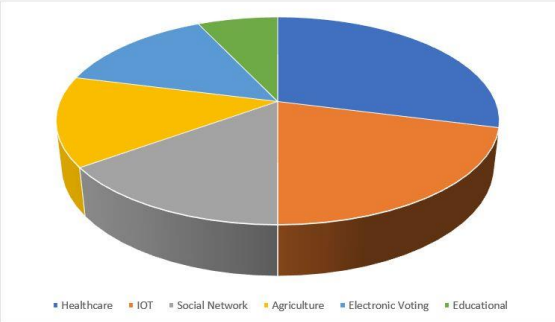

Fig. 4. Blockchain Application areas

IV. Healthcare Research Study

In this section, we discuss the most recent research studies that use Blockchain technology to solve challenges in healthcare systems. Healthcare systems have several security challenges, which has pushed Blockchain technology to the top of the list of options for healthcare systems. Fig. 5 shows these issues summarized in data security, privacy, integrity, and interoperability problems.

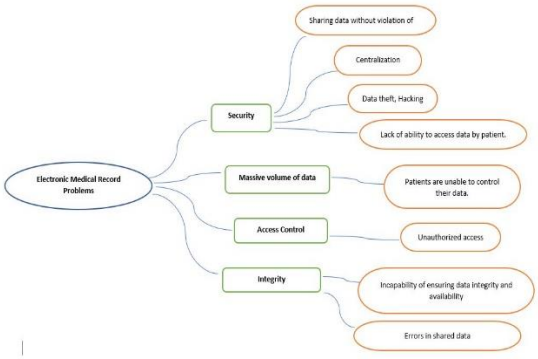

Fig. 5. Electronic Record problems

Most research papers concentrated on security and privacy concerns. As in [3], the authors discussed how consortium Blockchains like MedChain, ModelChain, and BlockInsure are better suited for health information security. Blockchain lowers duplication and offers carers reliable patient records.

In [21] The authors investigated vulnerabilities on several levels of Healthcare systems and theoretically shown that Blockchain technology can solve Healthcare 4.0 privacy and security issues.

Healthcare data violations include 6% data loss, 10% unknown data, 10% data theft, 12% unauthorized access or disclosure, and 62% hacking /IT incidents. In [22] The authors demonstrated the research efforts made to manage the integrity of health data. The Blockchain was suggested as the most urgent data integrity method.

In [23] The authors presented a Blockchain-based Self-Sovereign Identity (SSI) architecture for healthcare. It gives users total control over their personally identifiable information across many authorities.

In [24] The authors propose an architecture based on Blockchain to give access control and safeguard the privacy of patients. They created a paradigm based on decreasing data redundancy by clustering miners, which reduces network overhead and makes transactions small enough to send via the Blockchain.

As shown in [25], there are two major security challenges arose in Electronic Health Systems (EHS). The first is the selecting of an access control method without the knowledge of previous users. The second factor is the volume of data that the Healthcare practitioner will share. The authors developed an access control model that is based on the amount of confidence between the user and the Healthcare service.

In [26] the authors investigate the problem of authorization and access control in a cloud-based Healthcare system. They suggest the Blockchain as a solution.

In [27] authors categorized Blockchain applications in Healthcare into four primary categories. The first category is an enhancement in Electronic Medical Record (EMR) administration for data duplication avoidance, scalability, and security. The second technique was to use smart contracts to enhance insurance claim procedures, such as smart health profiles and the exchange of medical information. The third goal was to make it easier for researchers to share data. The fourth category included applications that allowed private interactions between

customers and carers. In [28], The authors provided an overview of several Blockchain-based Healthcare applications. The MedRec application is a record management tool for medical access that allows patients to see their medical records, audit their care, and share data. The BlockHie data sharing concept is used to exchange Electronic Medical Records and Personal Healthcare Data.

To give patients access to their medical information, the authors in [29] presented a smart contract approach. To securely store, retrieve, and exchange patients' medical data, they used Interplanetary File Systems (IPFS) and trusted reputation-based re-encryption oracles.

According to the results of our poll, there is still a concern with patients' privacy and how to limit their data access in EMR. Access control is the most difficult challenge, accounting for 17% of all challenges.

V. Conclusion

Blockchain is a distributed and decentralized ledger technology. It is used to conduct transactions securely and to eliminate the Single Point of Failure. In this paper, we present a brief systematic overview of the Blockchain network and its consensus algorithms. The paper demonstrates the importance of Blockchain technology in several domains and how researchers and communities can benefit from Blockchain technology. We focus our study on its applications in Healthcare systems that still suffer from security and privacy issues. Achieving access control for patients over their health data is the current research point. In future work, we will propose our system architecture scenario for solving the problem of access control for patients and evaluate it by comparing with the mentioned proposed solutions.

References

[1] Bandara, E., Ng, W. K., De Zoysa, K., Fernando, N., Tharaka, S., Maurakirinathan, P., & Jayasuriya, N. (2018, December). Mystiko—blockchain meets big data. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 3024-3032). IEEE.

[2] Satoshi Nakamoto, "A Peer-to-Peer Electronic Cash System," 2020, www.bitcoin.org.

[3] Z. Alhadhrami, S. Alghfeli, M. Alghfeli, J. A. Abedlla, and K. Shuaib, "Introducing blockchains for healthcare," 2017 Int. Conf. Electr. Comput. Technol. Appl. ICECTA 2017, vol. 2018-Janua, pp. 1–4, 2017.

[4] Mohanty, S. N., Ramya, K. C., Rani, S. S., Gupta, D., Shankar, K., Lakshmanaprabu, S. K., & Khanna, A. (2020). An efficient Lightweight integrated Blockchain (ELIB) model for IoT security and privacy. *Future Generation Computer Systems*, *102*, 1027-1037.

[5] E. Indhuja and M. Venkatesulu, *A Survey of Blockchain Technology Applications and Consensus Algorithm*, vol. 55. Springer Singapore, 2021

[6] T. McGhin, K. K. R. Choo, C. Z. Liu, and D. He, "Blockchain in healthcare applications: Research challenges and opportunities," *J. Netw. Comput. Appl.*, vol. 135, no. February, pp. 62–75, 2019.

[7] Xie, S., Zheng, Z., Chen, W., Wu, J., Dai, H. N., & Imran, M. (2020). Blockchain for cloud exchange: A survey. *Computers & Electrical Engineering*, *81*, 106526.

[8] Chen, J., Lv, Z., & Song, H. (2019). Design of personnel big data management system based on blockchain. *Future Generation Computer Systems*, *101*, 1122-1129.

[9] Sifah, E.B., Xia, Q., Agyekum, K.OB.O. et al. (2018). Chain-based big data access control infrastructure. *The Journal of Supercomputing*, *74*(10), 4945-4964.

[10] Liu, Q., & Zou, X. (2019). Research on trust mechanism of cooperation innovation with big data processing based on blockchain. *EURASIP Journal on Wireless Communications and Networking*, *2019*(1), 1-11.

[11] Shahzad, B., & Crowcroft, J. (2019). Trustworthy electronic voting using adjusted blockchain technology. *IEEE Access*, *7*, 24477-24488.

[12] Wang, B., Sun, J., He, Y., Pang, D., & Lu, N. (2018). Large-scale election based on blockchain. *Procedia Computer Science*, *129*, 234-237.

[13] Kamilaris, A., Fonts, A., & Prenafeta-Boldú, F. X. (2019). The rise of blockchain technology in agriculture and food supply chains. *Trends in Food Science & Technology*, *91*, 640-652.

[14] Wang, Z., & Liu, P. (2019, July). Application of Blockchain Technology in Agricultural Product Traceability System. In *International Conference on Artificial Intelligence and Security* (pp. 81-90). Springer, Cham.

[15] Chen, Y., Xie, H., Lv, K., Wei, S., & Hu, C. (2019). DEPLEST: A blockchain-based privacy-preserving distributed database toward user behaviors in social networks. *Information Sciences*, *501*, 100-117

[16] Fan, K., Pan, Q., Zhang, K., Bai, Y., Sun, S., Li, H., & Yang, Y. (2020). A Secure and Verifiable Data Sharing Scheme Based on Blockchain in Vehicular Social Networks. *IEEE Transactions on Vehicular Technology*

[17] Al Omar, A., Bhuiyan, M. Z. A., Basu, A., Kiyomoto, S., & Rahman, M. S. (2019). Privacy-friendly platform for healthcare data in cloud based on blockchain environment. *Future generation computer systems*, *95*, 511-521.

[18] Chakraborty, S., Aich, S., & Kim, H. C. (2019, February). A secure healthcare system design framework using blockchain technology. In *2019 21st International Conference on Advanced Communication Technology (ICACT)* (pp. 260-264). IEEE.

[19] I. Yaqoob, K. Salah, R. Jayaraman, and Y. Al-Hammadi, "Blockchain for healthcare data management: opportunities, challenges, and future recommendations," *Neural Comput. Appl.*, vol. 0123456789, 2021.

[20] Srivastava, G., Crichigno, J., & Dhar, S. (2019, May). A light and secure healthcare blockchain for iot medical devices. In *2019 IEEE Canadian conference of electrical and computer engineering (CCECE)* (pp. 1-5). IEEE.

[21] J. J. Hathaliya and S. Tanwar, "An exhaustive survey on security and privacy issues in Healthcare 4 . 0," *Comput. Commun.*, vol. 153, no. January, pp. 311–335, 2020.

[22] A. K. Pandey *et al.*, "Key Issues in Healthcare Data Integrity: Analysis and Recommendations," *IEEE Access*, vol. 8, pp. 40612–40628, 2020.

[23] M. Shuaib, S. Alam, M. Shabbir Alam, and M. Shahnawaz Nasir, "Self-sovereign identity for healthcare using blockchain," *Mater. Today Proc.*, 2021.

[24] K. M. Hossein, M. E. Esmaeili, T. Dargahi, and A. Khonsari, "Blockchain-Based Privacy-Preserving Healthcare Architecture," *2019 IEEE Can. Conf. Electr. Comput. Eng. CCECE 2019*, pp. 1–4, 2019.

[25] A. Singh and K. Chatterjee, "An adaptive mutual trust based access control model for electronic healthcare system," *J. Ambient Intell. Humaniz. Comput.*, vol. 11, no. 5, pp. 2117–2136, 2020.

[26] L. J. Kittur, R. Mehra, and B. R. Chandavarkar, *The Dependency of Healthcare on Security: Issues and Challenges*, vol. 698. Springer Singapore, 2021.

[27] A. A. Mazlan, S. M. Daud, S. M. Sam, H. Abas, S. Z. A. Rasid, and M. F. Yusof, "Scalability Challenges in Healthcare Blockchain System-A Systematic Review," *IEEE Access*, vol. 8, pp. 23663–23673, 2020.

[28] H. Rathore, A. Mohamed, and M. Guizani, *Blockchain applications for healthcare*. Elsevier Inc., 2020.

[29] M. M. Madine *et al.*, "Blockchain for Giving Patients Control over Their Medical Records," *IEEE Access*, vol. 8, pp. 193102–193115, 2020.