Federated Learning Meets Blockchain in Decentralized Data Sharing: Healthcare Use Case

Saeed Hamood Alsamhi, Raushan Myrzashova, Ammar Hawbani, Santosh Kumar, *Member, IEEE*, Sumit Srivastava, Liang Zhao, Xi Wei, Mohsen Guizan, *Fellow, IEEE*, and Edward Curry

***Abstract*—In the era of data-driven healthcare, the amalgamation of blockchain and federated learning (FL) introduces a paradigm shift toward secure, collaborative, and patient-centric data sharing. This article pioneers the exploration of the conceptual framework and technical synergy of FL and blockchain for decentralized data sharing, aiming to strike a balance between data utility and privacy. FL, a decentralized machine learning paradigm, enables collaborative AI model training across multiple healthcare institutions without sharing raw patient data. Combined with blockchain, a transparent and immutable ledger, it establishes an ecosystem fostering trust, security, and data integrity. This article elucidates the technical foundations of FL and blockchain, unravelling their roles in reshaping healthcare data sharing. This article vividly illustrates the potential impact of this fusion on patient care. The proposed approach preserves patient privacy while granting healthcare providers and researchers access to diversified data sets, ultimately leading to more accurate models and improved diagnoses. The findings underscore the potential acceleration of medical research, improved treatment outcomes, and patient empowerment through data ownership. The synergy of FL and blockchain envisions a healthcare ecosystem that prioritizes individual privacy and propels advancements in medical science.**

***Index Terms*—Blockchain, data sharing, Dataspace 4.0, decentralized data sharing, federated learning (FL), healthcare, Industry 4.0, Industry 5.0, IoE.**

Manuscript received 27 November 2023; revised 5 February 2024; accepted 13 February 2024. Date of publication 19 February 2024; date of current version 23 May 2024. This work was supported by the Science Foundation Ireland under Grant SFI/12/RC/2289_P2. The work of Raushan Myrzashova was suuported by the ANSO Scholarship for Young Talents. The work of Xi Wei was supported by the Natural Science Foundation of Anhui Province under Grant BJ2060000039. *(Corresponding author: Saeed Hamood Alsamhi.)*

Saeed Hamood Alsamhi is with the Insight Centre for Data Analytics, University of Galway, Galway, H91 TK33 Ireland, and also with the Faculty of Engineering, IBB University, Ibb, Yemen (e-mail: Saeed.alsamhi@insight-centre.org).

Raushan Myrzashova is with the School of Computer Science and Technology, University of Science and Technology of China, Hefei 230026, Anhui, China (e-mail: rose5004@mail.ustc.edu.cn).

Ammar Hawbani and Liang Zhao are with the School of Computer Science, Shenyang Aerospace University, Shenyang 110136, China (e-mail: anmande@ustc.edu.cn; lzhao@sau.edu.cn).

Santosh Kumar is with the Department of Computer Science and Engineering, International Institute of Information Technology-Naya Raipur, Atal Nagar-Nava Raipur 493661, India (e-mail: santosh@iiitnr.edu.in).

Sumit Srivastava is with the Department of Electronics and Communication Engineering, FET, MJP Rohilkhand University, Bareilly 243001, India (e-mail: sumitsrivastava@mjpru.ac.in).

Xi Wei is with the Department of Chemistry, University of Science and Technology of China, Hefei 230026, Anhui, China (e-mail: wxi@ustc.edu.cn).

Mohsen Guizan is with the Machine Learning Department, Mohamed Bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE (e-mail: mguizani@ieee.org).

Edward Curry is with the Insight Centre for Data Analytics, University of Galway, Galway, H91 TK33 Ireland (e-mail: edward.curry@insight-centre.org).

Digital Object Identifier 10.1109/JIOT.2024.3367249

I. Introduction

THE RAPID development of the Internet of Things (IoT), cloud computing, and big data has led to Dataspace 4.0, a digital ecosystem where massive amounts of data from various sources are seamlessly integrated and shared among stakeholders. Dataspace 4.0, funded by the European Union, aims to establish shared principles for exchanging manufacturing data at the EU level; Dataspace 4.0 is to pave the way for a unified manufacturing data ecosystem and foster the formation of a cohesive European community focused on Dataspace 4.0 [1]. Therefore, data sharing is essential in Dataspace 4.0 to create a coherent European community and a unified industrial data environment. With the advent of the sixth generation (6G), the capabilities of Dataspace 4.0 are expected to be further enhanced, providing new opportunities for data-driven applications and services. Dataspace 4.0 refers to the next generation of data management systems expected to enable the integration and sharing of data across various industries and domains [2]. Varga et al. [3] discussed how advanced technologies and the needs set for 6G affect Industry 4.0 developments based on massive data. The foundation of Industry 4.0 is data sharing, which facilitates smooth communication between entities, machines, and processes, improving operational excellence, decision making, and resource usage. Furthermore, Han et al. [4] provided a vision for a 6G industrial digital twin (DT) ecosystem to bridge the gaps between machines, humans, and data infrastructure to enable numerous applications. As a result, data sharing is essential to achieving the full potential of Industry 4.0 and Dataspace 4.0, not merely necessary.

The safe and ethical sharing of private patient data is a crucial challenge when healthcare data is expanding exponentially, and there is an increasing demand for data-driven medical advancements. Healthcare institutions, researchers, and patients need to strike a delicate balance between the utility of aggregated medical data for scientific progress and the paramount importance of preserving individual privacy and data security. The challenge has spurred the emergence of innovative technologies poised to reshape the landscape of healthcare data sharing. Data sharing has become an essential component of modern society, enabling businesses,

governments, and individuals to access and analyze vast amounts of data for various purposes, such as research, decision making, and innovation. However, centralized data-sharing systems have limitations, such as data privacy and security issues [5], interoperability issues [6], and single points of failure [7]. To address these challenges, decentralized data sharing has emerged as a promising alternative that distributes data across multiple nodes or peers without needing a central authority or intermediary. In addition, decentralized data sharing offers several benefits, such as increased privacy and security, improved data ownership and control, and enhanced transparency and accountability [8].

Decentralized data sharing is an essential aspect of Dataspace 4.0, as it allows multiple parties to share data without needing a central authority or intermediary [9], leading to improved collaboration, increased data privacy and security, and the potential for new business models and revenue streams. Several decentralized data-sharing technologies and techniques, such as federated learning (FL) [10] and blockchain [11], have emerged as promising solutions to address these challenges. The technologies above have been applied in various domains, such as healthcare, finance, and the IoT, to address specific use cases and requirements. Two such technologies, FL and blockchain, have garnered significant attention for their potential to solve this conundrum. FL, a decentralized machine learning (ML) approach that Google pioneered [12] offers a novel paradigm for collaborative model training across a network of data sources without centralizing raw data. It inherently safeguards data privacy at its source, a crucial factor in healthcare, where data confidentiality is sacrosanct [13]. Initially developed as the underlying technology for cryptocurrencies like Bitcoin [14], blockchain has transcended its financial origins to become a secure and immutable ledger capable of ensuring data integrity and transparency. Its characteristics are well suited to address the need for trust and accountability in data-sharing ecosystems [15]. Despite the potential benefits of decentralized data sharing, several challenges and limitations are associated with the above technologies, such as scalability, interoperability, and regulatory compliance.

In this article, we explore the intersection between FL and blockchain in the context of decentralized data sharing, with a particular focus on the healthcare sector. Our objective is to unravel the synergies between these two technologies, shedding light on how they can be harnessed to revolutionize healthcare data sharing while preserving individual privacy and fostering collaboration. The significance of this article extends beyond theoretical exploration and embraces practical implications for healthcare institutions, researchers, and, ultimately, patients. The combination of blockchain technology and FL has become a game-changer in the quickly developing field of data-driven technologies, providing a fresh approach to decentralized data sharing. In the context of a decentralized data-sharing framework, this article examines the synergies between these two technologies, highlighting how they could transform collaborative data sharing while protecting individual privacy and promoting smooth collaboration.

A. Motivation and Contributions

Modern societies depend on data sharing because it promotes cooperation, spurs innovation, and increases industry transparency [16]. Although it is essential to research, development, and the welfare of society, the explosion in data generation—especially since the introduction of the 6G network and the spread of the IoT—brings new difficulties. Once shared, centralized data-sharing solutions now have privacy, security, and accessibility issues. To overcome these obstacles, this article proposes a paradigm shift toward decentralized data sharing by utilizing blockchain technology and FL. The synergy of blockchain and FL strategy guarantees enhanced security, privacy and a strong barrier against unwanted access and possible data breaches. Furthermore, It offers protection from changing cyber threats by sharing power and leveraging blockchain's advantages.

Moreover, the synergy of FL and blockchain gives stakeholders unparalleled control over data in addition to security [17]. It creates an environment of trust and accountability among players by protecting intellectual property rights and promoting openness. At the vanguard of transforming healthcare data exchange, the synergy strategy goes beyond satisfying urgent needs. Safe, effective, patient-centered data sharing will speed up medical research, enhance patient care and accelerate improvements in healthcare. Our proposed paradigm stands out for resolving the conventional tradeoff between privacy and data sharing. Not only does it comply with strict regulations, but it also dramatically increases productivity and openness in the healthcare industry. In addition to providing a comprehensive solution, our work establishes a new benchmark for the interchange of healthcare information. The combination of blockchain technology and FL promises to transform the healthcare industry by promoting scientific breakthroughs, enhancing patient care, and guaranteeing legal compliance.

Data sharing is pivotal in shaping modern societies, offering myriad benefits that span individuals, organizations, and the broader community [16]. It fosters collaboration, drives efficiencies, and fosters innovation across various sectors. Data sharing enhances transparency and accountability, acting as a bulwark against corruption and building trust among stakeholders [18]. It also streamlines resource utilization, leading to significant cost savings and productivity gains. In public services, data sharing catalyzes research and development, particularly in critical areas like healthcare, environmental conservation, and societal well-being. However, the landscape of data sharing is not without its complexities. With the proliferation of the IoT and the advent of the 6G network, there has been an exponential increase in data generation, presenting both opportunities and challenges. Data sharing in this context raises significant privacy, security, and interoperability concerns, necessitating a careful balance between innovation and risk mitigation. Centralized data-sharing models, traditionally prevalent, are increasingly seen as inadequate due to their inherent privacy and security limitations, reliance on singular management entities, and accessibility challenges. This article argues for a shift toward decentralized data sharing,

utilizing FL and blockchain technology. Such a decentralized approach leverages distributed computing for efficiency and scalability while harnessing blockchain's strengths in immutability and security. This method promises enhanced security and privacy, mitigating risks like unauthorized access and data breaches. It also empowers stakeholders by granting greater control over data, fostering transparency, and safeguarding intellectual property rights. Additionally, it promotes interoperability and seamless data exchange, thereby reducing fragmentation and improving collaboration.

Our work is at the forefront of reshaping healthcare data sharing by exploring the synergistic potential of FL and blockchain technologies. Our approach addresses the critical needs of secure, efficient, and patient-centric healthcare data sharing in a world increasingly driven by data. We propose an innovative framework that enables healthcare institutions, researchers, and patients to share data securely and efficiently. This approach not only enhances patient care and accelerates medical research but also promises greater accuracy in diagnoses, personalized treatment options, and rapid advancements in medical science. The primary driving force behind our work is the need to bridge the gap between collaborative healthcare research and the imperative to protect patient data privacy. Our proposed decentralized data-sharing model effectively resolves the traditional tradeoff between sharing and privacy. It aligns with stringent regulatory requirements while boosting efficiency, transparency, and trust in the healthcare sector. The main contributions of this article are encapsulated in the development of a groundbreaking, patient-centric framework for healthcare data sharing in the 6G era, integrating FL and blockchain technologies. This integration is poised to revolutionize the healthcare landscape, fostering advancements in research, improving patient care, and ensuring regulatory compliance, all while maintaining a steadfast focus on patient privacy. We offer a comprehensive solution to decentralized data sharing, setting a new standard in healthcare information exchange.

1) *Innovative Integration of Technologies:* We propose a novel framework combining FL and blockchain for healthcare data sharing. This synergy addresses complex challenges related to data security and efficient sharing, ensuring a patient-centric approach.
2) *Privacy-Preserving Data-Sharing Model:* Our work introduces a privacy-preserving framework for healthcare data sharing. By amalgamating FL's capability to train models without exposing raw data and blockchain's strength in maintaining data integrity, we ensure patient information's confidentiality and immutability.
3) *Empowerment of Patients in Data Sharing:* The proposed model enhances patient empowerment by allowing them to maintain control over their healthcare data. Blockchain technology enables patients to participate in medical research while actively preserving data ownership.
4) *Healthcare Use Case Application:* We demonstrate the practical applicability of our framework through a healthcare use case. This approach leads to more accurate medical models, personalized treatment options, and improved patient care, revolutionizing healthcare.
5) *Bridging the Privacy-Utility Gap in Healthcare:* Our framework addresses the critical balance between collaborative healthcare research and patient data privacy, aligning with stringent regulatory standards and enhancing transparency and trust in healthcare data sharing.

B. Related Work

Industry 4.0 is characterized by integrating several cutting-edge technologies, such as the Industrial IoT, artificial intelligence (AI)—including augmented intelligence, big data analytics, ML, and deep learning (DL)—and edge–fog cloud computing. These technologies are driving the next phase of digital transformation [28], [29], [30]. However, unlocking the full potential of IIoT requires cross-company collaboration, such as multiparty computation, pooled analyses, data sharing, and data exchanging within a network of collaborators or organizations, which is essential to overcome the significant fragmentation of data. Integrating FL, blockchain technology, and healthcare data sharing has been an increasing interest and research area. Numerous studies have examined the technologies individually and in conjunction to address the pressing challenges of healthcare data privacy, security, and collaborative research. Table I summarizes the comparison of existing related work.

FL in Healthcare: FL allows multiple parties to train an ML model collaboratively without sharing raw data. Liu et al. [31] proposed an FL-based approach for decentralized data sharing in the IIoT. The authors showed that their approach achieved better accuracy and reduced communication overhead compared to traditional centralized learning. However, FL still faces challenges, such as the privacy–utility tradeoff and communication efficiency [32]. The combination of homomorphic encryption and FL enables privacy-preserving healthcare data analysis, demonstrating the feasibility of collaborative model training without exposing sensitive patient data [13]. The challenges, methods, and prospects, including their applications in the healthcare domain, are discussed in [33] and [34]. Moreover, FL is a privacy-preserving paradigm in healthcare, emphasizing its potential in medical research and the development of diagnostic models [35].

Blockchain in Healthcare: Blockchain is a decentralized and tamper-proof ledger that records transactions and stores data securely and transparently. Blockchain has been proposed as a potential solution for decentralized data sharing due to its ability to provide data immutability, auditability, and transparency. Makhdoom et al. [36] proposed a blockchain-based decentralized data-sharing framework that addressed data privacy and security concerns. Blockchain's relevance in healthcare has been extensively investigated. Chen et al. [15] examined the patient-centric blockchain model in healthcare, highlighting its capacity for secure and transparent health data management and sharing. Fatima et al. [37] provided a comprehensive review of blockchain's role in healthcare privacy and data security, focusing on its applications in

TABLE I
COMPARISON OF EXISTING RELATED WORK

Ref.	highlighte	Applications	Domains
[19] (2020)	Decentralized tourism destinations recommendation system	Tourism	Blockchain, data-sharing
[20](2020)	Improving interorganizational information sharing for vendor managed inventory	Supply chain management	Blockchain, vendor-managed inventory
[21] (2019)	Building a secure biomedical data-sharing decentralized app	Biomedical research	Blockchain, data-sharing
[22] (2022)	Decentralized congestion control methods for vehicular communication	Vehicular networks	Blockchain, congestion control
[23](2021)	Decentralized trusted data-sharing management on IoVEC networks	Internet of Vehicle Edge Computing	Blockchain, data-sharing
[11] (2020)	Decentralized data-sharing infrastructure for off-grid networking	Off-grid networking	Blockchain, data-sharing
[24] (2019)	Framework of data-sharing system with decentralized network	General data-sharing	Blockchain, data-sharing
[25] (2017)	P2P platform for decentralized	Logistics	Peer-to-peer, decentralized logistics
[26] (2022)	Decentralized network secured data-sharing	General data-sharing	Blockchain, data-sharing
[27] (2020)	Unlocking the potential of AI in assisted reproduction	Assisted reproduction	Blockchain, AI, data-sharing

patient records, clinical trials, and supply chain management. Additionally, Fan et al. [38] explored secure multiparty computations using blockchain, with implications for privacy-preserving distributed prediction in healthcare analytics.

Integration of FL and Blockchain: While significant progress has been made in investigating FL and blockchain individually in healthcare, a notable gap in research exploring their synergistic potential exists. This article represents a pioneering effort to integrate these technologies specifically for decentralized healthcare data sharing. Our integration aims to harness the advantages of both approaches, such as FL's data privacy preservation and blockchain's data integrity, to address the challenges faced by traditional healthcare data-sharing methods.

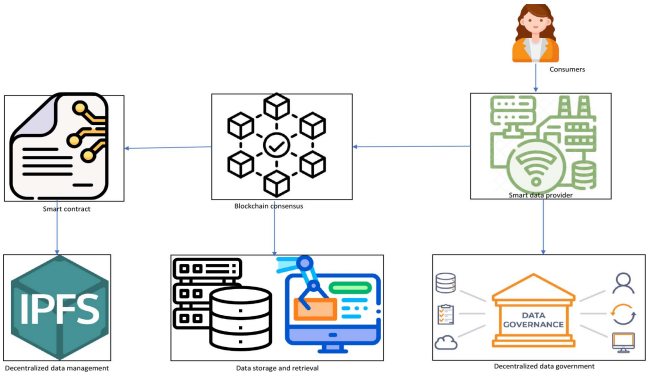

Fig. 1. Decentralized data sharing using blockchain.

II. OVERVIEW

A. Decentralized Data Sharing

Decentralized data sharing refers to distributing data across a network of independent participants rather than relying on a centralized authority to manage and control access to the data. In a decentralized data-sharing system, each participant has a copy of the data and is responsible for maintaining and updating their copy. In addition, participants share data with other participants, either directly or through a P2P network, and access data shared by other participants. Decentralized data sharing is designed with security and privacy in mind to protect against data breaches and unauthorized access to sensitive information. Decentralized data sharing involves encryption, access controls, and other security measures to safeguard the data [39].

Decentralized data sharing represents a groundbreaking departure from traditional data-sharing approaches, offering many compelling advantages. Primarily, it fortifies data security through its distributed structure, rendering it resistant to targeted cyber-attacks or data breaches [40], [41]. Unlike centralized systems, where all data resides in a single location vulnerable to hacking [42], decentralized data sharing scatters data across a network of nodes, bolstering protection measures with encryption and access controls. Each node possesses a private key [43], ensuring only intended recipients can access shared data, even if the network is compromised. Furthermore, consensus algorithms verify data accuracy [44], fortifying security and control over data access. Second, decentralized data sharing empowers individuals with heightened data privacy control. It eliminates the need for a central authority to manage data access, permitting individuals to grant access exclusively to trusted parties. Within this framework, data is distributed across nodes, safeguarded by cryptography. Each entity holds a private key for data encryption and decryption, assuring data privacy and thwarting unauthorized access. This approach significantly augments personal data privacy and control, aligning with contemporary demands for robust privacy measures to enable requirements of Industry 4.0 toward Industry 5.0 [45]. Fig. 1 illustrates the architecture of decentralized data-sharing using blockchain, including components such as smart contracts, blockchain databases, and data governance mechanisms.

Additionally, decentralized data sharing improves interoperability across diverse systems and organizations. It achieves this by embracing open standards and protocols that streamline data sharing among distinct platforms and applications. The result is reduced inefficiencies, redundancies, and delays in data exchange, facilitating seamless collaboration and resource optimization. Moreover, this decentralized approach enhances transparency by allowing all parties to access and validate shared data, cultivating trust and collaborative potential. Decentralized data-sharing systems use encryption to protect the data from unauthorized access or tampering. Each node in the network has a private key used to encrypt and decrypt data,

TABLE II
COMPARING CENTRALIZED AND DECENTRALIZED DATA SHARING

Items	Centralization	Decentralization
Data control	Controlled by a single organization or authority	Distributed across multiple nodes
Security	Centralized control creates security risks	Distributed network of nodes improves resilience
Privacy	Centralized control creates privacy concerns	Encryption, and smart contracts enhance privacy
Interoperability	Limited interoperability	Improved interoperability with the use of decentralized standards and protocols
Transparency	Limited transparency and accountability	Tamper-proof and transparent record of data-sharing activities

ensuring that only the intended recipient can access the data to prevent unauthorized access to the data and provide a greater level of security for the data. Therefore, decentralized data sharing improves resilience by creating a distributed network of nodes that continue to operate even if some nodes fail or are compromised and by using encryption to protect the data from unauthorized access or tampering, leading to a reduction of the risks associated with data sharing and enabling organizations to work more effectively and efficiently [46]. Table II outlines the differences between centralized and decentralized data-sharing, emphasizing the superior resilience, privacy, and interoperability of the latter.

B. Blockchain

Blockchain technology is a formidable decentralized and distributed data-sharing solution renowned for its robust security and transparency features. Functioning as a ledger system, it organizes data into immutable and chronological blocks, authenticated through consensus mechanisms among a network of nodes, ensuring its accuracy and timeliness [47]. The successful implementation of a blockchain-based decentralized data-sharing system hinges on several key considerations. It must accommodate substantial data volumes and transactions, necessitating high scalability and performance. Robust security measures, including encryption and tamper-proofing, are vital to data integrity and confidentiality. Additionally, the versatility to support various applications and use cases, spanning financial transactions, supply chain management, and digital identity verification, is paramount [14]. Blockchain's essential attributes position it as a pivotal player in the evolution of data-sharing systems, such as Dataspace 4.0 and 6G, offering a pathway to highly secure, efficient, and transparent decentralized data-sharing platforms [45], [48].

Blockchain technology's prowess extends to enhancing interoperability in decentralized data sharing, offering a unified framework for secure and effective interaction among diverse systems and organizations. Blockchain-based systems facilitate secure data exchange while preserving data integrity using common data structures and cryptographic algorithms. Transparency, another hallmark feature of blockchain, ensures all participants maintain a shared, comprehensive view of data and its historical changes. It achieves this through a distributed ledger, creating an immutable, transparent record of all data transactions. This heightened transparency fosters trust among parties, promotes accountability, and ensures compliance. Furthermore, blockchain's resilience factor is crucial in decentralized data sharing, guaranteeing data availability despite system failures or network disruptions. Advanced consensus mechanisms bolster this resilience, rendering the system less susceptible to malicious attacks or data breaches [49]. Blockchain's multifaceted potential is vividly evident in various industries, including supply chain management, healthcare, and financial services. It offers secure and transparent data recording and sharing capabilities, enhances efficiency, accountability, and transparency, and presents novel solutions to industry-specific challenges. While blockchain holds immense promise, it is essential to acknowledge and address challenges, such as scalability, energy consumption, and regulatory frameworks, to fully harness its potential for decentralized data sharing across a spectrum of applications [49].

Every node in a decentralized blockchain network has a copy of the ledger. A new transaction is announced to the network whenever one is proposed. The transaction is then independently verified by nodes using pre-established protocols and regulations. The consensus process's primary goal is to reach a consensus over the ledger's current status. This keeps any one node from intentionally or mistakenly changing the blockchain by requiring all nodes to verify and concur on the sequence and legitimacy of transactions. Every node in the network is equal and cooperates to keep the blockchain current. These nodes divide up the transaction processing, including consensus-building and validation. Blockchain's decentralization guarantees that no single entity controls the network. Rather, a democratic consensus is reached among nodes through the consensus process. To enhance security and resilience, no single organization can dictate changes to the blockchain. Blockchain technology's core feature is the distribution of processing among the network of nodes. It guarantees that the system is resilient to attacks, strong, and able to unite different people when trust is lacking. In conclusion, the blockchain's distributed bulk processing site highlights the decentralized character of consensus processes across the nodes. This decentralized processing enhances the blockchain's security, transparency, and reliability.

C. Federated Learning

FL presents an innovative approach to ML that prioritizes collaborative model training while preserving data privacy and security [50]. In this decentralized paradigm, each participating entity retains its data on its local device or network, eliminating the need to transmit sensitive information to a centralized repository. FL operates by having each participant train an ML model using their local data and sending model updates to a central server, reflecting the parameter differences post-training. The server aggregates these updates from all participants, typically through algorithms like averaging

or median computation. The central server then returns an updated global model to each participant. This iterative process of local training, update transmission, and model retrieval continues until the global model reaches an acceptable level of accuracy or satisfies other predetermined criteria. The inherent structure of FL facilitates participation from multiple parties in the ML process without necessitating the sharing of raw data. Utilizing a standardized ML model across all participants ensures consistent application, ultimately leading to a more accurate global model. However, the effectiveness of FL relies heavily on a robust communication infrastructure for efficient model exchange between participants and the central server. Weak infrastructure or connectivity can delay model updates and compromise learning processes [51].

FL significantly augments data security and privacy by retaining sensitive information locally, thereby reducing the risk of data breaches during transmission [51]. Since raw data remains on local devices, potential attackers face formidable challenges accessing sensitive information. Compromising multiple devices to reconstruct a complete data set is considerably more complex than targeting a single centralized server. Moreover, FL's design ensures only model updates, typically aggregated and abstracted information, are transmitted to the central server. These updates do not reveal the raw data from which they were derived, further fortifying data privacy [52]. Regarding privacy preservation, FL guarantees that user data remains private by avoiding central server sharing. Data remains confined to each user's device, rendering it inaccessible to third parties, including entities engaged in the learning process. FL incorporates privacy-enhancing techniques like differential privacy, introducing statistical noise into data or model updates, rendering reidentifying individuals based on shared information exceedingly tricky. This feature is precious in sectors governed by strict data privacy regulations, such as healthcare, finance, and telecommunications [53].

Furthermore, transparency plays a pivotal role in establishing trust among collaborating parties. Participants can verify that sensitive data remains unexposed during model-building [54]. In terms of resilience, FL enhances system robustness through various means. For instance, it ensures efficient data utilization even in environments with limited network connectivity [55]. Most computations occur on edge devices (i.e., locally), requiring only intermittent network access to transmit aggregated model updates. Additionally, FL is designed to handle device failures and data corruption robustly. Should a device go offline or experience data corruption, the FL process continues with minimal disruption, as it relies on numerous other devices that persist in their local computations. This redundancy significantly enhances the reliability of FL models, ensuring their functionality even in adverse circumstances [52]. For instance, there are three nodes (Node 1, Node 2, and Node 3) in the decentralized infrastructure layer, as shown in Fig. 2. Each node has its own instance of the FL framework, represented by the FL Framework layer. The nodes communicate with each other through the decentralized infrastructure to collaborate on training an ML model using their local data while ensuring data privacy and security through FL techniques.

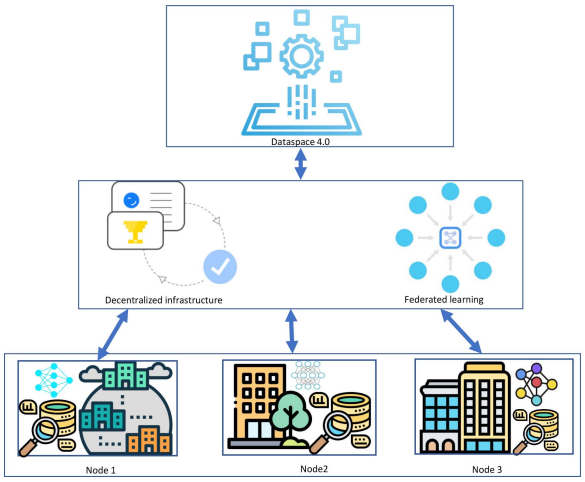

Fig. 2. Decentralized data sharing using FL.

In Fig. 2, the FL framework ensures that most processing happens in a distributed manner by distributing computational workloads across participating nodes. Every node in the network performs part of the total calculation. Using its local data set, each participating node trains its model independently. Therefore, it eliminates the need to send raw data to a central server by processing it locally on the node's device. Following local model training, the nodes only send the model updates—not the raw data—to a central server. Usually, these updates show the modifications or enhancements made to the models during local training. The central server combines these changes to create a global model. FL's decentralized architecture briefly contributes to a safe, private, and cooperative ML environment by distributing processing duties across nodes and ensuring that crucial operations like model training occur locally.

III. Combination of FL and Blockchain for Decentralized Data Sharing

The combination of FL and blockchain presents a robust solution for decentralized data sharing. FL enables secure, local model training across multiple parties without centralizing data, enhancing privacy and reducing network load. Blockchain complements this by providing a secure, transparent ledger for recording transactions and maintaining data integrity. Together, they create a powerful platform that enhances security, privacy, interoperability, and transparency in data sharing in healthcare [56]. Our approach uniquely addresses end-to-end data security, from local model training to secure data storage and sharing, promising substantial improvements in the efficiency and trustworthiness of collaborative data sharing. A comparative analysis of decentralized data-sharing when combining FL with blockchain technology reveals enhanced security, improved transparency, and efficient collaboration, as illustrated in Table III.

Enhanced Security: The combination of blockchain and FL ensures that data is encrypted, hashed, and distributed across a network of nodes, making it difficult for hackers to compromise the system. FL can enhance security by allowing

TABLE III
DECENTRALIZED DATA SHARING WHEN BLOCKCHAIN MEETS FL

Aspect	FL	Blockchain	Combination
Enhanced Security and Privacy	FL enables data to be trained locally, reducing the risk of data exposure during transmission. However, FL does not inherently address data security during transmission.	Blockchain provides tamper-proof and encrypted data storage, ensuring the security and privacy of shared data.	FL with blockchain ensures end-to-end security, from data training to storage and sharing.
Data Integrity and Transparency	FL focuses on model updates and consensus, ensuring that the shared model is accurate and reliable.	Blockchain's immutable and transparent ledger guarantees the integrity of shared data.	Combining FL's model updates with the blockchain's data record, both models and data can be verified for authenticity.
Interoperability and Standardization	FL promotes collaboration across diverse devices and platforms for model training.	Blockchain establishes standardized protocols and smart contracts for data access.	Combining both ensures interoperable data-sharing mechanisms and a common data usage framework.
Decentralized Governance	FL allows data owners to retain control over their data and contribute to model training.	Blockchain's decentralized consensus empowers participants to collectively agree on data-sharing terms.	Combining FL and blockchain extend this control to model updates and data access.
Resilience and Fault Tolerance	FL's distributed nature ensures system resilience against participant failures.	Blockchain's redundant data storage enhances resilience.	Combining both mitigates risks associated with individual participant failures.
Efficient Collaboration	FL facilitates collaborative model development.	Blockchain's transparent and automated smart contracts streamline data-sharing agreements.	FL and blockchain enhance efficient and trustworthy collaboration.
Data Monetization and Incentives	FL enables data owners to contribute to model training and earn incentives.	Blockchain's tokenization and incentive mechanisms extend these rewards to data-sharing.	The combination encourages active data contribution.

TABLE IV
ADVANTAGES OF USING FL AND BLOCKCHAIN FOR DECENTRALIZED DATA SHARING

Feature	FL	FL with Blockchain
Data privacy	Data is kept private by each party, but may be vulnerable to attacks during transmission	Data is kept private by each party and is secured by the tamper-proof nature of the blockchain
Security	Requires trust between parties, and may be vulnerable to attacks or malicious behavior	Provides a secure and transparent record of the training process, making it more resistant to attacks or malicious behavior
Scalability	Can scale to large datasets, but may be limited by the communication bandwidth and computational resources of each party	Can scale to large datasets, but may be limited by the computational resources required to perform blockchain transactions
Cost	Lower cost compared to centralized training, but may still require significant resources and coordination between parties	Higher cost due to the computational resources required for blockchain transactions, but may provide increased security and transparency that justifies the cost
Accuracy	Can produce high accuracy if each party has representative data but may be affected by data heterogeneity or class imbalance	Can produce high accuracy if each party has representative data, and the blockchain can provide a mechanism for identifying and addressing data heterogeneity or class imbalance

local model training on the user's device without transferring data to a central server.

Improved Privacy: By using blockchain to store data in an encrypted and distributed manner, users can retain control over their data and decide who can access it. FL can also improve privacy by allowing local model training on user devices without centralized data collection.

Improved Interoperability: Blockchain and FL can enable interoperability between different systems and platforms, allowing seamless data sharing across different networks. FL can also improve interoperability by aggregating locally trained models across different devices and platforms.

Greater Transparency: The use of blockchain can provide greater transparency in data sharing by providing an immutable record of all transactions. The combination can further enhance transparency by enabling users to verify the authenticity of data and model outputs. FL can also improve transparency by allowing for the inspection of locally trained models by independent auditors.

Improved Resilience: The combination of blockchain and FL can ensure that data and models are distributed across a decentralized network of nodes, making the system more resilient to failures and attacks. FL can also improve resilience by allowing local model training on user devices, reducing the reliance on centralized servers. Using blockchain with FL can increase security and privacy while ensuring a transparent and fair training process. However, it may also require additional computational resources and coordination between parties and may not always be necessary or practical, depending on the specific use case. Table IV compares FL with and without blockchain for decentralized data sharing.

Fig. 3 shows the combination of FL and blockchain for decentralized data sharing. Data sources represent data sources that can be used in Dataspace 4.0. These can include sensors, devices, databases, and other sources. At the same time, FL represents the ML algorithms used for training models on distributed data. FL allows models to be trained without the need for centralized data storage. Data labeling and model training represent the processes of labeling data and training ML models on the labeled data. This process can be done in a decentralized manner using FL. Blockchain consensus and validation of transactions represent the use of blockchain for consensus and validation of transactions in Dataspace 4.0. Blockchain provides a decentralized mechanism for validating and verifying data transactions.

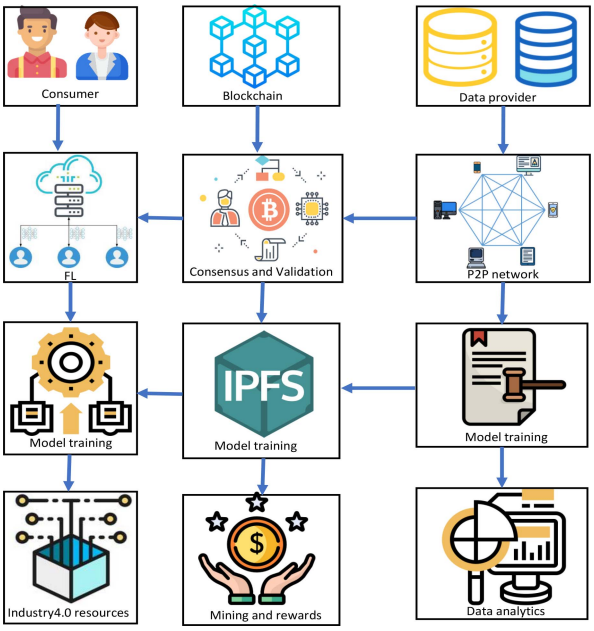

Fig. 3. Combination of FL and blockchain for decentralized data sharing.

Decentralized Data Management represents using the interplanetary file system (IPFS) for decentralized data management. IPFS allows data to be stored and accessed decentralized without relying on a central server. Mining Mechanism and Rewards represent the mechanism for mining data and rewarding data contributors. The rewards can be in the form of tokens or other incentives. Data analytics and reporting represent data analytics and tools to analyze and visualize data in Dataspace 4.0. These tools can be used to gain insights and make data-driven decisions. Data Governance represents using smart contracts for data governance in Dataspace 4.0. Smart contracts can be utilized to enforce rules and regulations for data sharing and access. Data consumers and smart data providers represent the users of Dataspace 4.0 who consume and provide data. Table V presents the security, privacy, interoperability, transparency, and resilience benefits of FL and blockchain technologies individually and in synergy within different industrial applications.

Nodes representing patients, researchers, and healthcare organizations must be put up to create a local, decentralized network for sharing medical data. By starting a blockchain, the nodes create a visible and safe ledger. Smart contracts are used to automate governance and guarantee compliance. Patients voluntarily supply personal health data, academics offer analytical models, and healthcare facilities contribute data sets. Nodes validate transactions using consensus procedures, keeping an accurate record. The network encourages cooperation by enabling a range of inputs without centralizing unprocessed data. By creating a safe and effective environment for healthcare data sharing, participants get access to a larger pool of data for research, improved privacy management, and transparent governance.

By distributing blockchain nodes across medical facilities, researchers, and patients, a distributed ledger is created to integrate blockchain technology into the local decentralized network. For automated governance, smart contracts enforce compliance with pre-established guidelines. By reaching a consensus on the ledger's current state, consensus mechanisms—like Proof of Authority or Proof of Stake—validate transactions and preserve data integrity. Blockchain improves security by guaranteeing data confidentiality and limiting unwanted access. Offering an unchangeable and auditable record of transactions encourages openness and builds participant confidence. Data immutability is a significant advantage as it offers a solid basis for healthcare data exchange inside the local network since it cannot be changed once data is stored on the blockchain.

An essential component of the infrastructure of the local network is the use of IPFS for decentralized data management. Instead of depending on a single server, IPFS functions as a distributed file system where data is saved among several nodes. It functions as a peer-to-peer network, enabling direct data storage and retrieval for any member of the healthcare ecosystem. Using a content-addressed architecture, IPFS ensures data integrity and minimizes redundancy by assigning a unique hash to each piece of data depending on its content. Because the data is spread across several nodes, IPFS has improved resilience, making the system resistant to failures. By enabling direct data retrieval from other network users, IPFS improves data accessibility and encourages a decentralized and effective method.

The community that the local decentralized network in healthcare serves benefits greatly. First, it allows hospitals, researchers, and patients to safely and effectively share medical data, improving patient care. The cooperative method improves the precision of medical diagnosis and available treatments. Second, the network expedites medical research by giving interested parties access to a large and varied data set while protecting personal privacy [57]. It encourages advancements in medical research and the creation of more realistic models. With its robust consensus processes, blockchain guarantees data security and privacy when integrated, while IPFS increases accessibility by decentralizing data storage and retrieval. In conclusion, cooperative data sharing on a local decentralized network advances healthcare, and IPFS and blockchain are essential for guaranteeing security, privacy, and accessibility for all parties involved [58].

IV. Decentralized Data Sharing in Healthcare: Use Case

Our methodology presents a decentralized approach in an era dominated by centralized data repositories. Let H represent the set of hospitals, where each hospital $h \in H$ maintains its independent data set. Integrating FL and blockchain in our framework presents a powerful combination. FL facilitates the initial stages of data preprocessing and distribution between entities like Hospital A and Hospital B. Meanwhile, blockchain serves as the decentralized ledger, ensuring subsequent data transactions' transparency, security, and immutability. By leveraging the strengths of both paradigms, we enhance the privacy, security, and efficiency of

TABLE V
BENEFITS OF DECENTRALIZED DATA SHARING IN DIFFERENT INDUSTRIES WITHIN THE CONTEXT OF INDUSTRY 4.0

Technology	Security	Privacy	Interoperability	Transparency	Resilience
FL	Encryption of data during transmission and storage	Data kept on local devices	Compatibility with different data formats	Limited transparency due to decentralized nature	Resilient to system failures
Blockchain	Immutable data storage	Decentralized control and verification	Ability to work across different systems	Publicly verifiable transactions	Resilient to tampering and attacks
Synergy of FL and blockchain	Multiple layers of encryption and verification	Data kept on local devices	Compatibility with different data formats	Publicly verifiable transactions	Resilient to tampering, attacks, and system failures

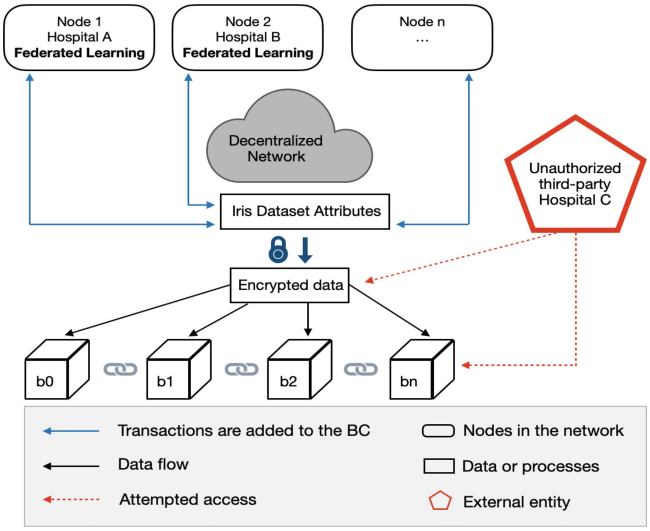

Fig. 4. Decentralized data sharing in healthcare use case.

Fig. 5. Nodes initialization.

decentralized data sharing. The schematic in Fig. 4 depicts the decentralized data-sharing process in a healthcare use case, highlighting the role of federated learning and the protection against unauthorized access.

The processing in our BCFL system is highly distributed across multiple nodes. Each node operates autonomously within the decentralized infrastructure, conducting computations using its local data. This design is foundational to the FL framework we have implemented. It allows for a resilient and reliable process, as each node independently contributes to the overarching ML model without centralizing data, thus preserving privacy and minimizing the risk of data corruption or loss. For instance, our framework involves multiple nodes collaborating through a decentralized network to train a ML model. The local computations at each node mean that even if one device goes offline or experiences data corruption, the FL process experiences minimal disruption. This not only enhances the reliability of the FL models but also ensures their functionality even in adverse circumstances. In essence, the bulk of the processing in our BCFL system occurs distributedly. Each node in the network takes on a portion of the computational load, with local data being processed at the edge, close to the data sources. The FL framework ensures that processing occurs locally at each node, particularly the computationally intensive model training tasks. This distributed processing approach is crucial for maintaining the system's integrity, ensuring data privacy, and enabling collaborative model training across various nodes. The local processing at the nodes is complemented by the blockchain, which provides a secure and transparent way to record and validate the model updates contributed by each node.

A. FL for Data Preprocessing and Distribution

In our approach, FL plays a pivotal role in the initial stages. Hospitals A and B utilize FL for data preprocessing while ensuring the raw data set remains securely within their respective premises. Through FL, both hospitals, despite retaining the actual data locally, collaboratively develop a model using shared insights and updates. The goal here is to benefit from the data available across both entities, and by the time any information gets ready for the blockchain, it is not the raw data but its processed encrypted attributes. The overall flow can be described as follows.

1) Hospital A and Hospital B each start with their local data sets. Fig. 5 provides an example of the node initialization process in a blockchain network, detailing the assigned hashes and the encryption keys for each node.
2) An FL cycle is initiated, where both hospitals collaborate to preprocess the data. Fig. 6 shows a sequence diagram for transactions between hospitals in a blockchain network, emphasizing the encryption, signature generation, and verification processes.
3) The processed data, now in a standardized format, is integrated into the blockchain for subsequent decentralized transactions.

It is worth noting that by utilizing FL at this stage, the integrity and privacy of the hospital data is maintained. Only aggregated updates are exchanged, ensuring data privacy.

B. Sharing Iris Data Set

The Iris data set, a widely used data set in ML and data analysis was employed as the primary data set for this

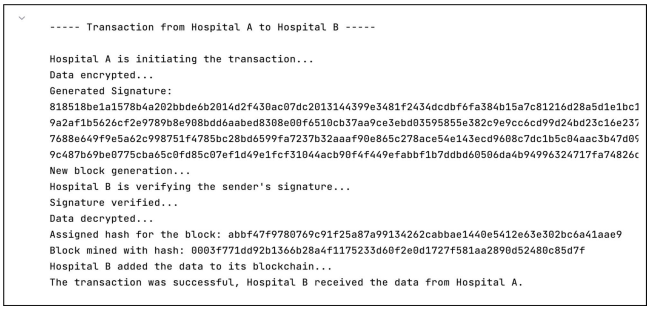

Fig. 6. Transactions between hospitals.

research. This data set consists of 150 samples from three species of Iris flowers (Iris setosa, Iris virginica, and Iris versicolor). Four features were measured from each sample: the lengths and the widths of the sepals and petals. Given its rich history in data analysis and ML, the Iris data set served as an ideal foundation for demonstrating the feasibility and effectiveness of our decentralized data-sharing mechanism.

1) Data Representation in Federated Learning: FL ensures that the participating nodes, like hospitals, retain their local data without exposing the raw data set to others. However, essential attributes or insights derived from the data might undergo encryption and be shared for collaborative learning. These shared attributes, rather than the actual data, get recorded on the blockchain, ensuring transparency, security, and consistency.

2) Data Representation in Blockchain: While the actual data sets, like the Iris data set, do not leave the respective hospitals, specific data attributes are processed and then encrypted for sharing on the blockchain. Specifically, the attributes of the Iris data set—sepal length, sepal width, and petal length—are encrypted using the recipient's public key. Additionally, the species label acts as metadata, which is not encrypted, allowing for querying based on species without requiring decryption. The complexity in these sections is centered around data attribute encryption and decryption. The encryption process used for the Iris data set attributes, like sepal length and width, is based on public-key cryptography. The time complexity for such operations typically depends on the critical size and the algorithm used, often being polynomial concerning the key length

$$T = \{\text{encrypt}(s_l, pk), \text{encrypt}(s_w, pk) \\ \text{encrypt}(p_l, pk), \text{encrypt}(p_w, pk), \text{species}\} \quad (1)$$

where:
1) s_l is the sepal length;
2) s_w is the sepal width;
3) p_l is the petal length;
4) p_w is the petal width.

3) Data Retrieval and Analysis: To retrieve specific data attributes from the blockchain, we implement Algorithm 1. The retrieval algorithm's complexity depends on the filtered data's size and the decryption process's efficiency. If n represents the number of transactions and d represents the decryption time, the total time complexity would be $O(n * d)$, assuming the filter operation's complexity is less than or equal to $O(n)$. After retrieving the data, standard data analysis or ML techniques can be applied to the decrypted data set.

Algorithm 1 Data Attributes Retrieval

```
function RETRIEVEDATA(species, sk)
    filtered_data ← filter_by_species(species)
    decrypted_data ← []
    for T in filtered_data do
        append(decrypted_data, decrypt(T, sk))
    end forreturn decrypted_data
end function
```

4) Data Structure (Blockchain): Each hospital's blockchain can be represented as a sequence of blocks

$$B = \{b_0, b_1, b_2, \ldots, b_n\} \quad (2)$$

where b_0 is the genesis block and b_n is the latest block. Each block b_i contains

$$b_i = \{T, h(b_{i-1}), \text{nonce}\} \quad (3)$$

where:
1) T is the transaction data;
2) $h(b_{i-1})$ is a cryptographic hash of the previous block;
3) nonce is a variable adjusted during the proof-of-work process.

The blockchain structure comprises a sequence of blocks, each linking to its predecessor through a hash. The complexity of adding a new block involves calculating the hash and performing the proof of work, which has a complexity of $O(2^k)$ on average, where k is the number of bits required by the difficulty target D.

C. Data Transaction

Given a message M, the encrypted message E for a recipient with public key pk is

$$E = \text{encrypt}(M, pk). \quad (4)$$

The signature S using the sender's private key sk is

$$S = \text{sign}(M, sk). \quad (5)$$

The data transaction process involves encryption and signing operations. Both operations are considered polynomial time complexity based on the key sizes used for encryption and signing. The transmission complexity depends on network factors and is typically considered $O(1)$ in the context of algorithmic analysis. Algorithm 2 outlines the process for sending encrypted data and the corresponding digital signature in a blockchain-based data transaction.

D. Consensus Mechanism: Proof of Work

The proof-of-work consensus mechanism aims to find a nonce such that

$$h(T, h(b_{i-1}), \text{nonce}) < D \quad (6)$$

where:

Algorithm 2 Data Transaction

procedure SENDDATA(M, $pk_{recipient}$, sk_{sender})
    $E \leftarrow$ encrypt(M, $pk_{recipient}$)
    $S \leftarrow$ sign(M, sk_{sender})
    transmit(E, S) ▷ Send encrypted data and signature
end procedure

Algorithm 3 Proof of Work

procedure MINEBLOCK(T, $h(b_{i-1})$, D)
    nonce $\leftarrow 0$
    while $h(T, h(b_{i-1}), \text{nonce}) \geq D$ **do**
        nonce $\leftarrow$ nonce $+ 1$
    end whilereturn nonce
end procedure

Fig. 7. Requests for data access between hospitals.

1) D represents the target difficulty;
2) h is the hashing function.

Proof of work is inherently designed to be computationally intensive. The complexity is not fixed and is adjusted by the difficulty target D. The average time complexity of finding a valid nonce is proportional to the difficulty target, which is typically exponential concerning the number of leading zeros required in the hash output. Algorithm 3 describes the Proof of Work (PoW) process, essential for maintaining the integrity and trust in blockchain operations.

E. Authorization Mechanism

Let the centralized registry R be a set of tuples

$$R = \{(id_1, pk_1), (id_2, pk_2), \ldots, (id_n, pk_n)\} \tag{7}$$

where:

1) id_i is the unique identifier of hospital h_i;
2) pk_i is the public key of hospital h_i.

The authorization check function, is Authorized(pk), verifies if a given public key exists in the registry R. In Fig. 7, the transactional workflow for requesting and granting data access between hospitals is depicted, demonstrating the use of encryption and blockchain verification.

The authorization check involves searching through a registry for a matching public key. If the registry is unsorted and has n entries, this operation has a worst case time complexity of $O(n)$. If the registry is sorted or hashed, the time complexity could be reduced to $O(\log\ n)$ or even $O(1)$, respectively.

Algorithm 4 Authorization Check

function ISAUTHORIZED(pk)
    if $\exists(id, pk) \in R$ **then return** True
    elsereturn False
    end if
end function

Algorithm 5 Potential Attack Methods

function REPLAYTRANSACTION(interceptedTransaction)
    send(interceptedTransaction)
end function

function MASQUERADE(fakeID, transactionData)
    fakeSignature $\leftarrow$ forgeSignature(transactionData)
    send(transactionData, fakeSignature, fakeID)
end function

function INTERCEPTANDALTER(transaction)
    interceptedData $\leftarrow$ transaction.data
    alteredData $\leftarrow$ modify(interceptedData)
    forward(alteredData)
end function

Algorithm 4 presents a method for checking authorization of a participant in a blockchain network using a public key.

F. Adversarial Simulation: Hospital C

For our research, we introduced a malicious third-party entity termed Hospital C. This entity was not part of the authorized hospital's list and acted as an adversary, simulating various attack vectors to compromise the system's security. Algorithm 5 enumerates potential attack methods within a blockchain network, including replay, masquerade, and intercept and alter attacks.

1) *Replay Attack:* Hospital C eavesdrops on the transactions between Hospital A and Hospital B. It tries to resend intercepted transactions, aiming to reinsert data or initiate unauthorized data requests.
2) *Identity Masquerade:* Hospital C attempts to masquerade as Hospital A or Hospital B by forging signatures or manipulating its IP address.
3) *Man-in-the-Middle Attack:* Hospital C places itself between Hospital A and Hospital B, intercepting and potentially altering the data being exchanged.

G. Defense Mechanisms

Against the backdrop of these simulated attacks, our blockchain implementation showcased several defense mechanisms.

1) *Nonce and Hash Verification:* Every block contains a nonce value, ensuring the block's hash matches a particular pattern. Replay attacks get detected as the blockchain verifies the nonce and hash values, and a reused nonce value indicates a replay attempt.
2) *Digital Signatures and IP Verification:* Our system uses RSA-based digital signatures to verify the authenticity of

transactions. The digital signature verification will fail if Hospital C masquerades Hospital A or B. Additionally, IP address checks were implemented to add an extra layer of verification, further thwarting identity masquerade attempts.

3) *End-to-End Encryption:* Data exchanged between hospitals is encrypted using the recipient's public key. This ensures that even if Hospital C intercepts the data in a man-in-the-middle attack, it cannot decrypt or modify it without the corresponding private key.

Through these defense mechanisms, our decentralized data-sharing blockchain system demonstrated resilience against the common threats posed by adversarial entities.

H. Evaluation

Evaluation of the effectiveness of the defense mechanisms against poisoning attacks conducted by the adversarial entity is as follows.

1) *Replay Attack:* The defence mechanism includes nonce and hash verification within the blockchain. When Hospital C attempts to resend intercepted transactions, the system checks for nonce values. A reused nonce indicates a replay attempt, which the blockchain is designed to detect. The graph in Fig. 8 shows a low success rate for replay attacks, remaining consistently low across multiple attempts. This indicates the system's effective detection and prevention of replay attempts attributed to the robust verification process.
2) *Identity Masquerade:* The system uses RSA-based digital signatures and IP verification to ensure the authenticity of transactions. Hospital C's attempts to forge signatures or manipulate its IP address will likely be unsuccessful due to these stringent checks. The graph in Fig. 8 corroborates this: the success rate for identity masquerade attacks is also low and does not significantly increase with more attempts. This reflects the strength of the digital signature verification and IP checks in preventing unauthorized entity masquerading.
3) *Man-in-the-Middle Attack:* With end-to-end encryption, it cannot decrypt or alter the information even if Hospital C intercepts the data without the corresponding private key. The graph in Fig. 8 suggests that man-in-the-middle attacks have a slightly higher success rate than the other two types but remain relatively low. This slight increase could be due to the complexity of detecting and preventing active interception compared to the more straightforward detection of replay and identity attacks. Nonetheless, the encryption mechanism is a solid barrier, preventing Hospital C from gaining meaningful access to the data.

The overall low success rates across all attack types illustrate the robustness of the defence mechanisms. The nonce and hash checks, digital signature and IP verification, and end-to-end encryption collectively contribute to the resilience of the blockchain system, effectively mitigating the risk of poisoning attacks. This analysis, supported by the empirical data shown in Fig. 8, demonstrates that the defence strategies are sufficiently robust, and the system can be considered secure against the simulated adversarial actions.

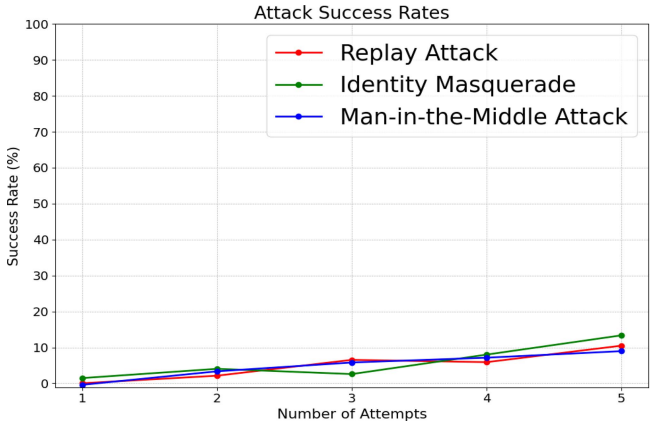

Fig. 8. Attacks' success rates.

V. Challenges, Opportunities, and Future Directions

A. Challenges

Decentralized data sharing presents several technical challenges that must be addressed to ensure its effectiveness and security. Some of these challenges include the following.

Interoperability: Different decentralized data-sharing systems may use different protocols and standards, making sharing data across different systems difficult. This requires standardization and interoperability between systems.

Scalability: Decentralized data-sharing systems must be designed to handle large amounts of data and many participants. This requires efficient data storage and retrieval mechanisms and distributed processing capabilities.

Consensus: Decentralized data-sharing systems rely on consensus mechanisms to ensure that all participants agree on the validity of shared data. This requires robust consensus algorithms to handle malicious attacks and ensure data integrity.

Security: Decentralized data-sharing systems must be designed to protect data from unauthorized access, tampering, and corruption. This requires robust authentication, encryption, and effective mechanisms for detecting and mitigating attacks.

Privacy: Decentralized data-sharing systems must protect the privacy of participants' data and sensitive personal and financial data. This requires effective mechanisms for anonymizing and protecting data and ensuring participants have control over the data.

Data Quality: Decentralized data-sharing systems must ensure the accuracy and reliability of shared data, especially in cases where data is collected from multiple sources. This requires effective data validation and verification mechanisms to resolve conflicts between data sources.

B. Opportunities

First, it enables businesses and organizations to access a broader range of data, leading to more comprehensive insights and improved decision making. This can lead to the development of new products and services and enhance

the competitiveness of companies. Second, decentralized data sharing promotes collaboration among participants, allowing them to work together to solve complex problems and develop new solutions. This can lead to new business models, partnerships, and ecosystems. Third, decentralized data sharing can facilitate the development of new technologies and applications, such as blockchain and edge computing, which can further enhance the capabilities of Dataspace 4.0. Fourth, it can lead to increased transparency and accountability, which is particularly important in healthcare and finance, where privacy and security are crucial. Finally, decentralized data sharing can give individuals more control over their data, increasing privacy and security. This can lead to the development of new services that provide individuals with more control over their personal information.

The combination of BCFL for decentralized data sharing presents a unique and promising use case in healthcare, particularly for remote monitoring applications.

1) *Remote Patient Monitoring (RPM):* It involves tracking patient health data outside of traditional clinical settings. This could include monitoring vital signs, blood sugar levels, heart rate, or other relevant health metrics through wearable devices or home-based equipment [59].
2) *Collaborative Research and Treatment Optimization:* BCFL can facilitate collaborative research among different healthcare entities while maintaining data privacy. This collaboration can lead to more comprehensive health models, benefiting treatment optimization [60].
3) *Regulatory Compliance and Consent Management:* Healthcare is a highly regulated sector, and BCFL can aid in complying with regulations like HIPAA, GDPR, and others concerning patient data protection [61].

C. Future Directions

Advancing decentralized data sharing requires multifaceted research efforts. Technical challenges, including data integration, interoperability, and security, demand the development of tailored algorithms and architectures. Legal and regulatory dimensions necessitate the exploration of frameworks safeguarding privacy amid data sharing. Investigating the potential of decentralized data sharing in industries like healthcare and finance involves identifying domain-specific use cases. Additionally, emerging technologies, such as blockchain and edge computing, require scrutiny for their performance in decentralized contexts. Finally, developing business models and ecosystems with incentives for collaboration is vital. Looking ahead, a focus on practical applications, exemplified through case studies in healthcare partnerships, aims to validate methodologies, address concerns about centralized control, and enhance flexibility for global applicability. The commitment to refining and verifying these approaches in real-world healthcare underscores a dedicated thrust for the evolution of decentralized data sharing.

VI. Conclusion

This article has introduced a groundbreaking exploration of the conceptual framework and technical synergy between FL and blockchain, signalling a paradigm shift toward secure, collaborative, and patient-centric decentralized data sharing in the data-driven healthcare era. The combination of FL's decentralized ML paradigm and blockchain's transparent and immutable ledger creates an ecosystem fostering trust, security, and data integrity. While a specific real-world healthcare use case is not presented, this article vividly outlines the potential impact of this fusion on patient care, emphasizing the preservation of patient privacy alongside granting healthcare providers and researchers access to diverse data sets. The proposed approach promises to accelerate medical research, improve treatment outcomes, and empower patients through data ownership. The synergy of FL and blockchain envisions a healthcare ecosystem that prioritizes individual privacy, fosters advancements in medical science, and sets the stage for a transformative shift in healthcare data sharing. This innovative approach addresses the challenges of balancing data utility and privacy and opens avenues for more accurate models, leading to enhanced diagnoses and ultimately contributing to the evolution of a patient-centric and collaborative healthcare landscape.

References

[1] "A common data space 4.0 for European manufacturing." DFA. Accessed: Oct. 16, 2023. [Online]. Available: https://digitalfactoryalliance.eu/moving-towards-a-common-data-space-4-0-for-european-manufacturing/

[2] "Dataspace 4.0." Accessed: Jun. 23, 2023. [Online]. Available: https://digitalfactoryalliance.eu/data-space-4-0-alliance/

[3] P. Varga et al., "Converging telco-grade solutions 5G and beyond to support production in industry 4.0," *Appl. Sci.*, vol. 12, no. 15, p. 7600, 2022.

[4] B. Han et al., "Digital twins for industry 4.0 in the 6G era," 2022, *arXiv:2210.08970*.

[5] T. White, E. Blok, and V. D. Calhoun "Data sharing and privacy issues in neuroimaging research: Opportunities, obstacles, challenges, and monsters under the bed," *Hum. Brain Map.*, vol. 43, no. 1, pp. 278–291, 2022.

[6] A. Torab-Miandoab, T. Samad-Soltani, A. Jodati, and P. Rezaei-Hachesu, "Interoperability of heterogeneous health information systems: A systematic literature review," *BMC Med. Informat. Decis. Mak.*, vol. 23, no. 1, p. 18, 2023.

[7] M. A. Uddin, A. Stranieri, and I. Gondal, "A survey on the adoption of blockchain in IoT: Challenges and solutions," *Blockchain, Res. Appl.*, vol. 2, no. 2, 2021, Art. no. 100006.

[8] V. Neumann et al., "Examining public views on decentralised health data sharing," *Plos One*, vol. 18, no. 3, 2023, Art. no. e0282257.

[9] E. Curry et al., "Data sharing spaces: The BDVA perspective," in *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage*, pp. 365–382. Cham, Switzerland: Springer Int. Publ., 2022.

[10] V. Pandi Chellapandi, A. Upadhyay, A. Hashemi, and S. H. Zak, "On the convergence of decentralized federated learning under imperfect information sharing," 2023, *arXiv:2303.10695*.

[11] H. Niavis, N. Papadis, V. Reddy, H. Rao, and L. Tassiulas, "A blockchain-based decentralized data sharing infrastructure for off-grid networking," in *Proc. IEEE Int. Conf. Blockchain Cryptocurr. (ICBC)*, 2020, pp. 1–5.

[12] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Artif. Intell. Statist.*, 2017, pp. 1273–1282.

[13] H. Fang and Q. Qian, "Privacy preserving machine learning with homomorphic encryption and federated learning," *Future Internet*, vol. 13, no. 4, p. 94, 2021.

[14] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," in *Proc. Decent. Bus. Rev.*, 2008, p. 21260

[15] H. S. Chen, J. T. Jarrell, K. A. Carpenter, D. S. Cohen, and X. Huang "Blockchain in healthcare: A patient-centered model," *Biomed. J. Sci. Techn. Res.*, vol. 20, no. 3, 2019, Art. no. 15017.

[16] S. Alansari, "A blockchain-based approach for secure, transparent and accountable personal data sharing," Ph.D. dissertation, Faculty of Eng., Sci. Math., University of Southampton, Southampton, U.K., 2020.

[17] S. H. Alsamhi et al., "Drones' edge intelligence over smart environments in B5G: Blockchain and federated learning synergy," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 1, pp. 295–312, Mar. 2022.

[18] I. Jao et al., "Research stakeholders' views on benefits and challenges for public health research data sharing in Kenya: the importance of trust and social relations." *PLoS ONE*, vol. 10, no. 9, 2015, Art. no. e0135545.

[19] Y. M. Arif, H. Nurhayati, F. Kurniawan, S. M. S. Nugroho, and M. Hariadi "Blockchain-based data sharing for decentralized tourism destinations recommendation system," *Int. J. Intell. Eng. Syst.*, vol. 13, no. 6, pp. 472–486, 2020.

[20] T. Guggenberger, A. Schweizer, and N. Urbach, "Improving interorganizational information sharing for vendor managed inventory: Toward a decentralized information hub using blockchain technology," *IEEE Trans. Eng. Manag.*, vol. 67, no. 4, pp. 1074–1085, Nov. 2020.

[21] M. Johnson, M. Jones, M. Shervey, J. T. Dudley, and N. Zimmerman, "Building a secure biomedical data sharing decentralized app (DApp): Tutorial," *J. Med. Internet Res.*, vol. 21, no. 10, 2019, Art. no. e13601.

[22] A. Balador, A. Bazzi, U. Hernandez-Jayo, I. de la Iglesia, and H. Ahmadvand, "A survey on vehicular communication for cooperative truck platooning application," *Veh. Commun.*, vol. 35, 2022, Apr. no. 100460.

[23] M. Firdaus, S. Rahmadika, and K. H. Rhee, "Decentralized trusted data sharing management on internet of vehicle edge computing (IoVEC) networks using consortium blockchain," *Sensors*, vol. 21, no. 7, p. 2410, 2021.

[24] P. Wang, W. Cui, and J. Li, "A framework of data sharing system with decentralized network," in *Proc. 1st Int. Conf. (BigSDM)*, 2019, pp. 255–262.

[25] O. Gallay, K. Korpela, N. Tapio, and J. K. Nurminen "A peer-to-peer platform for decentralized logistics," in *Proc. Hamburg Int. Conf. Logist. (HICL)*, 2017, pp. 19–34.

[26] S. Swetha and P. M. JoePrathap, "A study on a decentralized network secured data sharing using blockchain," in *Proc. 1st Int. Conf. Comput. Sci. Technol. (ICCST)*, 2022, pp. 620–624.

[27] C. F. L. Hickman et al., "Data sharing: Using blockchain and decentralized data technologies to unlock the potential of artificial intelligence: What can assisted reproduction learn from other areas of medicine?" *Fertil. Steril.*, vol. 114, no. 5, pp. 927–933, 2020.

[28] F. Firouzi, B. Farahani, and A. Marinšek, "The convergence and interplay of edge, fog, and cloud in the AI-driven Internet of Things (IoT)," *Inf. Syst.*, vol. 107, Jul. 2022, Art. no. 101840.

[29] F. Firouzi et al., "Fusion of IoT, AI, edge–fog–cloud, and blockchain: Challenges, solutions, and a case study in healthcare and medicine," *IEEE Internet Things J.*, vol. 10, no. 5, pp. 3686–3705, Mar. 2023.

[30] B. Farahani, F. Firouzi, and M. Luecking, "The convergence of IoT and distributed ledger technologies (DLT): Opportunities, challenges, and solutions," *J. Netw. Comput. Appl.*, vol. 177, 2021, Art. no. 102936.

[31] Y. Liu et al., "Deep anomaly detection for time-series data in Industrial IoT: A communication-efficient on-device federated learning approach," *IEEE Internet Things J.*, vol. 8, no. 8, pp. 6348–6358, Apr. 2021.

[32] Q. Meng, F. Zhou, H. Ren, T. Feng, G. Liu, and Y. Lin, "Improving federated learning face recognition via privacy-agnostic clusters," 2022, *arXiv:2201.12467*.

[33] T. Li, A. Kumar Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, May 2020.

[34] S. H. Alsamhi, A. V. Shvetsov, A. Hawbani, S. V. Shvetsova, S. Kumar, and L. Zhao, "Survey on federated learning enabling indoor navigation for industry 4.0 in B5G," *Future Gener. Comput. Syst.*, vol. 148, pp. 250–265, Nov. 2023.

[35] Y. Tian, S. Wang, J. Xiong, R. Bi, Z. Zhou, and M. Z. A. Bhuiyan, "Robust and privacy-preserving decentralized deep federated learning training: Focusing on digital healthcare applications," *IEEE/ACM Trans. Comput. Biol. Bioinformat.*, early access, Mar. 3, 2023, doi: 10.1109/TCBB.2023.3243932.

[36] I. Makhdoom, I. Zhou, M. Abolhasan, J. Lipman, and W. Ni, "PrivySharing: A blockchain-based framework for privacy-preserving and secure data sharing in smart cities. *Comput. Secur.*, vol. 88, Jan. 2020, Art. no. 101653.

[37] N. Fatima, P. Agarwal, and S. S. Sohail "Security and privacy issues of blockchain technology in health care—A review," in *ICT Analysis and Applications*. Singapore: Springer, pp. 193–201, 2022.

[38] H. Fan et al., "Privacy preserving ultra-short-term wind power prediction based on secure multi party computation," 2023, *arXiv:2301.13513*.

[39] Y. Ye, L. Zhang, W. You, and Y. Mu, "Secure decentralized access control policy for data sharing in smart grid," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, 2021, pp. 1–6.

[40] M. Sultana, A. Hossain, F. Laila, K. A. Taher, and M. N. Islam "Towards developing a secure medical image sharing system based on zero trust principles and blockchain technology," *BMC Med. Informat. Decis. Mak.*, vol. 20, no. 1, pp. 1–10, 2020.

[41] Y. Lu, X. Huang, Y. Dai, S. Maharjan and Y. Zhang, "Blockchain and federated learning for privacy-preserved data sharing in Industrial IoT," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 4177–4186, Jun. 2020.

[42] M. Stolpe, "The Internet of Things: Opportunities and challenges for distributed data analysis," *ACM SIGKDD Explor. Newslett.*, vol. 18, no. 1, pp. 15–34, 2016.

[43] R. E. Endeley, "End-to-end encryption in messaging services and national security—Case of Whatsapp messenger," *J. Inf. Secur.*, vol. 9, no. 1, p. 95, 2018.

[44] B. Lashkari and P. Musilek, "A comprehensive review of blockchain consensus mechanisms," *IEEE Access*, vol. 9, pp. 43620–43652, 2021.

[45] M. Pilkington, "11 blockchain technology: Principles and applications," in *Research handbook On Digital Transformations*, vol. 225. Cheltenham, U.K.: Edward Elgar Publ., 2016.

[46] E.-H. Diallo, O. Dib, and K. Al Agha. "A scalable blockchain-based scheme for traffic-related data sharing in VANETs," *Blockchain, Res. Appl.*, vol. 3, no. 3, 2022, Art. no. 100087.

[47] M. Pilkington, "Blockchain technology: Principles and applications," in *Research Handbook on Digital Transformations*. Cheltenham, U.K.: Edward Elgar Publ., 2016, pp. 225–253.

[48] H. Xu, P. V. Klaine, O. Onireti, B. Cao, M. Imran, and L. Zhang, "Blockchain-enabled resource management and sharing for 6G communications," *Digit. Commun. Netw.*, vol. 6, no. 3, pp. 261–269, 2020.

[49] N. Radziwill, "Blockchain revolution: How the technology behind bitcoin is changing money, business, and the world," *Qual. Manag. J.*, vol. 25, no. 1, pp. 64–65, 2018.

[50] J. Konecný, H. B. McMahan, F. X. Yu, P. Richtàrik, A. Theertha Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," 2016, *arXiv:1610.05492*.

[51] K. Bonawitz et al., "Towards federated learning at scale: System design," in *Proc. Mach. Learn. Syst.*, vol. 1, 2019, pp. 374–388.

[52] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proc. Mach. Learn. Syst.*, vol. 2, 2020, pp. 429–450.

[53] N. Rieke et al., "The future of digital health with federated learning." *NPJ Digit. Med.*, vol. 3, no. 1, p. 119, 2020.

[54] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–19, 2019.

[55] T. Nishio and R. Yonetani. "Client selection for federated learning with heterogeneous resources in mobile edge," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2019, pp. 1–7.

[56] R. Myrzashova, S. H. Alsamhi, A. V. Shvetsov, A. Hawbani, and X. Wei, "Blockchain meets federated learning in healthcare: A systematic review with challenges and opportunities," *IEEE Internet Things J.*, vol. 10, no. 16, pp. 14418–14437, Aug. 2023.

[57] Z. Zhou, C. Guo, X. Zhang, R. Wang, L. Zhang, and M. Imran, "A Blockchain-based data sharing marketplace with a federated learning use case," in *Proc. IEEE Int. Conf. Blockchain Cryptocurr. (ICBC)*, Dubai, UAE, 2023, pp. 1041–1044.

[58] J. Bang and M.-J. Choi, "Design of personal data protection decentralized model using blockchain and IPFS," in *Proc. 24st Asia–Pac. Netw. Oper. Manag. Symp. (APNOMS)*, Sejong, South Korea, 2023, pp. 251–254.

[59] D. K. Acharya, M. Shrivastava, and P. Padhi, "A decentralized blockchain-based IoT system for privacy-preserving data sharing," in *Proc. IEEE Int. Conf. Blockchain Distrib. Syst. Secur. (ICBDS)*, New Raipur, India, 2023, pp. 1–5.

[60] R. Song, B. Xiao, Y. Song, S. Guo, and Y. Yang, "A survey of blockchain-based schemes for data sharing and exchange," *IEEE Trans. Big Data*, vol. 9, no. 6, pp. 1477–1495, Dec. 2023.

[61] Y. Liu, P. Liu, W. Jing, and H. H. Song, "PD2S: A privacy-preserving differentiated data sharing scheme based on blockchain and federated learning," *IEEE Internet Things J.*, vol. 10, no. 24, pp. 21489–21501, Dec. 2023.