Computer Communications 222 (2024) 38–67

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Review

Privacy-preserving in Blockchain-based Federated Learning systems

Sameera K.M. [a], Serena Nicolazzo [b,*], Marco Arazzi [c], Antonino Nocera [c], Rafidha Rehiman K.A. [a], Vinod P. [a,d], Mauro Conti [d]

[a] Department of Computer Applications, Cochin University of Science and Technology, India
[b] Department of Computer Science, University of Milan, Italy
[c] Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Italy
[d] Department of Mathematics, University of Padua, Italy

ARTICLE INFO

Keywords:
Federated Learning
Blockchain
Privacy
Blockchain-enabled FL
IoT
Industry 5.0

Objective
Problem Identification
Problem-Solving Approach
Technology Integration for Efficiency
Research Gap and Potential Improvements

ABSTRACT

Federated Learning (FL) has recently arisen as a revolutionary approach to collaborative training Machine Learning models. According to this novel framework, multiple participants train a global model collaboratively, coordinating with a central aggregator without sharing their local data. As FL gains popularity in diverse domains, security, and privacy concerns arise due to the distributed nature of this solution. Therefore, integrating this strategy with Blockchain technology has been consolidated as a preferred choice to ensure the privacy and security of participants.

This paper explores the research efforts carried out by the scientific community to define privacy solutions in scenarios adopting Blockchain-Enabled FL. It comprehensively summarizes the background related to FL and Blockchain, evaluates existing architectures for their integration, and the primary attacks and possible countermeasures to guarantee privacy in this setting. Finally, it reviews the main application scenarios where Blockchain-Enabled FL approaches have been proficiently applied. This survey can help academia and industry practitioners understand which theories and techniques exist to improve the performance of FL through Blockchain to preserve privacy and which are the main challenges and future directions in this novel and still under-explored context. We believe this work provides a novel contribution concerning the previous surveys and is a valuable tool to explore the current landscape, understand perspectives, and pave the way for advancements or improvements in this amalgamation of Blockchain and Federated Learning.

Scope of work

Why merge?
limitation of FL + BlockChain=S

1. Introduction

Federated Learning (FL, hereafter) has undergone a significant surge in popularity in recent years. This novel strategy enables training Machine Learning models directly on user devices or at the edge without centralizing raw data. Because sensitive data stays on the user's device, the risk of exposing information to possible data breaches is lowered, and user privacy is preserved. Moreover, collaboration among workers provides access to a large amount of data, which enhances performance, making models more efficient and scalable. While FL offers several advantages, it also has limitations and challenges. The main FL characteristics that expose it to new threats are *(i)* system heterogeneity, *(ii)* the need for a trustworthy central authority for the coordination of the processing of locally trained models, *(iii)* vulnerability to data falsification and inference attack, *(iv)* the lack of incentive mechanism for the participating nodes, *(v)* communication security, and *(vi)* regulatory complaints [1,2].

Because current implementations of the FL system do not provide proper mechanisms to address these challenges, researchers have recently begun to study approaches that leverage Blockchain [3]. This new technology, derived from the decentralized cryptocurrency system, has attracted the interest of industry and academia for its countless potential. It relies on the possibility of performing authentic and traceable transactions without a trusted third party and ensuring secure data storage and tracking. Hence, integrating Blockchain and FL can empower this last paradigm, ensuring data privacy, trust, and model security in decentralized collaborative Learning environments.

This study comprehensively reviews Blockchain-enabled FL, focusing mainly on data privacy. Although several papers analyze different aspects of the Blockchain-enabled FL paradigm, a systematic review of existing works on privacy still needs to be included. Moreover, we provide a novel perspective examining the possible attacks menacing

* Corresponding author.
E-mail addresses: sameerakm@cusat.ac.in (Sameera K.M.), serena.nicolazzo@unimi.it (S. Nicolazzo), marco.arazzi01@universitadipavia.it (M. Arazzi), antonino.nocera@unipv.it (A. Nocera), rafidharehimanka@cusat.ac.in (Rafidha Rehiman K.A.), vinod.p@cusat.ac.in, vinod.puthuvath@unipd.it (Vinod P.), mauro.conti@unipd.it (M. Conti).

https://doi.org/10.1016/j.comcom.2024.04.024
Received 9 January 2024; Received in revised form 15 March 2024; Accepted 18 April 2024
Available online 20 April 2024

privacy in this scenario and all the current adopted countermeasures present in literature to guarantee privacy through a Blockchain-enabled FL system. Specifically, our main aim is to examine the existing literature on privacy attacks in Blockchain-enabled Federated Learning (BCFL) systems. Moreover, we organize related papers according to the type of solution they implemented for privacy preservation, such as differential privacy, homomorphic encryption, secure multiparty computation, reward-driven approaches, hybrid privacy approaches, and cross-chained Federated Learning. Lastly, we delve into the practical applications of BCFL in cutting-edge scenarios, including healthcare, Industry 5.0, and the Internet of Vehicles. This survey provides several contributions, namely:

- It introduces a conceptual introduction to both FL and Blockchain technologies. Moreover, it deep dives into the description of existing architectures for Blockchain-enabled FL, describing how Blockchain can tackle the current challenges for FL, especially those related to privacy.
- It identifies the primary attacks menacing data privacy in Blockchain-enabled FL systems and the recently investigated countermeasures involving privacy methods, such as homomorphic encryption, differential privacy, secure multiparty computation methods, reputation approaches, and solutions relying on cross-chain FL.
- It describes how several practical application scenarios in various industries can benefit from integrating Blockchain and FL.
- It discusses and examines Blockchain-enabled FL systems' future directions and open research problems.

This survey offers fresh perspectives on the new paradigm of Blockchain-enabled FL focused on privacy-preserving. We expect the conducted analysis to be helpful to practitioners and researchers in categorizing the high number of studies dealing with privacy-preserving Blockchain-enabled FL approaches and in highlighting potentially promising directions to motivate future research work.

The structure of this paper is outlined as follows. Section 2 introduces related survey studies, while Section 3 delves into the methodology employed for conducting this survey. In Section 4, we overview both FL and Blockchain main concepts. In Section 5, we analyze the State-of-the-Art regarding the integration of FL and Blockchain technology, describing the main architectures. Section 6 addresses the primary privacy threats within Blockchain-enabled FL. Sections 7 and 8 focus on the possible solutions to preserve privacy in such a domain. Section 9 is devoted to the analysis of several application scenarios that benefit from Blockchain-enabled FL (such as Healthcare, Industrial IoT (IIoT), and the Internet of Vehicles). In Section 11, we explore various unresolved challenges and provide insights for potential areas of future research. Ultimately, Section 12 encapsulates our concluding remarks on the survey.

2. Comparison with other survey articles

Although several related surveys have been conducted to explore the integration of Blockchain and FL from different perspectives, most of them focus on different aspects, issues, or application domains related to this combination [1,4–7].

For instance, Qu et al. [1] consider three problems of Blockchain-enabled FL, namely decentralization, incentive mechanism, and membership selection. They focus on attack categorization and evaluate the performance of existing countermeasures. The paper presented in [4] describes the structural designs of Blockchain-enabled FL, the deployed platforms, and possible industrial applications. Moreover, it analyses the aspects of Blockchain that allow an improvement of the FL system, such as the node incentive mechanisms.

The authors of [8] conduct a literature review on the integration of Blockchain in FL, analyzing 41 research studies published between the years 2016 to June 2022. They focus on several aspects of the BCFL system: security and privacy, record and reward, verification, and accountability.

In [5,7], the authors focus on Blockchain-based FL approaches for IoT applications. Specifically, [5] presents the notion of Blockchain, its application to IoT, and the privacy issues and possible countermeasures. Then, they introduce the FL application in IoT systems, devise a taxonomy, and present privacy threats in FL. The combination of these two paradigms is only briefly presented through an IoT-based use case. The work proposed by Nguyen et al. [6] instead focuses on the applications of BCFL in mobile-edge computing domains, analyzing some critical aspects of system design, including communication cost, resource allocation, incentive mechanism, as well as aspects related to security and the safeguarding of privacy.

Several works rely on custom taxonomies to categorize related literature on Blockchain-based FL [2,9,10]. In particular, in [9], the authors propose a taxonomy to categorize Blockchain-based FL systems referring to three distinct layers: the Blockchain, the training, and the aggregation layers. They briefly review and summarize representative work according to this taxonomy.

Similarly, the study presented in [10] analyses 41 research papers between 2018 and 2021 deal with Blockchain-based FL methodologies in smart environments, categorizing work in a custom taxonomy. In particular, FL methodologies are divided into public FL and private FL environments. In public Blockchain mechanisms, they investigate only vertical FL approaches, whereas, for private Blockchain mechanisms, they evaluate both horizontal FL and Federated Transfer Learning approaches.

Zhu et al. [2] rely on a categorization of BCFL models in three classes: decoupled, coupled, and overlapped, according to how the FL and Blockchain functions are integrated. Then, they use these classes to compare the advantages and disadvantages of the e state-of-the-art solution they considered.

The authors of [11] conduct a brief review of existing literature on Blockchain-based FL that addresses privacy challenges. They describe only 18 papers published mainly from 2019 to 2022 but do not consider all the possible approaches to guarantee privacy in BCFL.

Table 1 summarizes the main topics addressed by related surveys and makes a comparison with our contribution. As visible from the table, none of the existing contributions cover the topics presented in the survey or consider papers in the temporal span we analyzed. Table 2 outlines the distribution of articles, specifying the number of articles considered in each category related to existing surveys and our work, along with the corresponding publication years.

3. Methodology

3.1. Research approach

Blockchain-enabled Federated Learning has surfaced as a groundbreaking paradigm in the rapidly evolving technology landscape, presenting the prospect of decentralized and collaborative machine Learning while safeguarding data privacy. This study intricately explores the critical aspect of privacy within this innovative framework, meticulously examining potential threats and presenting effective mitigating strategies. It systematically explores Blockchain fundamentals and Federated Learning along with its categorization. It delves into relevant literature on privacy attacks and protection methods in BCFL. Moreover, it highlights the essential need for privacy preservation in BCFL-focused applications across domains. The study aims to identify areas of concern, and the paper thoroughly examines the open issues and limitations faced by Blockchain-enabled Federated Learning. Finally, it discusses the future direction in these areas, primarily focusing on enhancing BCFL's privacy.

Table 1
Summary of related surveys and their significant contributions to Our Work, specifically focused on Privacy attacks and privacy preservation approaches in BCFL.

Survey paper	Literature timeline*	Blockchain and FL background	Proposed general architecture	Privacy attack in BCFL	Privacy preserving approaches in BCFL							Applications
					C_1	C_2	C_3	C_4	C_5	C_6	C_7	
Ali et al. [5]	2019–2020	●	○	○	○	○	○	○	○	○	○	◐
Nguyen et al. [6]	2019–2021	●	●	○	◐	○	○	○	◐	○	○	◐
Huang et al. [9]	2020–2021	○	●	○	◐	○	○	○	◐	○	○	○
Li et al. [4]	2018–2021	●	●	○	○	○	○	○	◐	○	○	◐
Qu et al. [1]	2018–2021	●	●	○	●	○	○	○	◐	○	○	○
Issa et al. [7]	2019–2021	●	●	○	◐	○	○	○	○	○	○	○
Zhu et al. [2]	2019–2022	●	●	○	◐	○	○	○	◐	○	○	◐
Chhetri et al. [11]	2019–2022	◐	○	○	○	◐	◐	◐	○	○	○	○
Qammar et al. [8]	2019–2022	◐	●	○	●	○	○	○	◐	○	○	◐
Our Work	**2018–2023**	●	●	●	●	●	●	●	●	●	●	●

C_1: BCFL architectures for security and privacy protection, C_2: BCFL with differential privacy based approach, C_3: BCFL with homomorphic encryption based approach, C_4: BCFL with secure multiparty computation based approach, C_5: BCFL with reward-driven based approach, C_6: BCFL with hybrid privacy approach, C_7 : Using cross-chain based approach. ○ denotes that the corresponding aspect has not been discussed, ◐ indicates a partial discussion, and ● signifies a comprehensive exploration.
* The interval we reported is not related to the whole period analyzed in the survey but only to the interval for which the authors examined the integration of Blockchain and FL.

3.2. Search strategy

To acquire pertinent information concerning Federated Learning based on Blockchain, we devised a search plan in line with our research objectives. We started by thoroughly exploring Google Scholar and Web of Science. Then, we expanded our investigation to reputable academic repositories such as IEEE Xplore, ACM Digital Library, ScienceDirect, and SpringerLink. Our search spanned publications from 2018 to 2023 to ensure a thorough review of recent research. As a matter of fact, after a first global scan, we noticed that the amount of work about BCFL tends to be negligible before 2018, and we did not find significant publications before that date. Hence we chose a range of six years of analysis. Additionally, we formulated search terms using specific phrases and keywords to cover different aspects of Blockchain-enabled Federated Learning.

We consolidated the search terms using the conjunction operator (AND) to pinpoint relevant studies accurately. Key search terms included “Blockchain” AND “Federated Learning” AND “privacy” and “privacy-preserving in Blockchain-based Federated Learning”. In addition to these primary terms, we incorporated supplementary search terms such as “privacy attack”, “inference attack”, “homomorphic encryption”, “differential privacy”, “secure multiparty computation” and “privacy-preserving in Internet of Things” to enhance the comprehensiveness and scope of the search.

The PRISMA flow diagram in Fig. 1 visually illustrates the iterative screening process, depicting the counts of identified, excluded, and included research works.

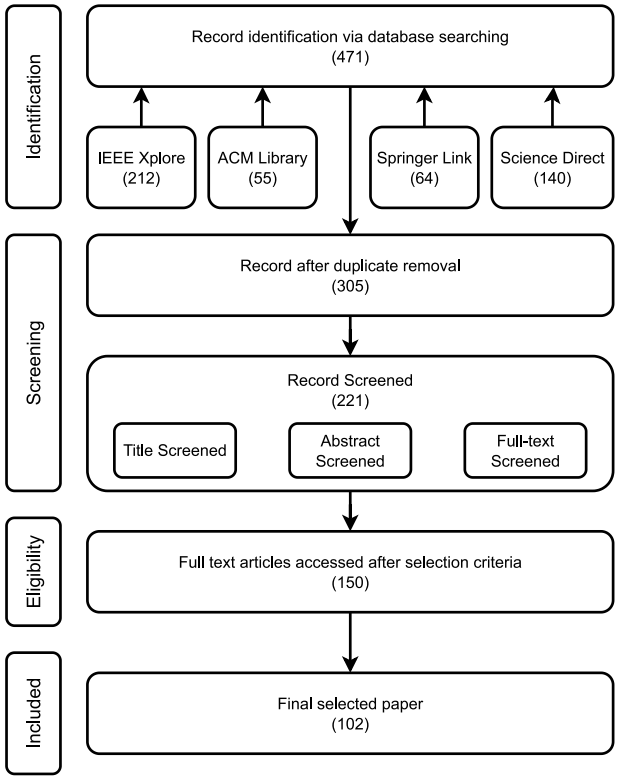


Fig. 1. The PRISMA flow diagram visually outlines the various phases of the systematic review process.

3.3. Selection criteria

This section outlines the criteria employed to assess the relevance and quality of scientific works selected for inclusion in this survey based on our search criteria. A paper is considered eligible for inclusion if it meets at least one of the following inclusion criteria and does not meet any exclusion criteria.

3.3.1. Inclusion criteria

In assessing the relevance of a paper for inclusion in this survey, we consider the following criteria:

- The citation count (primarily relying on Google Scholar [12] and Scopus [13] platforms);
- The age of the paper, we privileged more recent works;
- The paper’s publication venue significance is assessed using Scimago [14] and Core.edu [15] rankings for journals and conferences, respectively;
- The importance of the most representative author of the paper. In particular, we examined all the authors and verified that at least one meets the following: *(i)* a high H-Index (minimum 30), *(ii)* a high number of citations (more than 2,000), and *(iii)* she/he has published numerous papers with significant impact on the domain under analysis (i.e., accepted in good venues in terms of impact factor of the journal, higher than 2.75, and conference paper in core ranking).

Table 2
Number of articles compared with existing surveys and our work, specifically focused on privacy attacks and privacy preservation in BCFL.

Survey paper	Privacy attack in BCFL	Privacy preserving approaches in BCFL							Applications
		C_1	C_2	C_3	C_4	C_5	C_6	C_7	
Ali et al. [5]									3 (2018–2020)
Nguyen et al. [6]		6 (2020–2021)				4 (2019–2020)			5 (2018–2021)
Huang et al. [9]		4 (2021)				4 (2020–2021)			
Li et al. [4]						1 (2019–2021)			8 (2018–2021)
Qu et al. [1]		16 (2019–2021)				7 (2019–2021)			
Issa et al. [7]		8 (2019–2021)							
Zhu et al. [2]		5 (2019–2022)							2 (2019–2021)
Chhetri et al. [11]			5 (2020–2022)	5 (2019–2021)	7 (2019–2021)				
Qammar et al. [8]		4 (2020–2022)				1 (2022)			2 (2019–2022)
Our Work	2 (2020–2023)	14 (2018–2023)	19 (2019–2023)	14 (2021–2023)	5 (2019–2023)	17 (2018–2023)	9 (2020–2022)	4 (2021–2023)	31 (2018–2023)

C_1 : BCFL architectures for security and privacy protection, C_2 :BCFL with differential privacy based approach, C_3 :BCFL with homomorphic encryption based approach, C_4 :BCFL with secure multiparty computation based approach, C_5 :BCFL with reward-driven based approach, C_6 :BCFL with hybrid privacy approach, C_7 :Using cross-chain based approach.

Table 3
List of the acronyms used in the paper.

Acronyms	Description
BCFL	Blockchain-enabled FL
DP	Differential Privacy
FL	Federated Learning
HE	Homomorphic Encryption
IID	Independent and Identically Distributed
IPFS	Interplanetary File System
ML	Machine Learning
SMPC	Secure Multiparty Computation
IoT	Internet of Things

3.3.2. Exclusion criteria

Following the inclusion process, we apply the exclusion process, excluding a paper if it meets any of the following criteria:

- The paper is not peer-reviewed;
- the paper is written in a language other than English;
- The paper is not focused on solutions for Federated Learning-based Blockchain technology nor dealing with privacy;
- The paper lacks significance, as it represents an incremental improvement on a previously proposed approach, a duplicate publication, or an extended version of an already published key contribution;
- The paper's year of publication exceeds six years, and it is earlier than 2018.

4. Background knowledge

This section provides the essential background information to contextualize our survey. In particular, we describe the main concepts related to FL, its workflow, and the principal categorizations and challenges of such an approach. Moreover, we illustrate the fundamental notions about Blockchain. Table 3 summarizes the acronyms used in the paper.

4.1. Centralized learning, distributed learning, vs. Federated Learning

This part explains how ML architectures have evolved, progressing from centralized models to distributed on-site solutions and, most recently, up to Federated Learning (FL) [16].

The classical architecture, illustrated in Fig. 2, is called Centralized Learning. In this strategy, generated data is continuously streamed into the Cloud, where high-performance servers can process them and train models efficiently. Examples of the use of such an approach are provided by popular ML-As-A-Service providers, such as Amazon Web Services,[1] Google Cloud,[2] and Microsoft Azure.[3] In Centralized Learning, data is sent to the Cloud, where the ML model is built. A user uses the model through an API by requesting access to one of the available services. Within this architecture, abundant interactions generate a substantial volume of data. This can lead to privacy issues, latency, as data could be transmitted far away from the central server, and, consequently, high transfer costs.

Some ML tasks are moved to clients with powerful resources to overcome such drawbacks. This more recent strategy, visible in Fig. 3, is called Distributed On-Site Machine Learning architecture. Here, each device owns a local dataset through which it can build its model. After the first interaction with the Cloud to distribute a pre-trained or generic model to the devices, no more communication with the Cloud is needed. Hence, privacy is obtained as data does not leave its hosts. Although popular applications benefit from this architecture, such as medical solutions [17] and smart classrooms [18], models are local, and, therefore, they cannot take advantage of the results of their peers.

In Federated Learning, shown in Fig. 4, each device trains a local model leveraging local data and sends its parameters to the central curator for aggregation. Data is kept on-device, and knowledge is shared with peers through an aggregated model. In this way, FL combines all the advantages of the previous architectures. Indeed, it maintains data privacy while minimizing communication overhead by keeping raw data on devices and aggregating local model updates.

[1] https://aws.amazon.com/
[2] https://cloud.google.com/
[3] https://azure.microsoft.com/

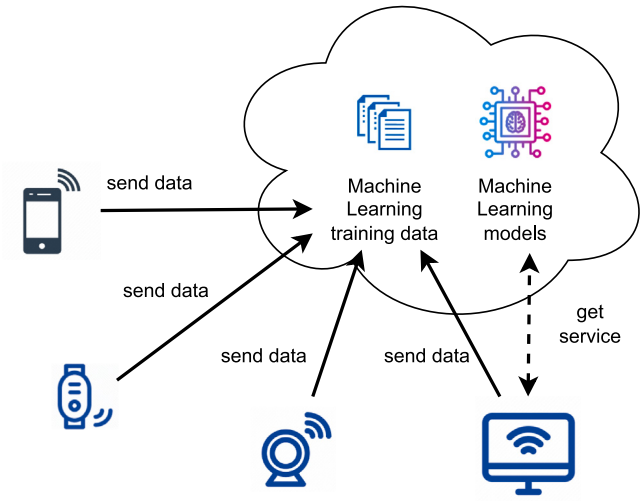


Fig. 2. Centralized ML Architecture.

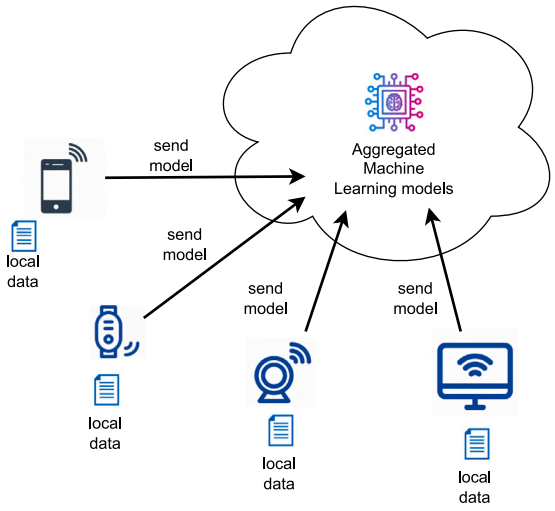


Fig. 4. Federated Learning Architecture.

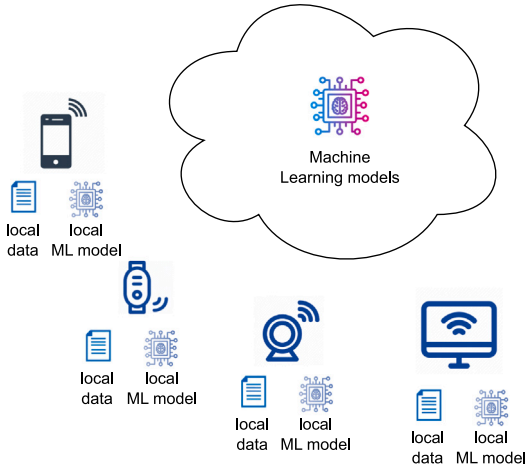


Fig. 3. Distributed On-Site ML Architecture.

4.2. Overview of Federated Learning

In the next sections, we deal with the key notions related to FL and its workflow. Finally, we discuss the primary categorizations and the main challenges inherent to such an approach.

4.2.1. Main concepts and workflow

As stated, FL is a Machine Learning strategy that allows a model to be trained across decentralized devices or servers holding local data samples while maintaining the localized data. This technique is beneficial if data cannot be efficiently centralized due to privacy regulations, network constraints, or large data volumes.

The involved entities are divided into two types: C devices known as "clients" or "workers" (e.g., IoT or remote devices) and a central server called "aggregator". Workers are individual devices, such as smartphones, IoT devices, or remote servers, that announce to the server that they are ready to run local training and participate in an FL task. Every client possesses its local dataset and utilizes it to train a dedicated local model. On the contrary, the central server or aggregator acts as the coordinating entity overseeing the Federated Learning process. More specifically, the goal of FL is to train a global model $\mathbf{w}$ by uploading the weights of local models from workers $\{\mathbf{w}^i \mid i \in C\}$ to a parametric aggregator trying to optimize a loss function:

$$\min_{\mathbf{w}} l(\mathbf{w}) = \sum_{i=1}^{n} \frac{k_i}{C} L_i(\mathbf{w}^i) \tag{1}$$

here $L_i(\mathbf{w}^i) = \frac{1}{k_i} \sum_{j \in I_i} l_j(\mathbf{w}^i, x_i)$ is the loss function, k_i is the local data size of the i-th worker, and I_i identifies the set of data indices with $|I_i| = k_i$, and x_j is a data point. The basic FL workflow consists of the following steps [19]:

- Model initialization: A global ML model $\mathbf{w}$ is initialized on a central server with random parameters. During this phase, workers (e.g., IoT devices or remote servers) are selected to participate in the FL process.
- Local model training and upload: Clients download the current global model to their local devices. Then, they perform local training using their data, which is kept private and not shared with the central server or other clients. The local training typically involves multiple iterations of gradient descent, back-propagation, or other optimization methods to improve the local model's performance. Following the local training, every client computes the model parameter updates and transmits them to the central server either in an aggregated form or encrypted. In particular, at the t-iteration, each client updates the global model by training with their datasets: $\mathbf{w}_t^i \leftarrow \mathbf{w}_t^i - \eta \frac{\partial L(\mathbf{w}_t, b)}{\partial \mathbf{w}_t^i}$, where η and b identify the learning rate and local batch, respectively.
- Global model aggregation and update: The central server collects and aggregates the model parameter updates from all the clients, namely, $\{\mathbf{w}^i \mid i \in C\}$. The central server can employ various aggregation methods like averaging, weighted averaging, or secure multi-party computation to incorporate the received updates from each client. This process enhances the performance of the global model by integrating diverse insights from the individual client models.

Fig. 5 illustrates a schematic diagram of FL workflow with the three phases described above. Observe that the last two steps of *(i)* iterative process of local model training and upload and *(ii)* global model aggregation and update are iterated across multiple epochs, continuously enhancing and refining the global model.

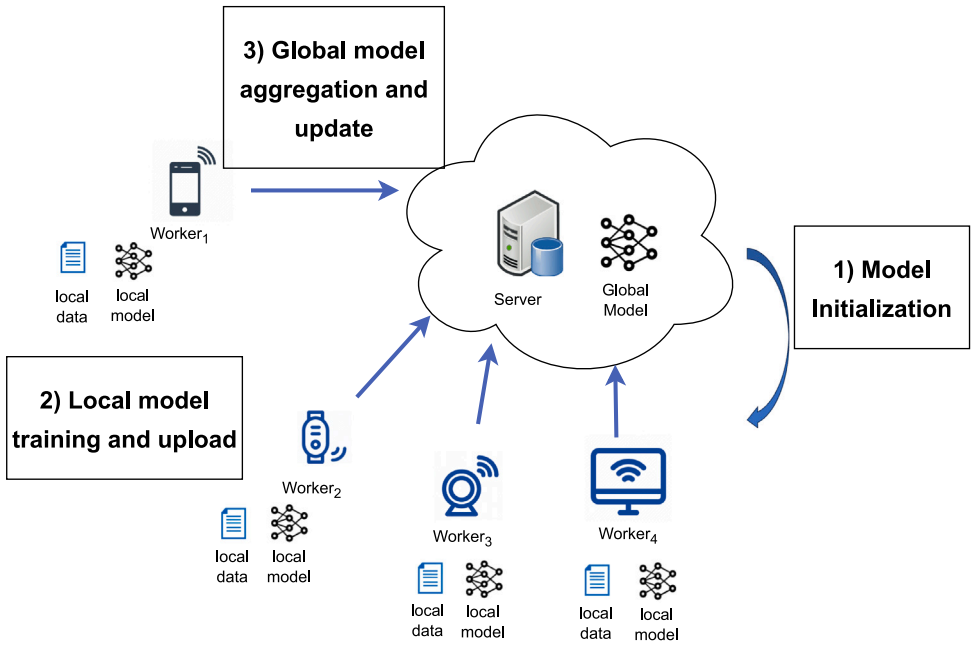
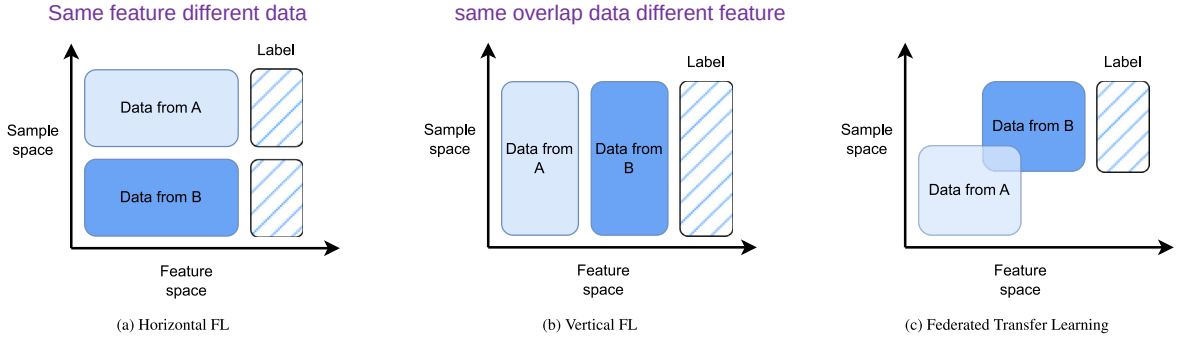


Fig. 5. A schematic diagram of the Federated Learning workflow.



(a) Horizontal FL (b) Vertical FL (c) Federated Transfer Learning

Fig. 6. The three categories of FL divided for feature and sample spaces.

4.2.2. Categorization of FL

This section deals with the different architectures for Federated Learning based on the feature and sample spaces shared by the workers and the aggregating server.

Vertical FL, Horizontal FL, and Federated Transfer Learning. A different perspective to classify FL relates to how data is distributed among the participating parties in the feature and sample spaces [19, 20]. According to this criterion, FL can be divided into Horizontal FL (HFL), Vertical FL (VFL), and Federated Transfer Learning (FTL). Fig. 6 shows a graphic representation of the three FL categories.

- Horizontal FL refers to scenarios where the parties share the same feature space but have different data samples. This schema can be also referred to HFL as sample-partitioned FL.
- Differently from HFL, Vertical FL applies to the case where the actors share overlapping data samples but differ in the feature space. We also refer to VFL as feature-partitioned FL.
- FTL is applicable for scenarios in which there is little overlapping in data samples and features. For instance, the case in which multiple subjects with heterogeneous distributions build models in a collaboratively way.

Types of data heterogeneity FL. In this part, we consider a possible FL classification according to the types of data heterogeneity. In FL, data can be Independent and Identically Distributed (IID) or Non-Independent and Non-Identically Distributed (Non-IID). These characteristics refer to how data is distributed among the different clients [21]. In the ideal scenario of an IID data distribution, the data on each client is assumed to be drawn from the same underlying probability distribution. In real-world scenarios, instead, data on different clients may have different statistical properties, feature distributions, label distributions, and data sizes. In FL, data heterogeneity refers to variations in data distributions among participants. During each federated iteration, participants are selected randomly to undertake the supervised task, incorporating features x and labels y. Subsequently, the local data distribution of the chosen client, denoted as $P_i(x, y)$, is utilized to extract feature-label pairs from (x, y). In particular, the authors of [22] consider the following categories:

- *Feature distribution skew.* It consists of an imbalance or non-uniformity in the distribution of features (input variables) across different devices, clients, or participants. Specifically, this happens when the distribution $P_i(x)$ of the features varies from participant to participant, but the distribution of the probability $P_i(y|x)$ is the same. For example, consider two participants, say i and j, collaborating to build an FL model using data from the EMNIST dataset. Suppose that i owns a training set mainly containing characters with a bold font; whereas j's training set contains mainly characters with an italic font. Therefore, in the data distribution for the first participant, the likelihood of the

character "A" being bold is higher, but for the second participant, it is more likely to be in italic font. Consequently, the feature distribution $P_i(x)$ differs between the two participants. However, $P_i(y|x)$ remains the same for all instances of the letter "A".

- *Label distribution skew*. It means that the distribution of labels $P_i(y)$ is different for different participants, but given $P_i(x|y)$ is the same. Label distribution may vary across participants even when they share the same label annotations. For example, consider two participants, denoted as i and j, containing data from the Fashion-MNIST dataset. In i's dataset, 80% images belong to the class "shirt", while the remaining 20% display other image types. Conversely, j's data illustrates that 85% of the images are "shirts", and the remaining 15% depict the other types. Consequently, the distribution of labels $P_i(y)$ differs among participants. However, focusing explicitly on images featuring "shirts", the probability of the associated features x portraying a shirt remains roughly equal for both participants. Hence, the probability distributions $P_i(x|y)$ are similar.
- *Quantity skew*. It is a common situation that causes data to deviate from a homogeneous distribution, and it refers to the significant difference in the quantity in different participant data $P_i(x, y)$. For instance, participant i has 500 samples, and participant j has 30,0000 samples for training. Therefore, the distribution of $P_i(x, y)$ differs significantly.

Cross-device and cross-silo FL. A further strategy to classify FL approaches is based on the participating clients and the training scale. According to this principle, FL can be divided into cross-device FL and cross-silo FL [23].

The first group consists of clients that are small distributed entities (e.g., smartphones, wearables, and IoT devices) holding few local data. Hence, to obtain good performance, many clients usually need to participate in the training process. Unlike the previous group, cross-silo FL clients are typically big companies or organizations (e.g., hospitals, transportation companies, and banks). In these environments, the number of participants is small (typically 2 to 100 clients), but each client usually participates in the entire training process.

4.2.3. Primary challenges to FL

Most of the scientific papers focusing on FL [24–26] investigate several core open challenges that still need to be addressed, such as:

- *Privacy protection*. One of the primary aims of FL is to guarantee the privacy and protection of data in ML solutions. It is essential that FL model training does not reveal users' private information. Most recent approaches often provide privacy at the cost of reduced model performance or system efficiency.
- *Security*. FL systems must be robust against adversarial attacks or clients with malicious intent. FL has been analyzed through an adversarial lens to study the vulnerability of the learning process to model-poisoning adversaries [27]. Since FL, in its classical form, is susceptible to adversarial attacks, poisoning resilience defense mechanisms should be investigated.
- *Data shortage*. ML algorithms usually demand extensive data for optimal performance, but in a distributed context (such as IoT), involved devices have limited data. Hence, FL needs local data utilization for training on each device, after which the resulting local models are sent to the server and consolidated into a global model.
- *Statistical heterogeneity*. Clients may have different data distributions, and data held by these devices may be non-IID. This makes it difficult to create a globally useful model that performs well on all clients.
- *Expensive communication*. Transmitting model updates between clients and the central server can be resource-intensive, especially in scenarios with high latency or limited bandwidth. Reducing both the total amount of communication rounds and the size of transmitted messages at each round are two main aspects to be considered.
- *Systems heterogeneity*. The presence of heterogeneous devices in terms of storage, computational, and communication capabilities leads to several challenges related to dropped devices in the network, a low amount of participation in the FL framework, and the design of scalable and flexible solutions.
- *Algorithm convergence*. The work presented in [28] describes a theoretical analysis of the convergence bounds of the gradient descent-based FL for convex loss functions. Anyhow, further studies on the optimum number of local workers and on the frequency of local updates and global aggregation to improve model performance and resource preservation should be deeply investigated [26].
- *Lack of incentive mechanisms*. Limited research acknowledges that participants in Federated Learning lack incentives to share their data and train models. As a result, task requesters face challenges in identifying and choosing trustworthy participants with high-quality data [29,30].

4.3. Overview of blockchain

This section is devoted to providing a background description of the Blockchain technology. In the next subsections, we will describe the main concept and workflow, the strategies to build the consensus mechanism, the smart contract technology, and the main Blockchain categories.

4.3.1. Concepts and workflow

In 2008, Nakamoto introduced the revolutionary Bitcoin cryptocurrency [3], which operates as a decentralized and transparent peer-to-peer system. Blockchain, the underlying technology supporting Bitcoin, finds extensive utility across many financial and industrial applications due to its remarkable characteristics. A Blockchain network's most prominent feature is its utilization of a publicly digitally distributed and immutable ledger of blocks, which is shared across all participants in the peer-to-peer network without relying on any centralized trusted third party [31]. Each participant in the Blockchain network retains an individual copy of the distributed ledger to ensure data integrity, and every block contains the previous block's digest and comprises multiple transactions, as illustrated in Fig. 7. Mathematically, a digest is produced by a hash function f, which must satisfy two properties *(i)* the size of the input space and the output space must be large, *(ii)* it must be practically impossible to find collisions, that is, finding two inputs x_1 and x_2 that produce the same output $f(x_1) = f(x_2)$.

In addition to the transaction data and the hash value of the previous block, each block includes a timestamp and a nonce, which is a random number for verifying the hash. Since hash values are unique, changes on any block in the chain would immediately change the respective hash values. Indeed, once generated, the information within each block cannot be altered, ensuring the network's immutability. Whenever a new transaction is generated, it undergoes validation and verification through a consensus protocol carried out by *miners*. If the majority of nodes in the network agree by a consensus mechanism on the validity of transactions included in a new block and on the validity of the block itself, this block is created and seamlessly integrated into the distributed ledger. In summary, as shown in the scheme illustrated in Fig. 8, once created by a client, a transaction goes through several steps [32], namely:

- *Propagation*. The transaction is propagated in a block towards the validating peers.

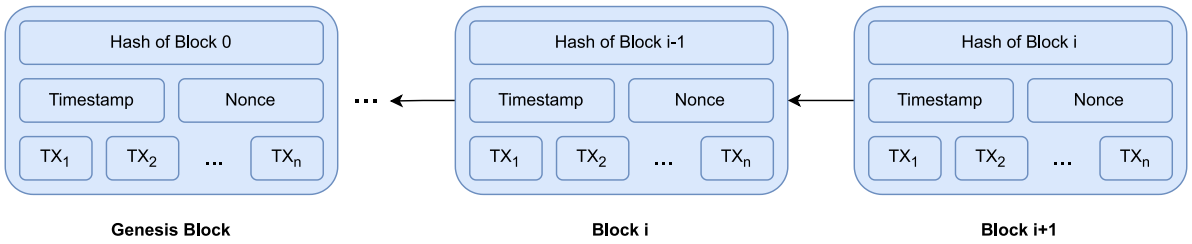


Fig. 7. Example of a Blockchain.

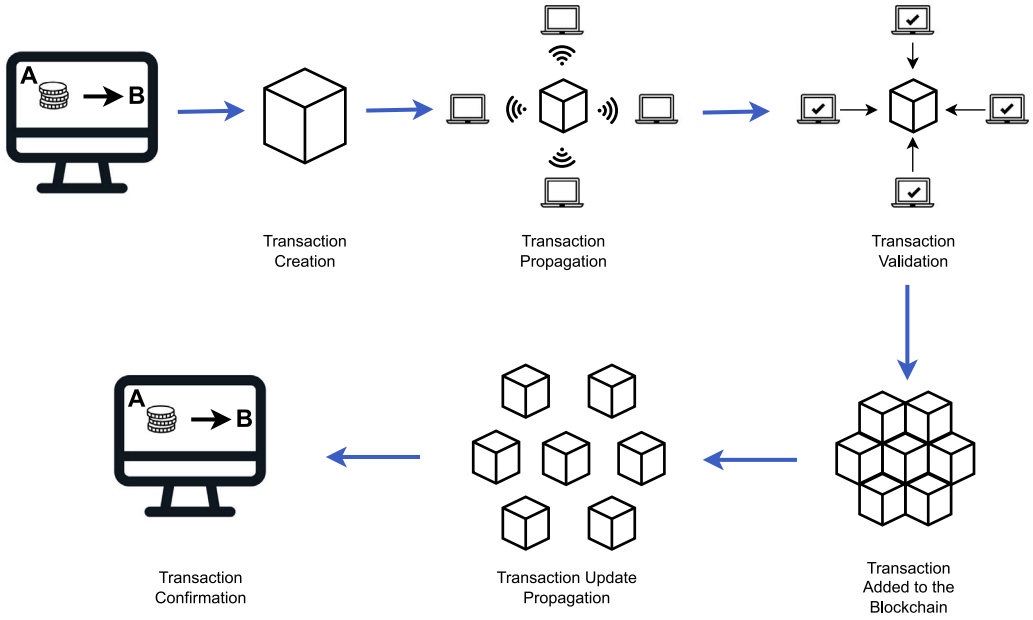


Fig. 8. Transactions workflow in Blockchains.

- *Validation*. The transactions collected in blocks must address the different phases of the consensus mechanism. Thereafter, the block of transactions can be attached to the Blockchain.
- *Update Propagation*. The valid transactions block is propagated throughout the network to let all nodes update their replica.
- *Confirmation*. The consensus procedure comes to an end, and the nodes have to agree on a single chain of blocks. Blocks of transactions are published on the Blockchain and are confirmed in the final version of the ledger, from which they may no longer be discarded.

The key features of this strategy can be summarized as follows:

- *Security*. Blockchain employs advanced cryptographic procedures to keep data secure. Once a transaction is written in a block, altering or deleting it is impractical. This makes Blockchain robust against attacks, such as fraud or tampering.
- *Decentralization*. Blockchain distributes data across a network of nodes. These nodes work together to validate and record transactions. In this way, all the drawbacks of centralized solutions can be avoided, i.e., bottleneck servers or high latency due to excessive resource contention.
- *Transparency*. All transactions on a Blockchain are publicly available to all the participants in the network. This transparency can help build trust among users.
- *Immutability*. Due to cryptographic hashing and chaining of blocks, once a block is added to the Blockchain, it cannot be changed, and the user cannot revoke it.

4.3.2. Consensus mechanism

As already stated in the previous section, Blockchain uses consensus mechanisms, such as Proof of Work (PoW) or Proof of Stake (PoS), to agree on the validity of transactions and ensure that all nodes in the network have a consistent view of the shared ledger [32].

PoW is the older and more widely adopted consensus mechanism, it has been adopted in Bitcoin and many other cryptocurrencies [33]. To validate a block in the PoW approach, miners should find a hash value of the block that meets a certain difficulty requirement as a mathematical puzzle. The winner of this competition can validate the block of transactions and is rewarded with cryptocurrency. PoW does not guarantee consensus finality; transactions can be considered as confirmed only when included in the longest chain. PoW is designed to consume a high amount of energy because of the miners' energy-intensive computations needed to solve puzzles.

On the contrary, PoS has recently gained popularity because it is a less energy-consuming alternative to PoW, and it is used in cryptocurrencies like Ethereum 2.0[4] [34]. PoS is adopted by a category of Blockchain algorithm where the consensus is achieved by stakes (e.g. digital assets) in the network. Validator nodes, which are participants who hold and "stake" a certain amount of cryptocurrency, are chosen in a deterministic and pseudo-random manner to create new blocks and validate transactions based on the amount of cryptocurrency they hold and are willing to "stake" as collateral. Validators are incentivized by earning transaction fees.

[4] https://ethereum.org/

4.3.3. Smart contract

Smart contracts, serving as executable codes, embody a mutual agreement between two or more parties. They operate atop Blockchain to enforce and execute agreements among parties that might lack trustworthiness. These contracts define the rules, conditions, and actions to be taken when certain conditions are satisfied [35]. Moreover, it stores information, processes inputs, and writes outputs thanks to its pre-defined functions. Smart contracts are replicated on each node of the Blockchain network to prevent contract tampering. Platforms like NXT,[5] Ethereum, and Hyperledger Fabric [36] are Blockchain-based development frameworks able to provide smart contracts to execute automatically events and actions.

Once deployed on a Blockchain, a smart contract operates autonomously. Usually, it is initiated by activating its constructor function via a transaction submitted to the Blockchain network. Each contract will be assigned to a unique address of 20 bytes. This constructor function is, then, executed, and the resulting smart contract code is permanently stored on the Blockchain. Once deployed, the creator of the smart contract receives essential parameters (e.g., the contract address). Subsequently, users can trigger any accessible functions within the smart contract by initiating transactions [37].

There are two main groups of smart contracts, namely, deterministic and non-deterministic. The smart contracts that belong to the first type do not require any information from an external party outside the Blockchain. Instead, a non-deterministic smart contract depends on information (called oracles or data feeds) from an external party.

4.3.4. Blockchain categorization

Depending on the characteristics of the Blockchain, researchers and industries have defined several categories.

Private and Public Blockchain. Both public Blockchain and private Blockchain networks are decentralized and shared among their clients to register all peer-to-peer transactions without the presence of a third-party authority. However, private Blockchains are restricted to authorized participants, and a centralized entity controls access. This leads to a very high transaction processing rate with few authorized participants. Moreover, a shorter time is required to get the consensus for the network, and more transactions can be processed within a time unit. Public Blockchain, on the other hand, is an open and permissionless network accessible to anyone, posing a risk to information privacy. However, since each transaction is open for the public to verify, they are very transparent, and the risk of hacking and data manipulation is lower when compared to private Blockchains. For this reason, it can be stated that public Blockchains are generally more secure [38]. A further type of Blockchain is represented by **Hybrid Blockchain** that combines elements of both public and private ones. Some parts of the network are public, while others are private. Finally, **Consortium Blockchains**, like hybrid Blockchains, have private and public features, but they involve various organizational members working together on a decentralized network.

Permissioned and Permissionless Blockchain. Permissioned Blockchains usually involve a consortium of organizations where transactions are grouped, accessed, and verified by authorized gatekeepers instead of anonymous miners. Their implementation is arising within the finance sector [39]. On the contrary, permissionless Blockchains, typically associated with public Blockchains, are open for anyone to join and participate without demanding prior authorization. They represent the first and oldest Blockchain development, in which the hashing of blocks of transactions relies on the work of many anonymous miners competing to solve a complex mathematical algorithm for that block of transactions via trial and error [40].

Cross-Chain Blockchain. Cross-Chain refers to a technology aiming to build a bridge of trust among chains realizing interoperability, interconnection, scalability, data and assets exchange between different Blockchain ecosystems [41]. To solve the challenge of Cross-Chain interaction of Blockchains and realize the free circulation of assets, different solutions have been proposed [42]. The first one is *notary mechanism*, which introduces a trusted third party elected among participants to verify and forward cross-chain messages. The project known as Interledger is the most important based on this mechanism [43]. *Sidechain/relay* implementation, instead, introduces a parallel Blockchain that serves as an intermediary to allow assets to be transferred between different chains. By contrast, the *hash-locking* solution does not request the presence of trusted intermediaries and achieves fair transactions by locking assets and setting both corresponding time and unlocking conditions through the use of smart contracts. Finally, *distributed private key control* relies on distributed nodes to control the private keys of various assets in the Blockchain system. To realize the asset circulation and value transfer between different Blockchain systems, the assets on the original chain are mapped to the cross-chain.

5. State-of-the-art: Integration of FL and blockchain

Blockchain is a promising technology, providing robust and secure solutions for various applications, even when dealing with untrusted entities. In FL, it primarily safeguards user privacy. Consequently, the amalgamation of FL and Blockchain, known as Blockchain-enabled Federated Learning (BCFL), enhances privacy and security in various distributed applications. These applications span sectors such as healthcare, cyber–physical systems, secure vehicular networks, pharmaceuticals, Industrial Internet of Things, and telemedicine [44–46]. BCFL effectively tackles the challenges associated with the FL paradigm by providing a range of valuable features. These include robust authentication and traceability, enhanced privacy, reliable availability, scalability, resilience against byzantine faults, resilience against inference attacks, long-term persistence, and anonymity [1].

5.1. Benefits and characteristics of blockchain-enabled FL

In this section, we explore the advantages of incorporating Blockchain in the FL process. The primary limitation of current FL systems is their dependence on centralized processing, which introduces vulnerabilities such as single-point failure and susceptibility to attacks [47–49]. Additionally, the extensive participation of edge devices contributes to network strain, leading to concerns about bandwidth availability and scalability [45]. Also, Blockchain technology offers a solution by providing decentralization, replacing the central server in FL applications with smart contract execution, enhancing security, and reducing the risk of malicious activities [50,51]. The decentralized nature of Blockchain, preventing any single entity from having control over the entire network, aligns seamlessly with the principles of FL. In FL, data remains on individual devices and only updated models are exchanged, thereby significantly enhancing security and privacy. Nguyen et al. [6] explored combining FL and Blockchain to create a decentralized, secure, privacy-enhanced intelligent edge network.

Furthermore, smart contracts automate and enforce governance rules in FL, ensuring participants adhere to predefined agreements and offering automated and transparent incentives for participants, miners, or validators based on their contributions. These agreements authenticate node contributions, perform global model computations, and facilitate node incentives based on their performance, enhancing the collaborative learning process's efficiency, audibility, and reliability [49,52]. Incentives provided through smart contracts enhance the security and functionality of the Blockchain infrastructure while maintaining transparency and accountability [53,54].

Transactions in BCFL enable participants to trace and verify the complete history of model updates, fostering a culture of accountability within the system [55]. Also, Blockchain's standardized protocols enhance interoperability in BCFL, allowing for seamless integration across various platforms and devices.

[5] https://nxtdocs.jelurida.com/Nxt_Whitepaper

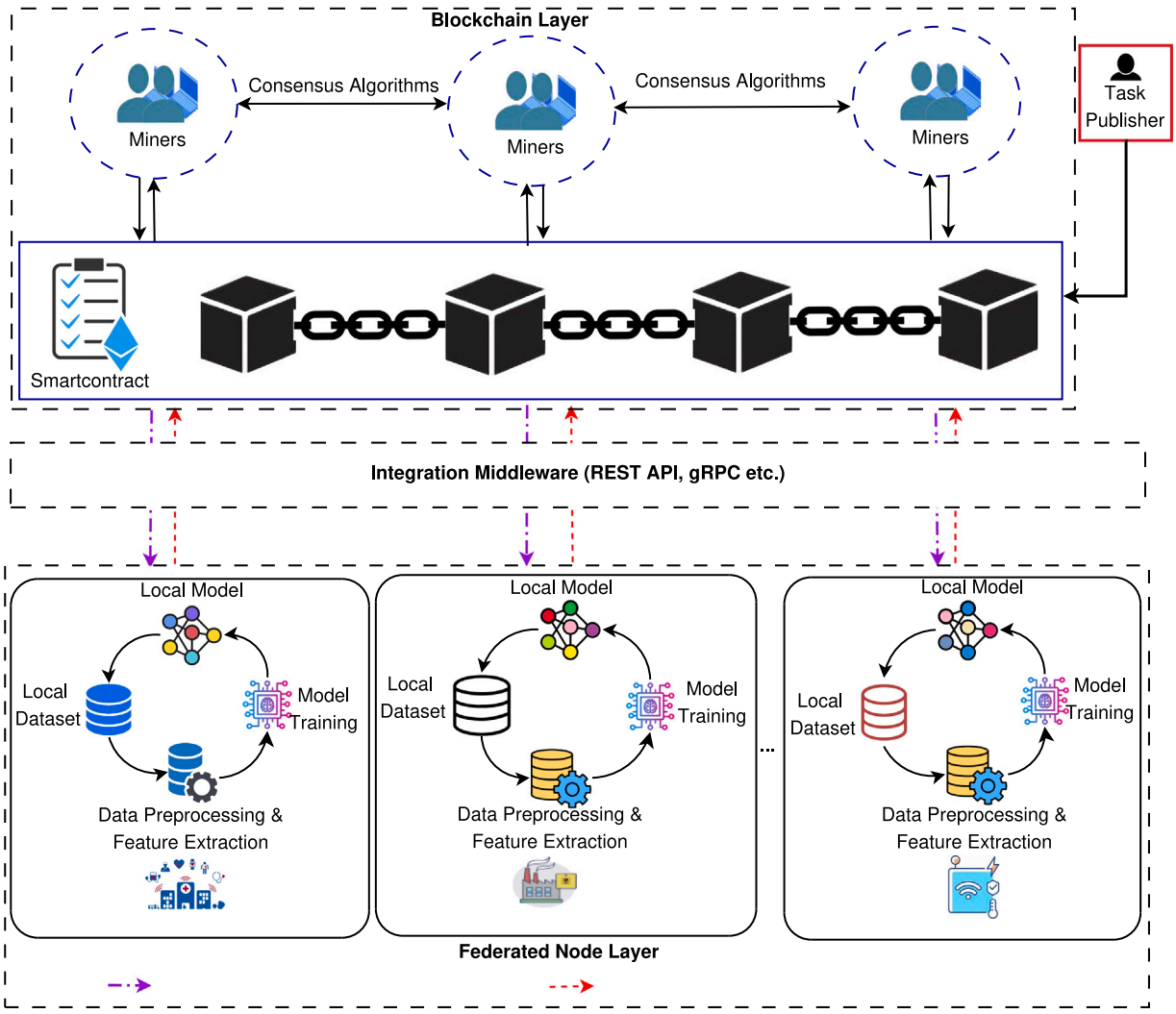


Fig. 9. General architecture for the Blockchain-enabled Federated Learning.

5.2. The general architecture of blockchain-enabled FL

The general abstracted architecture for Blockchain-enabled Federated Learning (BCFL) is illustrated in Fig. 9. This architectural framework comprises three distinct layers: the Federated node layer, integration middleware, and the Blockchain layer. In decentralized applications on a Blockchain, the task publisher, as an entity or user, initiates and creates tasks, actively defining and structuring them within the Blockchain. This role is pivotal in task creation and execution, starting by formally requesting a specific Federated Learning task and publishing the details into the Blockchain. Subsequently, participants expressing interest in the FL tasks retrieve the models from the Blockchain and contribute their trained models back to the Blockchain. The Blockchain then operates as a central server, employing smart contracts that aggregate the models from participants. A designated miner executes this operation and creates the new global FL model to fulfill the specific FL task. In the subsequent sections, we will briefly introduce each component in detail.

Task Publisher: The task publisher initiates the process by formally submitting a request for a specific Federated Learning task, meticulously outlining the parameters, requirements, and objectives. This encompasses the task publisher's identity, initialization details (such as the Machine Learning model type), targeted performance metrics for optimization, expected processing time, and other relevant information. Furthermore, it encompasses additional crucial parameters, such as the task's initiation time, the number of federation rounds, the total reward amount, and other relevant details. The task publisher submits details of the Federated Learning task into the Blockchain for securely and transparently storing information for participants interested in contributing to or downloading models related to the specified task. In [24], the manufacturer is a task publisher to develop a smart home system. In [56], the task publisher refers to enterprises, research institutes, or healthcare research units aiming to acquire a medical disease detection model.

Federated Node Layer: For the collaborative training of an ML model, the Federated node layer encompasses a varied group of participants, including diverse devices such as smartphones, wearables, servers, and other computing entities. Participants in the FL task download the model from the Blockchain. Each participant has their private local dataset and performs data preprocessing and feature extraction on its local dataset. Preprocessing may involve cleaning the data, normalizing, handling missing values, and extracting relevant features contributing to the model's learning process. Following this, participants individually train their models using their local datasets. After the training process, participants in FL produce personalized model updates specific to their datasets. Subsequently, participants submit these local model updates for verification and aggregation in the subsequent phase into the Blockchain.

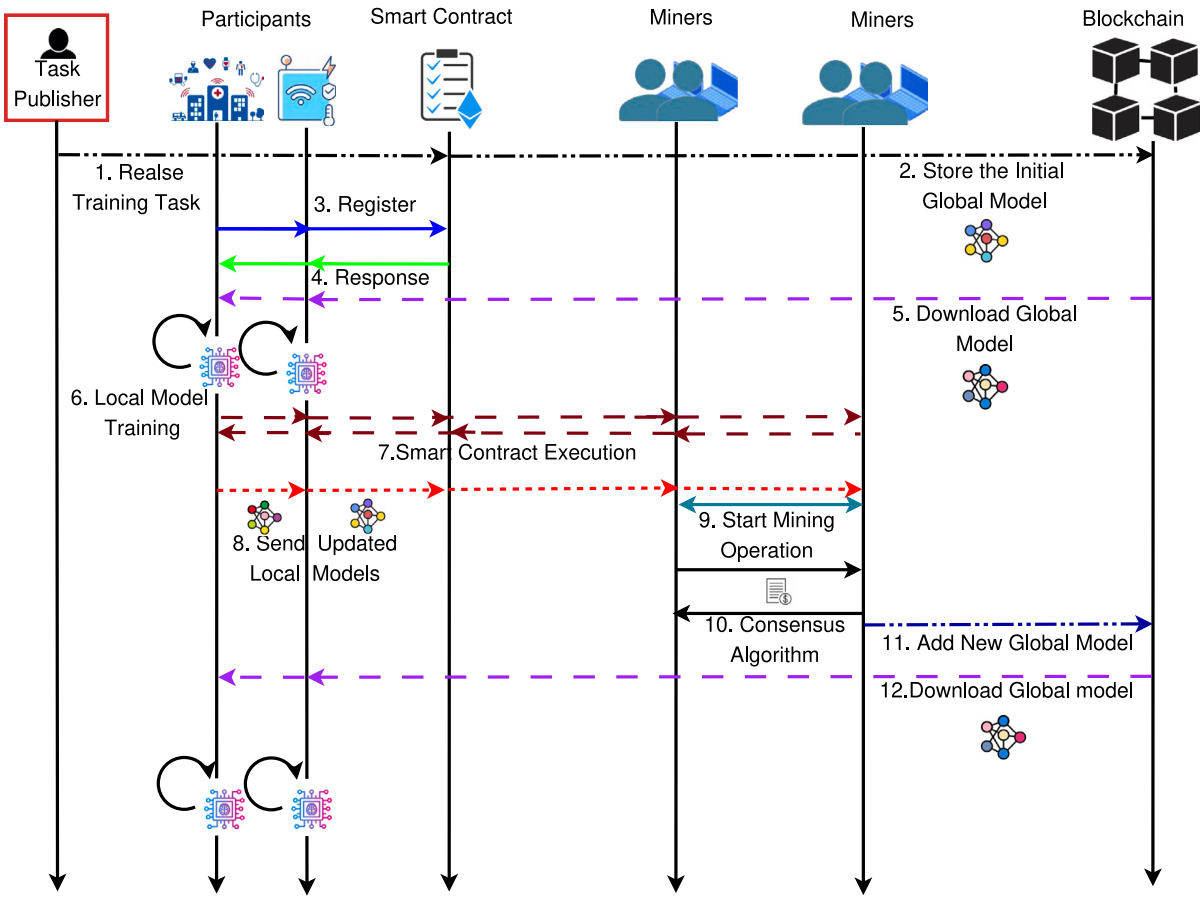


Fig. 10. The high-level workflow for a single epoch Blockchain-enabled Federated Learning.

Integration Middleware: The integration middleware bridges FL participants and the Blockchain. Lamken et al. [57] employed the Representational state transfer application Programming Interface (REST-API) to engage with Blockchain (Hyperledger Fabric) to enable a systematic allocation of network resources for recording and incentivizing gradient uploads. Additionally, the gRPC API, a remote procedure call (RPC) protocol developed by Google, facilitates model transfer between FL participants and the Blockchain network (Ethereum) [8].

Blockchain Layer: In the Blockchain layer, pivotal elements encompass smart contracts, miners, consensus protocols, and the underlying Blockchain networks. The smart contract, another key component in Blockchain networks, operates between parties to facilitate interactions within the decentralized system. The participants utilize smart contracts (Registration Contract) to register for FL model training, ensuring transparency and immutability of conditions. After a successful registration, the revised local model is transmitted to the miners. The miners, encompassing personal computers, cloud-based nodes, or standby servers, willingly adopt the mining software. Their primary responsibilities involve receiving local model updates (local weights or local gradients) transmitted by FL participants. Furthermore, miners verify and authenticate the trained local model using the consensus algorithm, which may involve Proof of Work (PoW), Practical Byzantine Fault Tolerance (PBFT), Proof of Stake (PoS), and more. Once verified, the connected miners receive updated local models from FL participants, aggregate these models, add the new updated model into the block, and subsequently upload the block onto the Blockchain network.

5.3. High-level workflow of blockchain-enabled Federated Learning

The high-level workflow for a single epoch in Blockchain-enabled Federated Learning is depicted in Fig. 10. The system iterates these procedures until the model converges or attains the designated federation round.

1. The task publisher initiates a service request by publishing a training task defining the parameters and details of the Federated Learning task. Following this, the task publisher deploys a smart contract to represent and regulate the Federated Learning task. This smart contract encapsulates the requisite rules, conditions, and parameters that govern the execution of the task.
2. Then, the task publisher publishes a training task on the Blockchain.
3. Participants willing to contribute to the Federated Learning task enroll through the smart contract. This registration process guarantees participants' adherence to the terms and conditions outlined in the smart contract.
4. The smart contract processes the registration request, validating participant information against predefined rules. If successful, it generates a response message to acknowledge the registration, given that it fulfills the required criteria.
5. The FL Participants download the global model from the Blockchain.
6. The FL Participants train the model utilizing their individually preprocessed local datasets.
7. During smart contract execution, the contract facilitates interactions among the Blockchain network, FL participants, and

miners. FL participants transfer their local model updates to the miners, and the smart contract verifies the registration and validity of participants.
8. Subsequently, participants upload their local model updates to the miners on the Blockchain. Once sufficient participants are reached, the miners, in turn, validate and authenticate these local model updates.
9. The miners receive updates to the local models from registered Federated Learning participants and subsequently verify the received local models.
10. Each miner actively engages in the consensus algorithm by competitively solving complex puzzles to earn the role of a temporary leader. These temporary leaders then execute a smart contract for local model aggregation, collectively generating a new block that encapsulates information about the updated global model. Subsequently, the newly created block is disseminated to all miners within the network.
11. Finally, a fresh block is appended to the Blockchain network, encapsulating details of the updated global model.
12. FL participants request and download the latest global model for further training.

The Blockchain functions as a secure and decentralized ledger, originating from the Bitcoin network, that permanently records transactions through a chain of blocks containing relevant information. There are two primary classifications for Blockchain storage: on-chain storage, which consolidates all records within a single ledger, and off-chain storage, where the trusted third party stores the data externally, notably through the InterPlanetary File System (IPFS), employs a decentralized and private storage system. IPFS, a peer-to-peer distributed file system, prioritizes content-based addresses, storing hashes on the Blockchain for efficient retrieval. It offers permanent data storage, version traceability, speed enhancements, reduced bandwidth waste, and serves as a decentralized cloud storage solution, mitigating the risks associated with centralized servers. Several researchers have successfully incorporated IPFS to store actual models (Local and Global models), ensuring immutability by sending the corresponding hash values to the Blockchain [24,58,59].

5.4. *Existing BCFL solutions and their limitations*

BCFL offers several benefits (as discussed in Section 5.1) but faces certain obstacles that could renounce its effectiveness. This section will discuss the key challenges of using Blockchain with Federated Learning across various applications. Designing a BCFL system involves trade-offs between security, privacy and trust, computational cost, training efficiency, performance, energy consumption, and latency.

Otoum et al. [60] developed an adaptive framework combining BCFL with reinforcement learning for secure IoT-based smart cities. It trains local models on end devices, storing local model updates in a Blockchain by applying a trust scoring mechanism to select the trusted devices. The framework maintains high accuracy and detection rates (0.93 and 0.96, respectively). Although the system provides security and trust, there is an upward trend in latency and energy consumption if the network size increases.

The solutions presented in [24,47,61] offer high global model accuracy and resist local model poisoning and privacy attacks through BCFL. However, they also exhibit limitations, including communication overhead, increased energy consumption, and security vulnerabilities as the number of participants increases. Nonetheless, the solution is impractical for resource-constrained devices due to the computational burdens associated with various privacy and security methodologies utilized in their studies. In [62], authors addressed security and privacy issues for the Internet of Vehicles, but this introduced the communication overhead and delays in communication.

The proposal in [63] prioritizes edge computing latency and communication latency during parameter transfer in BCFL. In contrast, [64] proposed schemes that analyzed communication costs by considering communication latency with miners, mining operation latency, and on-device training latency [65]. On the contrary, solutions in [64,65] demand considerable bandwidth and high energy consumption, which rely on the PoW consensus protocol mining operations [8].

In BCFL, exchanging local parameter updates among network peers can lead to privacy leaks [66]. While many approaches utilize differential privacy to address this challenge, this often impacts training convergence and overall system performance [67]. Integrating specific privacy preservation algorithms into BCFL significantly slows the system's processing speed, complicating the implementation of robust privacy protection mechanisms. These challenges in BCFL necessitate further exploration to achieve an optimal balance between algorithm performance and system efficiency in distributed environments. Imbalances may compromise model accuracy or fail to provide sufficient privacy for the participants [68]. The authors in [7] highlighted that more effective strategies are required to balance communication and computation cost, data utility, security, privacy, and trust in BCFL.

Additionally, hybrid privacy preservation techniques, such as combining BCFL with multiple methods like homomorphic encryption and differential privacy, show promise but still require substantial computing power and face trade-offs between global model performance and privacy level [69]. Homomorphic encryption increases computational demands compared to plaintext operations, leading to slower execution times and reduced system efficiency.

Communication hurdles arise due to unbalanced and non-IID training data across clients due to variations in the size and distribution of the data. If the number of clients exponentially grows, direct communication for parameter updating between clients and a server becomes a bottleneck, leading to network congestion [65]. This congestion causes delays and higher probabilities of data loss during communication rounds, resulting in slower and less accurate FL training convergence. Also, increasing skewness of the non-IID data has a progressively detrimental impact on the system performance.

Moreover, integrating Blockchain affects FL training by reducing transaction processing speed, necessitating efficient solutions. FL often handles sensitive data subject to regulatory requirements, such as the General Data Protection Regulation (GDPR), Health Insurance Portability and Accountability Act (HIPAA), or other financial regulations [70]. BCFL systems must ensure compliance with these regulations while operating in a decentralized and trustless environment. Adhering to regulations involves implementing strong security measures to safeguard sensitive data during BCFL aggregation. Techniques like DP and HE can ensure data privacy in BCFL and strong access restricting protocols such as Zero-Knowledge Proof (ZKP) for ensuring private access with extended computations [71]. Also, techniques such as FL with secure aggregation and secure multi-party computation play crucial roles in adhering to the regulations. Continuous research is vital to refine techniques and frameworks for achieving compliance while benefiting from BCFL, ensuring its secure and ethical adoption across industries. Implementing privacy-enhancing techniques and mechanisms for data governance and compliance auditing is essential to address regulatory challenges [72].

6. Attacks to privacy in blockchain-enabled FL

Existing works have shown that approaches based on FL are vulnerable to attacks against data privacy. In particular, malicious actors can be identified both in:

- The server aims to infer sensitive information from local updates of the participants over time.
- The workers can infer other participants' sensitive information.

Recent works have demonstrated that only by gradient observation a malicious attacker can successfully steal the training data, causing a deep leakage and revealing sensitive information both to a third party or the central server [73,74].

The known attacks against data privacy that may lead to information leakage and data breaches are *(i)* background knowledge attack, *(ii)* collusion attack, and *(iii)* inference attacks [1].

Background knowledge attack is a privacy-oriented attack in which an adversary leverages external information or prior knowledge to gain insights into the data used in the FL process [75]. In particular, a worker regularly obtaining updates to the global model from a central authority might initiate background knowledge attacks by exploiting the differences. This results in a certain degree of privacy compromise.

Collusion attack is a particular class of background knowledge attack, where malicious participants (that may be both devices or servers) collaborate to compromise the privacy of the model, leveraging background knowledge for aggregation [74]. If the main aim is to reveal sensitive information to reconstruct individual data samples or learn specific patterns present in the data, this attack realizes a model inversion. Instead, if participants aim to determine whether a specific data sample is part of the training set, this attack is a membership inference attack.

Inference attack aims at extracting sensitive information about the data used in the training process by examining its outputs [76]. It can be divided into reconstruction attacks and tracing attacks. In the former group, adversaries try to deduce sensitive information, specific attributes, or characteristics of the training data. For instance, in [77], a gradient inversion attack is presented. Instead, in tracing attacks, the attacker wants to determine the existence of an individual in a specific dataset. To this last group belongs the attack proposed by Shen et al. [78] that exploits unintended property leakage to enable a server to infer a set of participants with target properties. Unlike collusion attacks, where participants work together to compromise the system, inference attacks typically focus on using only information that can be deduced from the FL model's predictions or other outputs.

In the following sections, we describe recent works providing effective defenses to these attacks in scenarios involving Blockchain technology and adopting several privacy-preserving techniques while maintaining the collaborative nature of the FL paradigm.

7. Solutions for privacy preservation in blockchain-enabled FL

Ensuring the security and privacy of user data within the framework of Blockchain-enabled Federated Learning is a vital objective, as it necessitates a delicate equilibrium between collaborative machine learning and data protection. This section delves into various strategies and techniques to tackle the issues of safeguarding privacy in the FL empowered by Blockchain. However, these studies have collectively addressed the imperative challenge of safeguarding privacy through diverse means. Some suggested solutions involve employing homomorphic encryption, differential privacy, secure multiparty computations, incorporating reputation-aware BCFL, and, in some instances, combining these methods to fortify privacy measures. These approaches are designed to uphold the global model's accuracy, safeguard participants' privacy, and minimize the influence of malicious local updates.

7.1. Blockchain-enabled FL architectures for security and privacy protection

Numerous studies have introduced various approaches to incorporating Blockchain technology into Federated Learning, primarily focusing on enhancing privacy and security protection. This section explores strategies and techniques to address the challenges of ensuring privacy and security in FL empowered by Blockchain. These architectures guarantee user data protection while enabling collaborative machine learning within the Blockchain ecosystem.

Furthermore, the integration offers the benefit of reducing the risk of a single point of failure attributed to the centralized aggregation curator. For example, in [79], the authors introduced a BCFL system to bolster privacy and security and mitigate the risk of a single point of failure within fog computing. The study examines an attack model where adversaries attempt to manipulate training output by replacing the global model before update transmission. Achieves these goals by modifying fog servers to store global updates on the Blockchain, allowing end devices to maintain global learning models through distributed consensus, and saving only pointers on the Blockchain. In contrast, data is stored in an off-chain distributed hash table. Singh et al. [80] presented an alternative framework for safeguarding the privacy of IoT healthcare data through BCFL, aiming to minimize resource requirements while maintaining model accuracy and enabling fair compensation. The study also discusses the potential enhancement of this approach by not solely relying on the protocol but by incorporating a trust model and a novel consensus method within the Blockchain to support nodes. Xu et al. [81] presented a privacy-focused personalized reliability prediction model for IoT using Federated Learning Neural Collaborative Filtering (FNCF), offering user privacy protection and personalized predictions, along with context awareness and improved convergence speed via local model training.

The authors in [46] ensure privacy and network security within vehicular networks by enabling local model training on end devices, eliminating the need to share data with the edge server. They used practical Byzantine Fault Tolerance (pBFT) for reliable model training. The proposed framework has demonstrated remarkable performance, minimal energy consumption, low latency, high throughput, long lifetime rate, and high accuracy, approximately 97%. Lu et al. [63] introduced a system to reduce communication latency and enhance reliability in BCFL within edge computing. This system integrates Blockchain technology using a consensus mechanism called Delegated Proof of Stake (DPoS) to create a decentralized training network. The system assesses latency by considering local training costs, model aggregation, parameter transmission, and block verification. It employs a deep reinforcement learning algorithm with multi-agent to optimize latency while meeting learning accuracy and bandwidth constraints. The study in [82] presents a comprehensive framework for enhancing CT image recognition, focusing on COVID-19 detection while preserving privacy. It includes data normalization for diverse hospital data, uses Capsule Network-based segmentation and classification for precise patient identification, employs collaborative model training with BCFL, and achieves 98.68% accuracy. The authors in [83] present a lightweight encryption strategy based on Blockchain combined with Federated Learning. This integration aims to bolster the security and privacy of electronic health records (EHR) kept within a decentralized cloud system. The approach ensures protected access for authorized users, minimizing potential attacks on EHR data. They utilize active smart contracts to facilitate secure data transfer and validate the system's efficacy on an Ethereum-based testbed, showcasing its effectiveness. Moreover, they utilize Google Firebase to store the models.

Presently, solutions focus on favoring the selection of portable, honest local models rather than promptly and efficiently detecting Byzantine models and identifying attackers. This is mainly due to verification delays, exposing significant security risks, especially concerning untrustworthy edge and potential Byzantine attacks. To solve these issues, the authors in [84] proposed that BytoChain enhances model verification efficiency by employing verifiers to perform parallel verification workflows and employs a consensus mechanism called Proof-of-Accuracy (PoA) to detect byzantine attacks. It offloads the verification burden from miners by using verifiers for parallel verification workflows and introduces PoA to detect inferior models while preserving accuracy. The framework proposed in [85] addresses both byzantine-robustness and inference-resistance. Utilizes permissioned Blockchain to replace the central curator, ensuring decentralized trust and fairness while protecting participant privacy. It employs private

data collections in Fabric, supports multiple learning and prediction channels, and includes vertically partitioned secure aggregation to evaluate local model updates. This process calculates updated coordinate weights through Euclidean and cosine measures and determines new global model parameters via weighted averaging. Additionally, a secure prediction mechanism enables third-party applications to query the global model by securely processing raw data across peers before aggregating results for predictions.

Furthermore, Blockchain-based asynchronous FL aims to enhance reliability and security by introducing decentralized and transparent training processes [5]. However, traditional Blockchain consensus algorithms are either computationally intensive or communication-intensive, hindering efficiency, and committee-based algorithms like DPoS [86] may not be ideal for smart public transportation. The work of [87] presents a novel asynchronous BCFL system tailored for intelligent public transportation, integrating a dynamic scaling factor and a unique committee-based consensus algorithm to enhance reliability while minimizing communication overhead. Specifically, the committee leader, acting as the aggregation server, identifies low-accuracy local models from its local dataset to guard against poisoning attacks. Without requiring communication and voting, a new committee leader is periodically elected from roadside units based on the latest block's hash to reduce the vulnerability to DDoS attacks. Additionally, during the aggregation process, a dynamic scaling factor is employed to allocate suitable weights to local models based on their accuracy, subsequently improving the Learning performance of FL. Feng et al. [88] introduced BAFL as a novel asynchronous strategy designed to expedite Federated Learning. BAFL incorporates two policies to enhance its workflow: one regulates the block generation rate to minimize Federated Learning delays, and the other dynamically adapts training duration to avoid transaction overloads. In contrast to the conventional FedAvg, BAFL employs an entropy weight method to evaluate device participation and records it in the Blockchain for trust. It also employs Pareto optimization to reduce model energy consumption and local device delays, striking a balance between model update speed and transaction delays. Sarhan et al. [89] presents a hierarchical BCFL framework for secure and privacy-preserving collaborative IoT intrusion detection. With a smart contract, transactions (model updates) and processes are executed on a secure Blockchain, enhancing system security and reliability through task compliance verification.

The BCFL architecture removes the necessity for a trusted server in edge environments by utilizing Blockchain to enhance trust among participants and enable all users to verify the training process and maintain transparency. The authors in [90] proposed an innovative approach to facilitate collaborative learning in trustless edge computing environments. This strategy introduces a novel paradigm that includes a sandbox and a state channel for creating a secure FL environment, effectively tackling concerns regarding data privacy and quality. In addition, this approach employs smart contracts to incentivize local device and edge node participation to enhance node selection performance further. At the same time, they are utilizing a Deep Reinforcement Learning (DRL) node selection mechanism to enhance accuracy and efficiency. In [91], they proposed a decentralized BCFL architecture that enhances security and privacy by utilizing secure global aggregation and also employed a byzantine fault tolerance consensus protocol, which effectively safeguards against attacks from malicious servers and devices. However, the authors formulate a network optimization problem to mitigate potential long training latency that jointly considers bandwidth and power allocation. They propose transforming this problem into a Markov decision process and employing a DRL-based algorithm for adaptive and efficient resource allocation. It employs a twin delayed deep deterministic policy gradient algorithm that handles continuous optimization variables for long-term resource allocation.

7.2. Privacy preservation using BCFL with differential privacy approach

Differential privacy incorporates randomly generated noises into data to enhance privacy and prevent precise inference of sensitive information. In this approach, noise is introduced into the client's local parameters, ensuring the perturbation or encoding of responses independently before submission to the central curator and effectively thwarting adversaries from inferring sensitive data. However, this approach minimizes communication and computational overhead compared to cryptographic approaches. A random function $\mathcal{K}$ provides (ϵ, δ)- differentially private for $\delta \geq 0$ if, for any pair of datasets D and D' differing in at most one element, and all $C \subset Range(\mathcal{K})$ [92].

$$P[(K(D) \in C)] \leq e^{\epsilon} \times P[(\mathcal{K}(D') \in C)] + \delta \tag{2}$$

The equation shows a probabilistic inequality where the likelihood of random function $\mathcal{K}$ producing a result in set C with dataset D is limited by e^{ϵ} times the probability of obtaining a result in set C with a different dataset D', emphasizing that using D' increases the chance of results in C compared to D by a factor of e^{ϵ}. Where ϵ denotes the privacy loss and δ denotes the probability of error for the differential privacy algorithm. DP is categorized into Central Differential Privacy (CDP) and Local Differential Privacy (LDP). CDP relies on user trust in the data curator, incorporating random noise into the original aggregated model for privacy protection. Conversely, LDP ensures privacy without relying on trust by having individuals perturb or encode their local models. However, precise implementation of LDP is crucial to avoid inaccuracies in estimated frequencies, given that each individual independently perturbs their response [93].

To ensure privacy in BCFL, Lu et al. [44] suggests a novel framework for applications beyond 5G, emphasizing improved security and privacy through Blockchain integration. It also introduces a DRL optimization strategy to reduce resource costs, learning time, communication expenses, and parameter quality validation. Additionally, they added random noise in local updates for each participant to ensure privacy and tackle resource allocation challenges by optimizing resource consumption and learning quality. The system employs a DPoS protocol to validate transactions and, in the event that no consensus is achieved within a specified timeframe, diverts the computation to alternative edge servers for faster processing. Utilizes a DPoS protocol for verification, rerouting computation to alternative edge servers if consensus is not reached on time. This is followed by aggregating and verifying updates before uploading to the Blockchain, with added security through encryption using the aggregator's private key. In another work, Qi et al. [94] introduced an enhanced GRU neural network tailored for traffic flow prediction. They integrate a consortium Blockchain to decentralize the FL process, ensuring that local model updates are validated by trusted consensus nodes instead of relying on a vulnerable central server. This approach effectively mitigates security risks for the central server and participating individuals. Furthermore, they implement local differential privacy by introducing Gaussian noise to local model updates, significantly bolstering location privacy protection and thwarting malicious attempts at inferring participant information through membership inference attacks.

Wang et al. [95] proposed a secure and decentralized learning network for a mobile crowdsensing system utilizing unmanned aerial vehicles (UAVs), allowing UAVs to securely share model updates and verify contributions without needing a central server. Furthermore, it incorporates local differential privacy to safeguard the privacy of UAVs' updated local models and maintain privacy. Additionally, it incorporates a two-tier reinforcement learning-based incentive system to encourage the sharing of high-quality models among UAVs, even when network parameters are not explicitly disclosed. Xu et al. [45] proposed a novel BCFL model for the Industrial Internet of Things, incorporating adaptive differential privacy to safeguard local model privacy without compromising accuracy. They used the Laplace mechanism, which relies on local DP, to introduce noise into the intermediate

parameters during the model update phase. The cropping threshold can adapt automatically based on the training progress, effectively minimizing the influence of additional noise on model accuracy. The model further implements model parameter validation and proof of contribution consensus to effectively detect and prevent malicious node poisoning attacks, ensuring fairness through reputation and incentive mechanisms. A node reputation system is designed to assess participant reliability, calculated using a multi-subjective logic model. It serves as the basis for consensus committee election and incentives, enhancing overall fairness among participating nodes.

The authors in [24] designed a privacy-preserving BCFL for home appliances. They employed DP on locally trained customer models specifically to the gradient of the local model using the regularization method, and selected customers acted as miners to aggregate the model. They suggested using the IPFS for off-chain model storage to address limited storage, recording their hashes in the Blockchain. It also introduced a novel normalization technique for improved accuracy and proposed an incentive mechanism for rewarding honest customers. Attained a minimum accuracy of 90% but highlighted the existence of a trade-off between accuracy and the level of induced noise. In [96], it introduces a novel approach to tackle crucial challenges within edge computing environments by BCFL with DP facilitated by Wasserstein Generative Adversarial Networks (WGAN) in B5G networks. Minimize communication overhead between edge devices and the cloud, address data falsification concerns, and promote a collaborative data-sharing approach. WGAN generates controllable random noise that complies with DP requirements and is injected into model parameters, bolstering the privacy and security of local model data. Applying game theory to attain Nash Equilibrium among the generator, discriminator, and DP-identifier enhances the overall efficacy. In [97], they introduced a novel BCFL approach, integrating generative adversarial networks and differential privacy (GAN-DP) for privacy and decentralization in Delay-Tolerant (DT) networks. They used a modified Isolation Forest to detect and remove falsified local models. They employed an improved Markov decision process to select optimal DTs for flexible asynchronous aggregation. GAN-DP addressed privacy concerns and encouraged end devices to contribute sensitive data, enhancing system performance. It also supported local data augmentation, mitigating size and class balance issues, improving learning efficiency, and reducing operational costs. Cui et al. [98] designed an innovative GAN-driven differentially private algorithm to protect the privacy of local model parameters by adding controlled noise, ensuring compliance with differential privacy requirements while improving the utility of the anomaly detection model in IoT Infrastructures.

Safeguarding medical records data represents a crucial challenge in the modern digital age, demanding advanced protective measures as cyber threats evolve. In [99,100], proposed BCFL to enhance patient data privacy in healthcare applications along with DP noise added into the local models. The system addressed storage efficiency by storing only the hash value on IPFS within the Blockchain while the original data was kept locally. Liu et al. [101] introduced a cross-layer architecture, employing differential data sharing for origin data and model providers. Their targeted incentive mechanism, designed as a two-stage Stackelberg game, optimizes utility, enhancing privacy and speeding up performance, surpassing the simple shared model and data schemes by 1.72 and 2.59 s, respectively. Furthermore, Laplace differential privacy protects intermediate privacy parameters during aggregation. Li et al. [102] proposed an architecture to enhance FL privacy and security while dealing with lazy clients and SPoF issues. It introduces a bounded loss function to analyze the relationship between block creation and the impact of lazy clients on training efficiency. Optimizing the loss function improves performance despite the presence of lazy clients. Also, it provides learning incentives by optimizing computational resource allocation and ensuring data privacy through differential privacy. In [103], the authors have presented a lightweight authentication framework tailored for BCFL. This framework incorporates a flexible Blockchain consensus algorithm and zero-knowledge proof to validate the identity of participants. Furthermore, an adaptive model aggregation algorithm, considering both the model's quality and the contribution of each node, is employed to boost overall performance, thereby attaining a high level of training accuracy. The Laplacian mechanism for differential privacy protection is applied in intermediate gradients to protect local data privacy from inference assaults while reducing the possibility of data leaking.

Numerous researchers are developing custom Blockchains for various applications, including exchanging and verifying local model parameters in IoT-based Federated Learning. For instance, Salim et al. [67] developed a Python-based custom Blockchain for Blockchain-based Explainable Federated Learning (DP-BFL) to enhance security in IoT-based Social Media 3.0 networks. DP-BFL employs differential privacy to safeguard the exchanged local model updates and the aggregated global model from potential inference or membership attacks. Furthermore, this allows Internet-enabled devices to actively contribute to a globally preserved privacy model by uploading local updates to Blockchain miners. These miners evaluate and reward these contributions, with the added feature of introducing adaptable Gaussian noise to enhance privacy. Miao et al. [104] developed a secure data-sharing model using peer-to-peer FL with Blockchain-distributed ledgers to ensure data transparency and differential privacy for enhanced data privacy in IoT. They employed team-based data sharing with reward and punishment mechanisms to guarantee high-quality and reliable data sharing, where team sponsors initiate tasks and assess members' contributions, rewarding active participants and excluding poorly engaged members. They suggested a proof of model contribution consensus algorithm that relies on the contribution of the training model to enhance computational efficiency. Experimental results confirmed the effectiveness of their approach, highlighting high accuracy and improved privacy in IoT. Zhang et al. [105] present a privacy-protecting FL framework for IoT that employs Blockchain and committee consensus. Local updates are verified through Blockchain, ensuring data privacy with local differential privacy where Laplace noise is used. Committee nodes validate model parameters, and when sufficient validation responses are received, updates are aggregated through a smart contract for the next training round.

In [48] explores a permissioned Blockchain system with the Proof of Training Quality (PoQ) consensus process, optimizing node computing resources during data model training. The Laplace mechanism enhances local data model privacy and improves computing resource utilization and efficiency of the data-sharing scheme. Chen et al. [68] introduced an efficient Privacy-Preserving and Traceable FL framework with minimal overhead and high performance. Their innovative approach incorporates hierarchical aggregate Federated Learning, involving sub-aggregators and aggregators and adding noise to local model parameters using random seeds. The sub-aggregator can reconstruct pseudorandom weights with user IDs or decrypt subtracted parameters. After aggregating and encrypting the parameters, the sub-aggregator forwards them to the aggregator, which decrypts and combines parameters, subtracts user-added noise, and obtains global parameters sent to the server. In [106], PriModChain, a specialized FL architecture for Industrial Internet of Things networks, incorporates a differential privacy approach to add artificial noise to locally generated models, which reduces the risk of the identification of individual records. The secure transfer of the global ML model is facilitated through smart contracts, ensuring consensus on update verification and transparency in FL updates. Simulations in Python evaluate PriModChain's feasibility in terms of security, privacy, safety, reliability, and resilience, highlighting its innovative features in promoting unbiased and error-free data manipulations for enhanced FL safety and reliability against external data threats. The frameworks [68,106] integrate FL with Blockchain and IPFS, guaranteeing the traceability and immutability of model

Table 4
Privacy Preservation in BCFL using differential privacy approaches.

Reference paper	CDP/ LDP	Exponential distribution	Gaussian distribution	Laplace distribution	Random distribution	Parameter
[44]	LDP	✓				Local Gradient
[94,97,102]	LDP		✓			Local Weight
[95,100,101]	LDP			✓		Local model
[24,45,105]	LDP			✓		Local Gradient
[96]	LDP			✓		Local Weight
[48,98,103]	LDP			✓		Local Weight
[67,106]	LDP		✓			Local Gradient
[104]	CDP			✓		Global Gradient
[68]	LDP				✓	Global Gradient

CDP: Central Differential Privacy, LDP: Local Differential Privacy,

parameters, particularly suitable for Industrial Internet of Things scenarios. Table 4 comprehensively outlines the strategies employed for privacy preservation in BCFL by applying the differential privacy approach. The table details the diverse methods and techniques this privacy framework utilizes to ensure robust privacy measures in FL on the Blockchain.

7.3. Privacy preservation in BCFL relying on homomorphic encryption-based approaches

Homomorphic Encryption (HE) is a technique that enables computations on encrypted data, yielding encrypted results without requiring data decryption [107]. In FL, users can employ HE to secure their parameters while sharing them with the server, which protects data privacy and facilitates accurate model aggregation [108,109]. Typically, in FL, the server involves the processing function f, which aggregates parameters from local models across all participating nodes. The encryption computation utilizing HE is detailed in Eq. (3) as follows:

$$E(m_1) * E(m_2) * \cdots * E(m_n) = E(m_1 * m_2 * \cdots * m_n) \tag{3}$$

Where, $(m_1, m_2, m_3, \ldots, m_n)$ denotes the parameters and E represents the encryption algorithm.

Chen et al. [110] developed a data-sharing private model that utilizes BCFL. The study addresses data privacy by proposing a scheme based on FL and employs HE to safeguard user parameters during parameter updates. To alleviate storage issues and manage diverse data formats, the work combines Blockchain storage with off-Blockchain key–value storage, using Blockchain only for data pointers. An innovative on-chain data retrieval mechanism selects data providers for FL. Additionally, the research introduces a consensus mechanism called contribution authorizing Byzantine fault-tolerant algorithm (Con-dBFT), based on contribution, to improve fairness and efficiency in the system. Wang et al. [62] proposed a BCFL to address the security threats faced by the privacy-preserving FL, which enhances Multi-Krum technology by integrating it with HE, resulting in ciphertext-level model aggregation and filtering. This method ensures the verifiability of local models and preserves user privacy. In [111], it also protects the local model's gradients through encryption using the Threshold Paillier encryption algorithm. Furthermore, it introduces a reputation-based incentive mechanism within the Internet of Vehicles to incentivize honest participation in FL, and the authors used a semi-decentralized consortium Blockchain structure with an Elliptic Curve signature and Merkle tree to ensure data security. Sun et al. [112] proposed BCFL, which encrypts the local gradients using the Bresson-Catalano-Pointcheva (BCP) mechanism and then adds homomorphic noise to each encrypted gradient. The modified gradients are then gathered and assessed for quality using a joint audit algorithm. The system identifies any gradients that lead to the global model's degradation, effectively removing them from the model. It then aggregates the remaining gradients, generating a new global model with reduced processing time. However, the behavior and audit chains may become overwhelming as data owners increase, leading to delays and processing times, potentially limiting its practical use in large-scale Federated Learning scenarios. In another work, Miao et al. [113] created a BCFL-based byzantine robust model to ensure privacy and mitigate the system to infer the client's local data. They create a reliable global model by identifying malicious gradients and honest gradient vectors through cosine similarity. Additionally, they used the Cheon-Kim-Kim-Song (CKKS) scheme based on fully homomorphic encryption to safeguard privacy and encrypt local gradients. Furthermore, it significantly decreased the computation and communication overheads. In [54], researchers utilized a similar approach to safeguard the local model from inference attacks. Chen et al. [56] also an effective non-interactive designated decryptor function encryption method as a novel lightweight cryptography tool. The method effectively maintains the accuracy of the global model with comparatively low and efficient transmission costs. Sezer et al. [114] introduced the BCFL framework to guarantee the security and privacy of IoT sensor-based structures utilizing sampled data from electrochemical sensors. Within this architecture, they employed Federated models and cryptographic primitives to ensure user and data privacy in off-chain fog nodes with high accuracy, efficiency, and security.

However, existing HE-based systems face significant challenges, such as the reliance on trusted third parties for key management, increased complexity and vulnerability, and scalability issues with Deep Learning (DL) models due to computational constraints in encrypting and decrypting the trainable parameters [115]. The authors in [116] introduced a BCFL system empowered by edge computing for resource management in the Internet of Medical Things (IoMT). It employs an improved linear regressor model and Paillier encryption for gradient parameter security. Mobile devices act as initiators for model bootstrapping and local task initialization, while validators, selected based on computing capabilities, engage in Blockchain consensus processes, block verification, and validation. The computing threshold for validator miners is determined using maximum likelihood estimation, ensuring a data-driven approach to resource allocation. The resulting blocks are digitally signed, hashed, and encapsulated into the Blockchain, enhancing security features for IoMT and edge computing.

The approach presented by Qi et al. [53] guarantees gradient privacy using HE while tackling trust issues and Single Point of Failure (SPoF) through a reputation system based on smart contracts. Additionally, the model addresses Blockchain storage challenges by implementing an on/off-chain storage strategy. Li et al. [77] proposed a privacy-preserving FL system, employing distributed ElGamal encryption to safeguard gradient inversion attacks. The system recovers original data from local sign-based quantized gradients and utilizes smart contracts for secure self-aggregation among participants without reliance on a centralized server. Some works have focused on privacy in vertical FL, proposing a novel technique that utilizes DL and Blockchain to preserve the privacy of electronic health records by developing a secure logistic regression architecture [117].

In [118] uses a combination of FL, Blockchain, and HE to compute a global behavioral fingerprinting model for a target object in an IoT context. This fingerprint is derived from the interactions of an object with different peers and allows anomaly detection in the network to

Table 5
Privacy Preservation in BCFL using HE.

Reference paper	Encryption type	Privacy scheme	Parameter	Attack against	Adversary
[110]	PHE	Additive	Local Gradient	I	Server
[62]	PHE	Paillier additive	Local Weights	I&P	HbCS &MalC
[112]	FHE	BCP	Local Gradient	I&P	HbCS
[54,113]	FHE	CKKS	Local Gradient	I	Server &MalC
[56]	FE	NDD-FE	Local Weights	I	-
[116]	PFE	Paillier additive	Local Gradient, Global Gradient	Transaction Hacking, I,Impersonation&51%attack	Insider or Outsider
[77]	PHE	Distributed ElGamal	Local Gradient	Gradient Inversion	HbC clients
[120]	Encryption	Proxy re-encryption, ECC, SS,CH	Local Weights	I	HbC clients
[117]	Encryption	Proxy re-encryption	Global Weights	I	-
[49]	PFE	Paillier additive&Proxy re-encryption	Local Gradient	I	SHbCS
[53,111]	PHE	Paillier additive	Local Gradient	I&P	Insider/Outsider
[119]	FHE	Dijk-Gentry-Halevi-Vaikutanathan	Local Model& Global Model	I&P	MalC,HbCS & SHbCS

FHE: Fully Homomorphic Encryption, PHE: Partially Homormophic Encryption, HbCS: Honest-but-Curious Server,MalC: Malicious Client,
I: Inference attack, P: Poisoning attack, SS:Secret Sharing, ECC:Elliptic Curve Cryptography, CH: Chameleon hash, SHbCS: Semi Honest-but-Curious Server.

be performed. The underlying model, thanks to HE, guarantees the privacy of both the target object and the different workers, as well as the robustness of the strategy in the presence of attacks.

Li et al. enforced privacy safeguards in [119] by combining BCFL and HE within a traceable identity-based scheme, ensuring the records' integrity and traceability. They aimed to establish an anonymous identity-based scheme for safeguarding driver identity privacy by adopting FL and utilizing the classic Feige-Fiat-Shamir zero-knowledge-proof authentication.

Table 5 offers a comprehensive summary of privacy preservation within BCFL, utilizing homomorphic encryption with diverse approaches. Awan et al. [49] enhanced the Paillier cryptosystem, incorporating features like additive Homomorphic Encryption and proxy re-encryption to safeguard individual local model updates in FL. Their approach addresses issues such as random client dropouts through asynchronous recording on the Blockchain. Integrating BCFL mitigates multiparty dropout and enhances transparency, verifiability, and data privacy protection.

7.4. Privacy preservation using BCFL with secure multiparty computation approach

Secure Multiparty Computation (SMPC), introduced by Andrew Yao in 1982, forms the foundational protocol for secure computations [121]. It facilitates different parties $(P_1, P_2 \dots P_n)$, with private data $(d_1, d_2 \dots d_n)$, in jointly computing an objective function (f) on their private data $f(P_1, P_2 \dots P_n)$, thus preserving the confidentiality of the input data [107]. The authors in [51] present BCFL with novel committee consensus, utilizing Blockchain for global model storage and local updates. The innovative committee consensus minimizes computation and enhances security. A committee validates updates in each round, reinforcing the global model while rejecting incorrect ones. It allows flexible participation, enabling nodes to join or leave without disruption, and uses Smart Contracts driven by Blockchain transactions to execute the central server functions.

However, some studies emphasize persistent security concerns in key management, particularly regarding secret key ownership in adopted cryptographic systems. To tackle this issue, multiple studies, exemplified by [122,123], advocate for the adoption of the SecAgg protocol [124]. Within this protocol, secret keys are collaboratively shared and securely stored using Blockchain. Fang et al. [123] also address these concerns by employing Blockchain to verify global model gradients, effectively mitigating the potential risk of tampering attacks. Moreover, gradient compression methods are employed to alleviate communication overhead. In [122], a variant of ElGamal encryption was employed to validate the accuracy of aggregated results.

In the architecture proposed by [125], multiple smart hospitals in different regions are assumed, each equipped with a cluster of IoT medical devices and an edge server executing FL tasks. This verification involves encrypted inference through a SMPC protocol. Upon verification, the Blockchain node obtains the authenticated portion of the local model. Utilizing SMPC-based secure aggregation, the Blockchain and the hospital collaborate to reach a consensus on the global model, which is securely stored in the Blockchain. The tamper-proof storage system then disseminates the revised global model to all involved hospitals in the Federated Learning round.

In a Blockchain-based decentralized, secure multiparty Learning system outlined in [126], every client calculates and disseminates its local model via the Blockchain. Following a calibration process specifically designed for edge computing-based IoT applications, clients execute models received from other participants. The system employs a cooperative mining strategy, incorporating on-chain and off-chain mining, to address potential attacks during model broadcasting and calibration.

7.5. Privacy preservation using BCFL with reward-driven approaches

Integrating BCFL with incentive mechanisms not only addresses the challenge of preserving user privacy and encouraging active participation but also ensures the confidentiality and security of the BCFL system. By leveraging smart contracts, BCFL establishes a transparent and tamper-proof framework for fair and verifiable incentives, mitigating concerns about opaque reward structures in traditional BCFL platforms. This innovative integration promotes collaboration and significantly enhances the effectiveness and trustworthiness of the BCFL system [66,119]. BCFL's selection process is guided by a strong emphasis on client reputation. Higher-reputation clients are more likely to contribute reliable and high-quality training. After each training task, client reputations are updated based on their behavior, influencing client selection in subsequent training by considering their reputation records.

Assessing the contributions of diverse data providers is fundamental for fair profit allocation. Implementing reasonable contribution evaluation criteria enhances the incentive mechanism, attracting more participants to join. Clients' contributions can be distilled into two main categories: data quality and data quantity. For example, Salim et al. [67] introduced an incentive mechanism designed to combat free-riding attacks by proportionally rewarding participants based on the quality of their contributions. They implemented the Quality-Based Consensus (QBC) algorithm in DP-based BCFL, ensuring that only legitimate local updates contribute to the global model. QBC rewards participants for added updates, promoting high-quality contributions, and selects

Table 6
Privacy Preservation in BCFL using Reward Driven approaches.

Approach	Reference paper
Client Data Contribution	[24,62,67,111,127,128,130,131]
Auction theory-based schemes	[132–134]
Mechanism design-based schemes	[129]
Contract-theoretic approach	[29]
Game theory-based schemes	[101]
Smart contract-based schemes	[53,54,135]

the consensus leader based on the miner with the highest accuracy for inclusion of the most qualified models in the global update.

Furthermore, Qi et al. [127] proposed a mechanism to motivate data owners to provide high-quality data by establishing a distinct equilibrium by analyzing noncooperative games. A reputation layer utilizing Blockchain for collaborative assessment strengthens the equilibrium, which signifies that contributing the highest quality data leads to the highest reward. In the reward layer, incentives, determined by both the quantity and quality of contributions, are granted using a reputation-weighted algorithm to ensure fair distribution. The unique Nash equilibrium in the non-cooperative data-sharing game shows that data owners act selfishly to maximize their profits.

Additionally, in [111] proposed Deepchain, which also provides reward based on the data quantity. The system involves data owners collaborating to train a model and miners processing transactions for model updates on DeepChain. Data owners pay transaction fees based on their data quantity, with miners competing to process transactions and receive rewards. Value-based incentives promote correct participant behavior. Smart contracts regulate behavior and track attackers. The system assesses global model accuracy using local updates, penalizing invalid transactions and considering updates with decreased accuracy as potentially malicious. In [24,62], a customer-centric incentive system assesses contributions and calculates reputations using Multi-KRUM to eliminate unsatisfactory and malicious updates. In conjunction with this study, Abdel et al. [128] enforced a hybrid incentive strategy, incorporating Multi-KRUM for providing incentives. The authors in [129] introduce a fair and incentive-aware mechanism. Workers actively choose their top k previous models during each round, assigning precisely one vote to each model. The smart contract then calculates aggregated votes, determines worker counts from the preceding round, and allocates rewards in descending order based on these counts.

Rewards for edge nodes, tied to their contributions to the global model, may lack fairness and reasonability. This imbalance arises because edge nodes with substantial datasets and robust computational resources enjoy an unfair advantage, resulting in uneven reward distribution. However, [54] introduced the forward bidding mechanism, which selects the top k edge nodes within the FL task publisher/server budget and compensates them accordingly. To prevent edge nodes from withdrawing during model training, they must submit a fixed security amount, refunded upon successful convergence of the global model along with the reward.

In certain studies, a consensus mechanism has been introduced to fairly reward legitimate users across cross-silos using the model quality. Participants earn a reputation by staking cryptocurrency deposits or their existing reputation in the Proof-of-Federated Deep-Learning (PoFDL) consensus mechanism proposed in [130]. This approach enhances trust among participants and reinforces the immutability of the Blockchain. Participants who take on the role of validator nodes gain reputation through their active involvement in the PoFDL process, establishing a mechanism where contributions to the Federated Learning system increase reputation within the network. Furthermore, Kashyap et al. [131] introduced Proof of Interpretation and Selection (PoIS), a consensus mechanism for participant incentives. PoIS assesses individual contributions using label-wise model interpretation through Shapley value, detecting adversaries through feature attribution aggregation.

The authors in [135] proposed the "Balanced Sign SGD ", a 1-bit gradient compression method that emphasizes privacy by exchanging only the signs of gradients, excluding the gradients themselves. Additionally, it introduces a novel committee-based consensus algorithm featuring a personalized incentive mechanism. It also ensures that every contributing participant is rewarded based on their distinct contributions to enhancing the model. Committee members engage in global aggregation and achieve consensus through cross-validation, with the first finisher receiving additional rewards. Other committee members are rewarded based on their response times, working as evidence of effectiveness. Participants contributing to local models receive rewards based on the cosine distance of their contributions to the global model, with rewards increasing proportionally as the cosine distance approaches predetermined thresholds. In Qi et al. [53], a smart contract-based reputation scheme uses the Reputation Contract (RC) and Hunter Contract (HC) to establish trust. The RC assigns reputation scores, rewarding positive actions and penalizing negatives. Simultaneously, the HC guards against malicious nodes by verifying weights' accuracy and reporting dishonest behavior to the RC, contributing to a trustworthy system.

Some studies incorporate an auction-based mechanism to reward participants efficiently, ensuring a fair and transparent compensation system for their contributions. For example, Batool et al. [132] proposed a multidimensional auction-based reward mechanism that utilizes a smart contract to compensate participating clients with cryptocurrencies. This auction considers factors like computational and network resources and local data quality. The reward distribution is based on the Shapley value, ensuring fairness by measuring the relative contribution of each client. Kang et al. introduced a Subjective Logic approach, as outlined in [133], to assess individual reputations in the context of vehicular networks. This framework for probabilistic information fusion relies on subjective beliefs and operates by evaluating interactions as the basis for reputation assessment. In [29], the study extends [133] by introducing a multi-subjective logic function to enhance the reward approach. The authors also propose a worker selection scheme for dependable Federated Learning, incorporating a multiweight subjective logic model for reputation assessment. Blockchain integration ensures secure decentralized reputation management with nonrepudiation and tamper-resistant properties. Additionally, the incentive mechanism, blending reputation and contract theory, encourages high-reputation mobile devices with quality data to engage in model learning actively. Kang et al. [134] proposed Multiweight subjective logic to enhance reputation calculation in BCFL, considering interaction attributes like frequency, timelines, and effects.

In [101], the study proposes an incentive mechanism for a privacy-preserved data-sharing system, formulating it as a two-stage Stackelberg game. The mechanism is designed to maximize the utility of data requesters and two types of data providers, considering their distinct roles and contributions. The non-cooperative nature of the interactions justifies the choice of a Stackelberg game model, the hierarchical relationship between requesters and providers, and the one-to-many data-sharing structure. Table 6 presents an overview of privacy preservation in BCFL by enforcing a reward-driven approach using various methodologies.

7.6. Privacy protection using BCFL with hybrid privacy approaches

Several studies indicate that integrating diverse privacy approaches helps mitigate security and privacy attacks in BCFL. This section explores hybrid approaches that provide privacy by combining various privacy-preserving techniques. For instance, integrating differential privacy for initial data aggregation and applying homomorphic encryption could yield a more resilient solution. The amalgamation of HE and

Table 7
Privacy Preservation in BCFL using hybrid privacy approaches.

Reference paper	Privacy scheme used	Parameter	Attack against	Adversary
[137]	HE & SMPC	Local Gradient	I	MalC, Malicious Miners
[120]	HE & SMPC	Local models	I	HbC, Clients
[61]	HE & SMPC	Local Models	I&P	HbCS
[136]	HE & SMPC	Local Model	I	MalC
[130]	DP & SMPC	Local Gradient	Byzantine and Sybil attacks, Model inversion, I, Model theft attacks	HbCS, HbCC, MalC
[138]	DP & SMPC	Local Gradient	I	HbC, MalC
[69]	DP & HE	Local weights	Model extraction attack, Model reverse attack	
[139]	DP& HE & SMPC	Local Gradient	Collusion attack, Sybil attack, I, & P	HbCC, MalC
[140]	SS, Combine Paillier and ElGamal based scheme	Local Gradient	I	Internal or External Adversary

HbCS: Honest-but-Curious Server, MalC: Malicious Client, I: Inference attack, P: Poisoning attack, HbCC: Honest-but-Curious Client, SS: Secret Sharing.

SMPC in BCFL markedly enhances the confidentiality and privacy of the FL process within a transparent and decentralized Blockchain framework. This integration fosters trust and security in data sharing and model training, as exemplified by [61,136]. HE enables computations on encrypted data, preserving the privacy of individual contributions, while SMPC ensures secure collaboration among participants without exposing their raw data. Table 7 summarizes privacy preservation in BCFL by employing various privacy approaches. In the privacy-focused collaborative training proposed by Zhu et al. [137], participants protect their local gradients using the Paillier cryptosystem with threshold decryption and a secure multi-party aggregation algorithm. This method ensures data privacy during collaborative training by transforming gradients into a secure form.

Furthermore, in [120] introduced a flexible and trustworthy framework for industrial intelligence, integrating autonomous FL and secure data-sharing on the Blockchain. The proposed approach preserves privacy through a combination of HE and SMPC approaches, which can enhance the security of sensitive data. Their approach involves an autonomous Federated extreme gradient boosting Learning algorithm for privacy protection, verifiability of aggregated results, and model reliability. They also introduced a secure and trusted trading mechanism for controlled on-demand data sharing, a threshold aggregation signature for model ownership assurance, and proxy re-encryption and retrieval to facilitate controllable and reliable data sharing with high accuracy and performance. Feng et al. [61] presents a framework for decentralized cross-domain FL in 5G-enabled UAVs, leveraging Blockchain technology. It utilizes multi-signature smart contracts for dynamic cross-domain authentication, enhancing collaborative Learning. The framework employs decentralized smart contracts for model aggregation, addressing security concerns related to centralized servers. Additional security measures, such as homomorphic encryption and multiparty computation, are applied to protect against local update attacks.

FL presents a promising avenue for developing energy-efficient consensus algorithms, addressing the resource-intensive nature of traditional methods like PoW. Integrating the consensus process with FL eliminates the need for extra computational resources dedicated to separate consensus algorithms, potentially leading to substantial energy savings. From a communication standpoint, public Blockchains often require miners to broadcast their local model parameters, resulting in considerable communication overhead, especially as the number of miners grows. The authors in [136] proposed a method to mitigate these challenges using a novel consensus protocol like Proof-of-Federated-Learning (PoFL), leveraging the computational overhead of local training in Federated Learning as proof for consensus. PoFL significantly reduces mining power wastage and trims computational overhead while ensuring efficient consensus processes without reference to external sources. Moreover, it proposed a novel method utilizing a reverse game-based data trading mechanism to enhance data privacy by determining optimal data trading probabilities and pricing strategies. This approach encourages data pools with high privacy risks to trade less data at a higher cost, incentivizing them to train models without data leakage. Additionally, a privacy-preserving model verification mechanism consists of HE-based label prediction and SMPC with two-party-based label comparison, ensuring model accuracy while preserving privacy for both the task requester's test data and the pool's submitted model.

In [130], they explored the integration of secure multi-party computation and differential privacy to enhance system privacy. Also, a permissioned Blockchain and private peer-to-peer channels are utilized in their approach. Encourage cross-silo FL using the lightweight and energy-efficient consensus Proof-of-Federated Deep-Learning protocol, effectively detecting and classifying IIoT attacks in Non-IID and IID scenarios. Bai et al. [140] proposed a Blockchain-based privacy-preserving approach using no trusted third-party Federated Learning. They employ a conference key agreement to negotiate keys between the initiator and partners, eliminating the need for a trusted third party. A double-layer encryption mechanism ensures privacy encrypts local and global models, preventing partners from accessing each other's private information. The decentralized nature of Blockchain enhances transparency, traceability, and resilience against SPoF. Additionally, they used an efficient secret-sharing scheme to encrypt model parameters, reducing communication costs and computation time compared to Paillier and ElGamal-based schemes and secure aggregation protocols.

Bolstering security against Sybil attacks, poisoning attacks, and inference attacks, Shayan et al. [138] incorporate differential privacy and encryption approach within BCFL with secure and private multi-party ML. In each iteration, peers compute local model updates, keeping them private by masking with differentially private noise obtained from a set of peers identified through a verifiable random function. Verification committees validate these masked updates to prevent poisoning. If the majority of the committee approves, the updates are divided into Shamir's secret shares and passed to an aggregation committee. This committee securely aggregates the unmasked updates, with contributing peers and committee members receiving additional stake in the system. The aggregated updates are then added to the global model within a newly created Blockchain block and shared with all peers, and the process repeats with the updated global model and stake.

The studies outlined in [69] focus on establishing a secure data-sharing mechanism to uphold privacy among numerous distributed users. It also suggests a data protection aggregation approach that utilizes distributed K-means clustering with DP and HE, random forest with DP, and AdaBoost with HE to enhance data protection in Industrial IoT scenarios. Sun et al. [139] address the challenge of enhancing security and privacy in their work. They use a Blockchain to record each global model update, ensuring the verifiability and traceability of local

Table 8
Summary of studies on integration of Blockchain enabled Federated Learning, elucidating their privacy preservation methods, types of Blockchain used, the Blockchain frameworks integrated within Federated Learning systems, consensus algorithms employed, block storage, and data distribution used.

	Techniques	Reference paper
Privacy Approach	Differential privacy	[24,44,45,48,67,79,84,94–103,105,106,141–143]
	Homomorphic encryption	[53,56,62,77,83,110–112,114,116,117,119,136,144,145]
	Secure multi-party computation	[49,51,122,123,125,126]
	Reward driven approaches	[29,45,62,63,67,80,88,90,95,101,102,104,111,127,128,130,132–134]
	Hybrid privacy approaches	[61,68,69,120,130,136–140]
Consensus Protocol	PoW	[65,67,79,80,82,88,95,96,98,102,116,119,125,126,143,145,146]
	PoS	[56]
	DPoS	[44,63]
	pBFT	[29,46,53,61,62,90,91,94,101,134,139]
	PoA	[55,84,100,143,145,147]
	PoQ	[48]
	PoF	[84,97,136,138]
	PoC	[114]
	PoFL	[136]
	RAFT	[61,69]
	Con-dBFT	[110]
	Algorand	[24,111,128]
	PoV	[142]
Blockchain Type	Public	[44,46,56,69,82–84,96–98,100,106,117,128,132,143,144,146,148,149]
	Private	[113,125,145]
	Permissioned	[48,63,77,82,85,89,120,130,142]
	Consortium	[24,29,53,55,61,62,79,87,90,94,95,97,101,110,112,116,127,133,134,140,141,147,150–153]
Blockchain Platform	Ethereum	[49,54,77,83,100,104,106,113,116,120,122,128,132,137]
	Hyperledger Fabric	[61,62,79,85,87,90,112,127,139]
	Custom Blockchain	[67]
Blockchain Storage	off-chain	[24,46,49,53–55,61,68,79,83,84,99–101,103,106,110,114,132,138,147,151]
	on-chain	[44,55,62,69,77,87,88,90,96,98,101,105,112,114,116,125,130,141,143–146,152,154]
Data Distribution	IID	[24,44–46,48,56,63,65,67,68,79,83,84,87,88,91,94,96,98,101,103,106,110–113,125,127,128,130,136,138,142,150,155,156]
	Non-IID	[55,62,82,85,90,91,97,102,104,127,128,130,131,137,146,148,152]

updates through permanent records. It also enables an incentive mechanism tailored to user contributions. Additionally, HE secures users' local model updates. A validation process precedes local update aggregation to thwart poisoning attacks, and privacy is maintained with differential privacy noise. Ultimately, they establish a secure aggregation scheme for local updates using the Shamir secret sharing technique, balancing utility and privacy compared to differential privacy.

Table 8 elucidates the overview of studies specifically in BCFL, highlighting their privacy approach, Blockchain types, the Blockchain frameworks utilized within Federated Learning systems, consensus algorithms, and block storage techniques. It is possible to see that most of the existing literature preferred Consortium Blockchain, primarily utilizing the Ethereum platform, PoW consensus, and DP for privacy. Furthermore, IID data distribution and on-chain storage are commonly chosen by researchers. However, as we discuss in Section 9, such choices are mainly related to the specific use case scenario considered in the solution.

8. Privacy preservation using cross-chained FL approaches

In this section, we have delved into the intricacies of cross-chain-enabled Federated Learning as a mechanism for preserving privacy. The discussion thoroughly explores how leveraging cross-chain capabilities enhances Federated Learning methodologies to uphold and safeguard privacy. A brief introduction to Cross-Chain technology has been provided in Section 4.3.4.

8.1. Overview of cross-chained FL

Recent studies indicate that BCFL systems preserve the system's privacy. Still, the limited scalability of a single Blockchain becomes evident as the number of FL training tasks increases, resulting in the simultaneous generation of numerous blocks and subsequent queuing for block verification. This scalability challenge emerges due to the difficulty of managing massive block data with a limited number of miners, leading to constrained throughput, reduced efficiency, and slower FL training processes [157]. Additionally, BCFL incurs a substantial communication cost for model update transmission, requiring multiple rounds of communication to achieve the desired accuracy level. This arises from frequent gradient exchanges among peers over limited bandwidth channels, and as the block data size increases, so does the flow of model updates across the Blockchain network, posing significant communication challenges. Moreover, Blockchain-enabled Federated Learning encounters numerous challenges, including selecting efficient miners, consensus algorithm implementation, and chain validation [51,158,159]. Cross-chain technology enables data exchange among multiple Blockchains. Which also facilitates secure data transfers while maintaining the same machine-learning models throughout various Blockchain networks [1,160]. The following highlights the major benefits and key advantages of cross-chained enabled FL [161].

- **Higher Scalability:** Cross-chained FL outperforms single Blockchain in efficiency and scalability. Unlike single Blockchain limitations in managing FL training tasks, cross-chained systems efficiently distribute workloads, mitigating bottlenecks. The parallel processing capability ensures optimal scalability, seamlessly accommodating growing FL task demands. Multiple interconnected Blockchains enhance resource management, improving system efficiency and security compared to a singular Blockchain system. The cross-chain integration in FL enables global collaboration, fostering diverse participation and data federation across regions and industries.

- **Low Communication Cost:** Blockchain-based FL requires frequent gradient exchanges to synchronize model updates among peers, utilizing limited bandwidth channels. Cross-chained FL networks employ a compressed gradient strategy, ensuring cost-effectiveness and high accuracy. Due to the compression of gradients, this scheme fortifies the safeguarding of training data privacy by reducing the efficacy of gradient leakage attacks when there is an inadequate amount of gradient information [73].
- **Reduced Single-Point-of-Failure Risks:** Cross-chain Federated Learning mitigates the risks associated with a single-point-of-failure. Distributing the learning process across multiple Blockchain's makes the system more resilient to potential disruptions or attacks on a single chain.

8.2. Solutions for privacy preservation using cross-chain approaches

In this section, we explored the intricacies of the cross-chained network, which has diverse privacy solutions meticulously crafted to safeguard the system's privacy by integrating cross-chain-enabled Federated Learning. Kang et al. [161] introduced an innovative cross-chain powered FL framework with parallel Blockchains designed to handle model updates securely, with scalability and flexibility, eliminating the constraints of conventional single BCFL systems. Their approach incorporated a two-phase commit protocol to validate and authenticate block data across multiple Blockchains for Artificial Intelligence of Things in 6G. Furthermore, they utilized a mixed-precision local training strategy combined with flexible model update compression to improve communication efficiency without compromising accuracy. In the "Prepare"phase, the system establishes the groundwork by deploying model training and payment smart contracts on the source and destination parachains. The task publisher calls the training smart contract, sends a cross-chain request, and collaborates with validators and collators across parachains for legitimacy. Simultaneously, the payment smart contract activates to secure assets for worker rewards upon model training completion. Transitioning to the "Commit"phase, the trained model undergoes quality evaluation, triggering the training smart contract to generate a Simplified Payment Verification (SPV) proof and block header. Verified by the relay chain's validator group, they reach a consensus on the model training's legitimacy. Successful validation leads to worker compensation, with payment records logged in the payment chain. Discrepancies prompt a rollback, releasing locked assets. This two-phase process ensures the secure execution of cross-chain-enabled Federated Learning, managing complexities across interconnected Blockchains.

The prevailing BCFL system encounters data sparsity issues despite its commendable system efficiency. To tackle these concerns, Jin et al. [162] introduced a cross-cluster Blockchain-enabled FL framework employing a cross-chain approach for the Internet of Medical Things. Their proposal includes the integration of two Blockchain consensus algorithms to facilitate secure model exchange across clusters using PBFT and a two-phase cross-chain consensus mechanism. Additionally, they advocate for model aggregation within each BCFL cluster and subsequent transmission to the other cluster, resulting in a remarkable enhancement of system efficiency and accuracy, with performance increased from 39.3% to 75.8%. This places a significant burden on computational and communication resources, so researchers suggested using it with edge computing instead of end devices. Kang et al. [163] introduced a privacy framework that employed a hierarchical cross-chain structure for healthcare metaverses. The proposed system empowers users to safeguard sensitive data in the physical space and contribute non-sensitive data for metaverse tasks. Also, a data freshness-based incentive mechanism inspired by prospect theory [164] is used for user-centric data sharing, and a Paillier homomorphic encryption algorithm is used to provide security and privacy. Their approach achieved 93.71% accuracy in breast cancer prediction via vertical FL training.

Xu et al. [165] present a hierarchical micro chained fabric, denoted as μDFL, designed for decentralized, Federated Learning across devices in edge networks. The microchain consensus protocol, built upon a partially decentralized Blockchain utilizing Proof-of-Credit (PoC), ensures the transparency and privacy of data sharing during local model training. The proposed μDFL introduces a hierarchical Internet of Things network fabric, incorporating lightweight microchains. Each microchain adopts a hybrid approach involving PoC block generation and a Voting-based Chain Finality consensus to enhance efficiency and privacy. The Federated structure of μDFL is achieved through an inter-chain network employing Byzantine Fault Tolerance. Validation through a proof-of-concept prototype demonstrates the effectiveness of μDFL in cross-device Federated Learning environments, emphasizing efficiency, security, and privacy.

9. Application of blockchain-enabled FL for privacy preservation

In the following sections, we describe how several approaches leveraging FL and Blockchain for privacy preservation are used in different application scenarios, such as Healthcare, Industrial IoT (IIoT), and the Internet of Vehicles.

9.1. Healthcare

The analysis of health data using ML techniques can result in therapies and procedures with lower risks and better outcomes for patients, thus increasing the quality of care [166]. Healthcare data are usually spread across various sources such as hospitals, clinics, and wearable devices, which are characterized by highly sensitive information demanding to keep patients' data as private as possible. The decentralized nature of Blockchain technology and the ability of FL solutions to train models locally, while sharing only model parameters, has made the combination of the two approaches well-suited for healthcare. Moreover, Blockchain-based FL not only overcomes challenges associated with the outflow of confidential medical data efficiently, but can *(i)* reward FL members for their contribution to the network *(ii)* monitor that the centralized FL server accurately aggregates the global model.

In this context, multiple IoT devices, including weight meters, blood pressure, glucose meters, insulin pumps, and others are connected to patients and aim at acquiring specific data they are meant to be gathered from the human body, such as temperature, heartbeat, electrocardiograph, and many others. These devices communicate data to smart systems such as smart monitors, laptops, and mobiles to be analyzed and visualized. The main goals of the different Blockchain-based FL approaches for healthcare can be summarized as follows [167]:

- Management of medical records, also thanks to the cooperation between multiple hospitals/systems;
- Tracking disease outbreak;
- enhanced monitoring of patients thanks to a wider amount of data to be analyzed;
- Improving sensors' performance;
- Pharmaceutical clinical trials.

In the system presented in [80], IoT devices, before communicating with third-party components, send data to a Blockchain network for validation and, after this step, data is forwarded to other systems. Moreover, Blockchain provides large independent storage for healthcare data, recording usage behavior and ensuring authenticity. Multiple actors collaborate to provide a privacy-preserving solution, namely: *(i)* the sub-feature manager, which vertically partitions the aggregated data into different datasets; *(ii)* the different clients, which provide data to the federation manager and receives a sub-model for the training; *(iii)* the privacy broker, in charge of solving privacy issues; *(iv)* the integrity manager, which maintains the result integrity by avoiding errors inside sub-models.

The authors in [125] proposed Blockchain architecture assumes the presence of numerous smart hospitals situated in diverse regions, each equipped with a cluster of IoT medical devices. These devices use edge servers for FL tasks with privacy-preserving verification via SMPC before aggregation. After verification, the local model is sent to the Blockchain for SMPC-based secure aggregation. Once a consensus is reached, the global model is stored in the Blockchain, and tamper-proof storage shares it with all FL round hospitals.

Also, the works presented in [150,155] are related to the Internet of Medical Things (IoMT) devices. In particular, [155] proposes a Blockchain-enabled Federated Learning in the context of IoMT, with privacy-preservation and fraud detection characteristics. The solution is intended for healthcare applications in a fog-cloud-assisted network. The authors of [150] introduce a real-time medical data processing multi-agent system that utilizes Blockchain for sharing and safeguarding private data.

Similarly to the previous approach, the architecture presented in [82] considers multiple hospitals leveraging FL to keep their data private, thus sharing only weights and gradients. In this case, the aim is to recognize the presence of COVID-19 infection from lung Computed Tomography (CT, hereafter) scans. Each hospitals use Blockchain technology to distribute data, with each hospital storing an actual CT scan, and Blockchain facilitating the retrieval of the trained model. Privacy is ensured through encryption and the storage of unique identifiers for each hospital.

In the realm of COVID-19 diagnosis, [141] introduces a Federated Blockchain-powered medical system, termed FedMedChain. The primary objective of this system is to distribute COVID-19 information and establish a collaborative diagnosis model while safeguarding the privacy of data owners.

In the paper presented in [151], the authors propose a framework called Blockchain Vertical Federated Learning E-Medical Recommendation (BVFLEMR). It adopts a decentralized digital ledger system for Electronic Health Records (EHR) storage, LightGBM, and N-Gram models to recommend tailored treatments for the patients based on their EHR. In this way, it achieves private storage and management of patients' sensitive data in EHRs, such as diagnosis, treatment, medication, surgery, and diet specifications.

Privacy of EHRs is taken into account also by the authors of [117], who propose a framework called CNN_BC_Cryp_FL. It consists of *(i)*, a CNN-based secure classification component able to classify normal and abnormal users using the available dataset; *(ii)* a Blockchain-integrated cryptography-based FL used to restrict the accessibility of the database to abnormal users.

Several studies leverage Blockchain to incentivize participants to contribute their local data in training FL tasks [55,56,152]. In particular, the work presented in [56] describes the system called ESB-FL that can train a model while protecting the privacy of local training data using a function encryption scheme called non-interactive designated decryptor function encryption (NDD-FE). It also integrates Blockchain to support the fair payment between the task publisher and all participants, thus guaranteeing that each participant gets a reward if the trained model satisfies the task requirements. Instead, the authors of [55] aim to improve the fairness of the federated learned model and the trustworthiness of medical diagnostic image analyses to detect COVID-19. Table 9 summarizes the main aim of the papers analyzed in this section.

The diagram in Fig. 11 illustrates the most used configuration of the different factors allowing a privacy-preserving BCFL in Healthcare applications. The analyzed features are the Blockchain type (e.g., private, public, or consortium), the employed consensus protocol (e.g., PoW, PoA, or PoS), the used privacy technique, the Blockchain storage type (on-chain or off-chain), the information related to the data distribution in FL (if it is IID or non-IID), and if the proposal leverages smart contracts. In the diagram, we report the values of a property if it is different from zero.

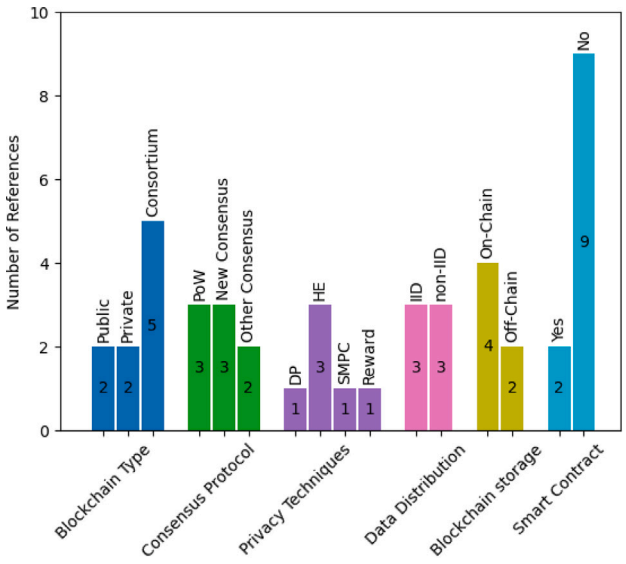


Fig. 11. BCFL configuration for Healthcare Applications.

By observing this diagram, we can state that most of the works dealing with Healthcare approaches choose a Consortium Blockchain, which is a semi-decentralized solution where multiple organizations or entities work together to maintain the ledger. As for the consensus protocol, both the classical PoW and custom solutions have been adopted. Whereas, different privacy techniques and FL data distribution have been chosen depending on the approach. The most adopted storage is the on-chain storage, which consolidates all records within a single ledger. Finally, only a few works leverage smart contracts.

9.2. Industry 5.0

Industry 5.0 is a new concept that focuses on the cooperation between humans and machines to create sustainable industrial products and services. The main principles inspiring this innovative scenario (namely, sustainability, human-centeredness, and resilience) are obtained thanks to the integration of digital technologies, the Industrial Internet of Things (IIoT, hereafter), artificial intelligence, and other advanced technologies into the manufacturing and industrial processes [168]. In this context, the combination of FL and Blockchain could provide powerful solutions for industries seeking to leverage data for innovation while ensuring privacy, security, and efficiency.

In [130], the authors present a framework called PPSS to protect privacy and defend against cyber attacks in the context of industry 4.0/5.0. PPSS includes two modules: *(i)* a Blockchain-enabled FL scheme, leveraging a differentially private training strategy, with an energy-efficient consensus protocol, named Proof-of-Federated Deep-Learning (PoFDL), and *(ii)* a privacy-preserving intrusion detection scheme using Convolutional Neural Networks for attack identification.

Similarly, [148] proposes a Federated threat-hunting model in IIoT networks to identify anomalous behavior while preserving the privacy of IIoT devices related to Blockchain-based smart factories.

The goal of the works proposed in [48,68,69] is to design a secure data-sharing mechanism that can share data among multiple distributed users while maintaining data privacy. In particular, the paper presented in [48] integrates FL in the consensus process of a permissioned Blockchain, so that the computing work for consensus can also be used for Federated training tasks. Whereas, [68] uses FL to obtain privacy-preserving model training, the InterPlanetary File System (IPFS) distributed storage system for storing model parameters and generating corresponding addresses based on the content, and Blockchain to provide the provenance and immutability of the parameters. Instead, the

Table 9
Healthcare applications.

Reference paper	Aim	Type of devices generating data	Type of data
Singh et al. [80]	Private storage, Health alert	IoT sensors (weight meters, blood pressure, glucose meter, insulin pump)	Temperature, heartbeat, blood pressure, electrocardiograph
Kalapaaking et al. [125], Polap et al. [150]	Privacy-preserving analysis from multiple hospitals	Internet of Medical Things devices	Medical datasets
Lakhan et al. [155]	Fraud analysis, and Data validation	Internet of Medical Things (IoMT) devices	Medical datasets
Kumar et al. [82]	Diagnosis of COVID-19	CT device	Lung Computed Tomography scans
Samuel et al. [141]	Privacy-preserving Diagnosis and Dissemination of COVID-19	Internet of Medical Things (IoMT) devices	Medical datasets
Hai et al. [151]	EHR Private storage, Recommendation for tailored treatment	Manual data insertion	Electronic Health Records (EHR)
Alzubi er al. [117]	Abnormal users identification and Database access	Manual data insertion	Electronic Health Records (EHR)
Chen et al. [56]	Privacy-preserving image detection, and incentive mechanism	Manual data insertion	Chest X-ray images
Liu et al. [152]	Privacy-preserving image detection, and incentive mechanism	Manual data insertion	Skin Cancer images
Lo et al. [55]	Diagnosis of COVID-19, and incentive mechanism	Manual data insertion	X-rays images

work of [69] proposes a data protection aggregation scheme based on three ML methods (i.e., distributed K-means clustering based on differential privacy and homomorphic encryption, distributed random forest with differential privacy, and distributed AdaBoost with homomorphic encryption) to enable multiple data protection in IIoT scenarios.

Always in the context of secure data sharing, the paper described in [144] tackles the problem of privacy-preserved credit data sharing. The combined credit data storage mechanism with a Deletable Bloom filter (DLBF) guarantees the traceability of the entire credit data sharing process in industrial applications. Moreover, they leverage homomorphic encryption, FL, and Blockchain to avoid data leakage. The paper presented in [146] has a different goal. It focuses on preserving the privacy of the client data (e.g., usage frequency and time), adopting FL to train the models locally on the client to detect possible device failures in the network. Moreover, to resolve disputes between the central organization and client organizations about failure causes, the architecture leverages a combination of Blockchain and Merkle-tree to enable verifiable integrity of client data.

The authors of [147] develop a framework for FL tasks to preserve privacy among various industrial departments. Decentralized secure storage is provided by the Distributed Hash Table (DHT) at the cloud layer of the proposed scheme, while the Blockchain network provides data authentication and validation.

Industry 5.0 is also expected to reshape the agriculture industry, as already done in the past, and promote the fourth agricultural revolution [169]. In this context, the authors of [153] propose an intrusion detection system called FELIDS for securing agricultural IoT infrastructures. It aims to protect data privacy through FL, employing three deep Learning classifiers, namely, Deep Neural Networks, Convolutional Neural Networks, and Recurrent Neural Networks against Agricultural IoT attacks. Moreover, Blockchain helps network members to track relevant information for supply chain management. Table 10 summarizes the main aim of the papers analyzed in this section.

The diagram reported in Fig. 12 illustrates the most used configuration of the different factors allowing BCFL to preserve privacy in Industry 5.0 application scenarios. The analyzed features are always the Blockchain type (e.g., private, public, or consortium), the employed consensus protocol (e.g., PoW, PoA, or PoS), the privacy technique used, the Blockchain storage type (on-chain or off-chain), the information related to the data distribution in FL (if it is IID or non-IID), and if the proposal leverages smart contracts. In the diagram, we report the values of a property if it is different from zero.

Interestingly, from the observation of this figure, we can state that most of the works referring to Industry 5.0 have chosen a Public Blockchain instead of a Consortium (as done for the healthcare scenarios), maybe because of the different nature of data exchanged. As for the consensus protocol and the FL data distribution, distinct solutions have been adopted depending on the approach. Whereas, the most used privacy techniques in this context are Homomorphic Encryption and Differential Privacy. Also, in this case, *(i)* the most adopted storage solution is the on-chain one, which consolidates all records within a single ledger, and *(ii)* only a few works leverage smart contracts.

Table 10
Industry 5.0 applications.

Reference paper	Main aim
Hamouda et al. [130]	Privacy-preserving FL, and Intrusion Detection
Yazdinejad et al. [148], Friha et al. [153]	Privacy-preserving Anomalies Detection
Lu et al. [48], Chen et al. [68], Jia et al. [69]	Privacy-preserving Data Sharing
Yang et al. [144]	Privacy-preserving Credit Data Sharing
Zang et al. [146]	Privacy-preserving Device Failure Detection
Singh et al. [147]	Privacy-preserving FL

9.3. Internet of Vehicles

The Internet of Vehicles (IoV, hereafter) defines the evolution of conventional Vehicle Ad-hoc Networks. It enables real-time information exchange among all the actors traveling through streets (e.g., vehicles, drivers, pedestrians) and road infrastructure through vehicle-to-everything (V2X) communication. The objective of IoV is to realize the convergence of mobile communication technology, intelligent transportation, and information systems [170,171]. Because this scenario allows for quick and efficient exchange of large amounts of data containing private information (i.e., location and user preferences), approaches that rely on the combination of FL and Blockchain have been investigated. The arisen challenges are the following:

- Strengthening privacy protection mechanisms;
- Preventing hostile intelligent connected vehicles (ICVs) and edge servers from faking FL aggregate results with verification mechanisms;
- Reducing the high communication overhead of FL.

The papers proposed in [44,62,142] provide approaches for data sharing among vehicles for collaborative analysis to enhance service quality and driving experience. In particular, in [142], a Blockchain-enabled and privacy-preserving FL framework called BV-ICVs is presented. In this system, smart contracts are used to prevent malicious

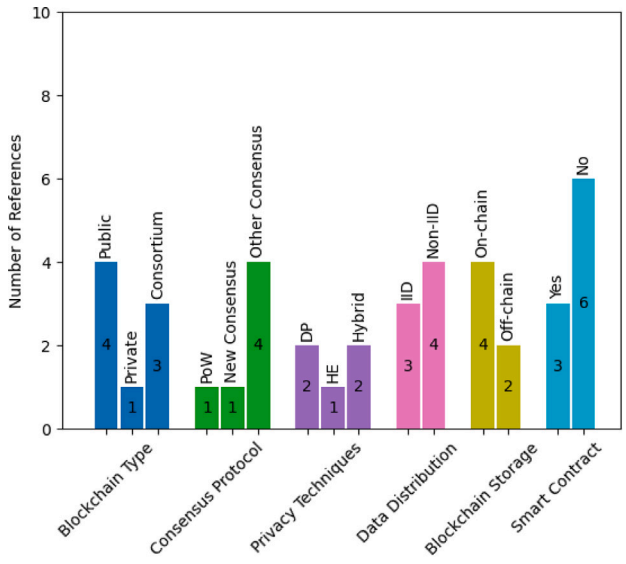


Fig. 12. BCFL configuration for Industry 5.0 Applications.

ICVs from uploading unreliable, erroneous, or low-quality FL model updates. The authors of [44] use FL to relieve transmission load and address privacy concerns of providers and a hybrid Blockchain architecture, which consists of a permissioned Blockchain and a local Directed Acyclic Graph (DAG), executing a two-stage verification to obtain the reliability of shared data. The scheme illustrated in [62], in addition to a Blockchain-based Privacy-preserving Federated Learning approach, also proposes a reputation-based model as an incentive mechanism to encourage users of IoV to participate in FL tasks actively.

The work presented in [94] describes a framework for traffic flow prediction. To avoid using a centralized model coordinator, a consortium Blockchain-based FL framework is proposed to enable decentralized and secure FL. The model updates from distributed vehicles are verified by miners and stored on the Blockchain. Moreover, to preserve model privacy on the Blockchain, a differential privacy method with a noise-adding mechanism is used. Likewise, the system proposed in [65] aims to remove the FL centralized global server and use a Blockchain to exchange local model updates from vehicles while providing and verifying their corresponding rewards.

The authors of [149] focus on an approach to provide a hierarchical Blockchain-enabled FL algorithm for knowledge sharing in IoV. Moreover, they formulate a lightweight Proof-of-Knowledge (PoK) consensus mechanism to reduce the computation consumption. The works presented in [143,156] have the different goal of designing a cooperative intrusion detection mechanism that offloads the training model to IoV devices. [143] distributes the FL computation to reduce the resource utilization of a central server while assuring security and privacy, and it relies on Blockchain to ensure the security of the aggregation model, store, and share the training models. Instead, [156] uses Blockchain to store and share models from the previous steps in a smart contract and return the updated models to the vehicles.

An IoV-related application scenario is that of Drone Edge Intelligence, which refers to the ability of unmanned aerial vehicles (UAVs), or drones, to process and analyze data directly at the source or edge rather than relying on a centralized computing system. Drones' characteristics, such as line of sight, ease of deployment, and capture of high-resolution images, make them the efficient solution for disaster mitigation, security surveillance, environmental monitoring, and recovery [172]. FL allows drones to execute decentralized collaborative learning by computing local models. Only model parameters

Table 11
Internet of Vehicles applications.

Reference paper	Main aim
Smahi et al.[142], Lu et al. [44], Wang et al. [62]	Privacy-preserving and verifiable FL
Qi et al. [94]	Privacy-preserving traffic flow prediction
Pokhrel et al. [65]	Privacy-preserving distributed FL
Chai et al. [149]	Privacy-preserving knowledge sharing
Liu et al. [143], Moulahi et al.[156]	Privacy-preserving and cooperative intrusion detection
Akram et al. [145]	Privacy-preserving malicious drone detection

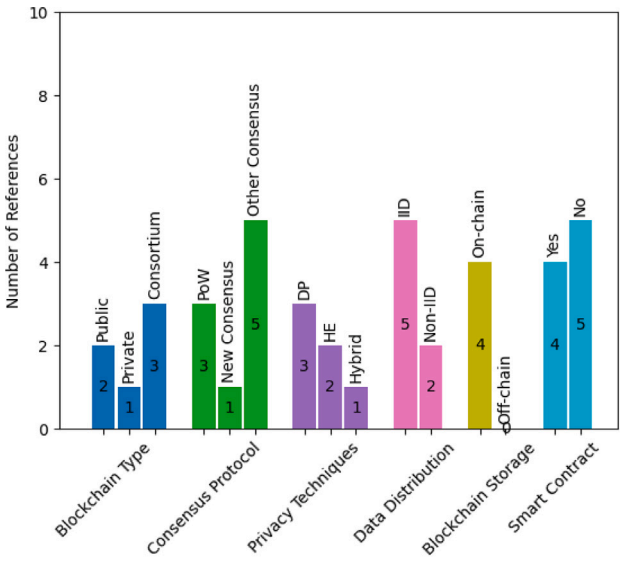


Fig. 13. BCFL configuration for Internet of Vehicles Applications.

are shared with neighbors and the centralized unit to improve global model accuracy while still keeping local data private. On the other hand, Blockchain can enable privacy-preserving data sharing in a distributed manner. However, combining the two solutions raises several challenges, such as scalability, energy efficiency, and transaction capacity [154]. The paper presented in [145] relies on Blockchain technology and FL for privacy-preserving malicious node detection in the Internet of Drone Things (IoDTs).

Table 11 summarizes the main aim of the paper analyzed in this section.

Fig. 13, instead, illustrates the most used configuration of the different factors allowing BCFL to preserve privacy in the Internet of Vehicles application scenarios. As before, the considered features are the Blockchain type (e.g., private, public, hybrid, or consortium), the employed consensus protocol (e.g., PoW, PoA, or PoS), the privacy technique used, the Blockchain storage type (on-chain or off-chain), the information related to the data distribution in FL (if it is IID or non-IID), and if the proposal leverages smart contracts. From this figure, we can affirm that, for this context, the typology of Blockchain and the chosen consensus protocol depend on the approach. The most used privacy technique in this case is Differential Privacy, and the most adopted storage solution is the on-chain one, which includes all records within a single ledger. Finally, due to the peculiarity of this application scenario, which usually involves homogeneous vehicles, FL data are assumed independently and identically distributed (IID), and several works leverage smart contracts.

10. Lesson learned

This section thoroughly explores our insights on the analysis carried out in this paper. The extracted key findings from our survey are summarized as follows:

- Public Blockchains are known for their scalability with a growing number of participant nodes, but they face privacy challenges. In contrast, private Blockchain ensures node privacy but struggles with scalability as the number of nodes increases. Smart contracts have recently been utilized to authenticate and regulate node access to public Blockchain, enhancing their privacy. As for our considered BCFL domain, most of the analyzed approaches make use of consortium-based networks to balance scalability and privacy.
- Traditional Blockchain consensus algorithms are often computationally intensive and resource-consuming, hampering overall efficiency. Committee-based algorithms may not be suitable for resource-constrained applications such as IoT. Innovative lightweight committee consensus algorithms in BCFL prioritize reliability while minimizing communication overhead, thus being useful across diverse constrained use cases.
- Off-chain solutions are utilized by researchers to address space-related constraints and store the local and global models outside the Blockchain. Off-chain storage employs a decentralized storage solution and offers enhanced traceability, as well as a fast and secure retrieval mechanism with low bandwidth. The approach mitigates the risk associated with centralized servers.
- Numerous studies focused on privacy preservation schemes using Differential Privacy in BCFL across diverse applications. Most of these studies inject Laplace noise into the participants' local model to protect against inference attacks. Differential Privacy is one of the lightweight solutions available in the literature and is used in constrained environments.
- Homomorphic encryption-based BCFL enhances the system's privacy by utilizing encryption methods such as PHE and FHE, countering different inference and poisoning attacks. The majority of the analyzed works used this approach to protect local models.
- BCFL with incentive mechanisms ensures privacy, confidentiality, and security of the BCFL system in different application domains. In BCFL, rewards are determined through various methods such as Client Data Contribution, Auction theory-based schemes, Contract-theoretic approaches, Game theory-based schemes, and Smart contract-based mechanisms.
- Numerous studies highlight the effectiveness of hybrid approaches, which combine multiple privacy-preserving techniques to mitigate security and privacy attacks in BCFL. Integrating HE and SMPC substantially enhances confidentiality and privacy within a transparent, decentralized framework. Furthermore, incorporating differential privacy for initial data aggregation and applying homomorphic encryption could yield an even more resilient solution.
- Many studies focus on preserving privacy in Blockchain-based Federated Learning by applying distinct privacy approaches to the local model (either local weight or local gradient) rather than global models.
- Despite the security enhancements and data-sharing capabilities offered by Blockchain for FL, BCFL remains susceptible to various attacks. Blockchain's decentralized and immutable nature can bolster security, but it remains open to attacks, including 51% attacks and those targeting smart contracts or consensus mechanisms.
- Smart contracts play a key role in BCFL, validating client contributions, computing global models, recording participant contributions, and incentivizing clients based on model effectiveness. Despite their importance, smart contracts are vulnerable to security threats from bugs in their code, necessitating attention to programming language issues and robust security measures.
- Recently, researchers introduced an innovative cross-chain powered FL framework with parallel Blockchains. The Blockchains are designed to handle model updates securely, with scalability and flexibility, eliminating constraints such as limited scalability, performance bottlenecks, security vulnerabilities, and lack of flexibility in accommodating diverse use cases, thereby eliminating the limitations of conventional single BCFL systems.

11. Open issues and future directions

Integrating Blockchain technology into Federated Learning is a noteworthy research area, offering substantial improvements in protecting privacy models. As highlighted earlier, BCFL is crucial in supporting various domain applications. In this section, we address issues in BCFL and propose potential solutions to shed light on future research directions for readers and researchers in this evolving domain.

- **Privacy Issues in BCFL:** Preserving data privacy in BCFL involves a delicate balance between cryptographic tools and lightweight techniques. While private aggregation using cryptographic tools provides robust parameter privacy, it is computationally intensive and limited in arithmetic operations. On the other hand, noise perturbation methods, such as adding noise and gradient compression, offer a more lightweight approach but come with a trade-off between model performance and data privacy.
  Current research predominantly emphasizes using HE to safeguard against inference attacks. However, a noteworthy drawback of HE is its limited computational efficiency and inability to handle large and complex operations efficiently. This limitation presents challenges in ensuring privacy for extensive and intricate datasets. Additionally, it is essential to note that HE does not inherently guard against collusion attacks, and this vulnerability remains a concern even when utilizing HE for privacy protection in data-intensive scenarios [173].
  Integrating SMPC in BCFL presents notable issues. Firstly, SMPC's interactive nature clashes with the noninteractive protocol required for secure aggregation in BCFL. Additionally, the susceptibility to collusion attacks poses a significant threat even when employing SMPC to protect the model, undermining the overall security and integrity of the FL process within the Blockchain environment. Addressing these issues is crucial for fortifying BCFL against collaboration threats and maintaining trust in the collaborative model training process. In BCFL, employing HE or SMC-based methods may prove impractical for large-scale scenarios due to the considerable increase in communication and computation expenses.
  As a future research direction, explore the seamless integration of zero-knowledge proofs to preserve privacy within the Blockchain-enabled FL framework. This forward-looking methodology empowers participants to validate the accuracy of their updates without revealing the raw data, ensuring an elevated standard of confidentiality and privacy throughout the learning process. Moreover, it has the potential to substantially alleviate the verification burden on clients, thereby paving the way for more efficient and secure systems with enhanced privacy measures. Future research should focus on integrating zero-knowledge proofs in BCFL to enhance privacy, allowing participants to validate updates without revealing raw data. This approach not only ensures elevated confidentiality throughout the learning process but also has the potential to significantly reduce the verification burden on clients, paving the way for more efficient and secure systems with enhanced privacy measures.
- **Computation Overheads in BCFL:** Researchers' incorporation of encryption methods aims to bolster privacy but comes with computational overhead as encrypted gradients are transmitted on the Blockchain. The size of local model gradients plays a

role in determining the communication overhead. In addressing this, some studies employ gradient compression methods to reduce overhead, though this introduces potential issues such as removing pertinent information during compression. These compression-related challenges may have repercussions on the performance of the global model [174].

- **Gas consumption in HE based system:** In [53], an incentive mechanism has been introduced to the BCFL system, leveraging HE applied to local gradients to provide rewards and preserve participant privacy simultaneously. However, the system faces a substantial hurdle, such as unexpectedly high gas consumption during PHE-related smart contract execution. The varying gas costs associated with encryption, additive, and decryption functions, with the additive operation incurring the highest cost, have emerged as a bottleneck, impacting both the economic feasibility and scalability of the BCFL system. These challenges underscore the pressing need to optimize gas consumption for PHE-related functions to ensure the continued effectiveness of the incentivization mechanism while maintaining privacy through HE.

  In the pursuit of advancing the BCFL system, a crucial area for exploration involves optimizing gas consumption for encryption-related smart contract functions. Currently, there is a notable absence of research addressing the reduction of gas consumption in the BCFL system, emphasizing privacy provision at a low cost. Investigating and implementing strategies to achieve low-cost and low gas consumption in the context of privacy-preserving BCFL operations represents a valuable and unexplored avenue for future research and system improvement.
- **Addressing Vulnerabilities in Smart Contracts:** For future directions in this area, it is crucial to prioritize research that delves explicitly into the vulnerabilities of smart contracts within BCFL. Given the inherent risks posed by faulty implementations, leading to persistent vulnerabilities and potential compromise of security and privacy in the model, a focused investigation is needed. Future efforts should aim to comprehensively identify and address these vulnerabilities, offering solutions to enhance the robustness of smart contracts in BCFL. This research would contribute to establishing a more secure foundation for the execution of logic and storage of final states in smart contracts, thereby mitigating risks associated with false data and bolstering overall security in BCFL models. Also, conduct a comprehensive security audit of the smart contract to identify and rectify system performance vulnerabilities.

12. Summary and conclusions

objective

Conribution

Blockchain-enabled FL (BCFL) systems are emerging approaches that combine the principles of FL with Blockchain technology to address various challenges. The main objective of these solutions is to guarantee privacy, security, and trust in decentralized collaborative learning environments while providing a trustworthy and transparent framework for participants. By adopting a privacy perspective, this survey paper presents a systematic overview of the fundamental concepts of BCFL architectures and explores the opportunities and challenges that arise from their development. This survey gives a novel contribution to the present literature because it analyzes in detail the existing attacks on privacy in BCFL, along with state-of-the-art solutions relying on differential privacy, homomorphic encryption, secure multiparty computation, reward-driven approaches, multiple methods, and cross-chained FL. Finally, we investigate the BCFL application in real-case scenarios, such as healthcare, Industry 5.0, and the Internet of Vehicles.

In summary, we analyzed 102 articles published in renowned international conferences, journals, symposiums, and workshops from 2018 to 2023 and focused on privacy aspects of Blockchain-enabled FL. Table 12 represents a quantitative overview of the reviewed research papers divided into topics, whereas Fig. 14 pictures the analyzed number of articles published per year in the reference period.

Table 12
Amount of papers analyzed per topic.

Topic	Amount of papers
BCFL Architecture	16
Attacks to privacy in BCFL	2
BCFL Architectures for Security and Privacy Protection	14
BCFL with Differential Privacy Approaches	19
BCFL with HE Approaches	14
BCFL with SMPC Approaches	5
BCFL with Reward-driven Approaches	17
BCFL with Hybrid Privacy Approaches	9
Cross-chained FL Approaches for privacy	4
Application of BCFL for privacy	31

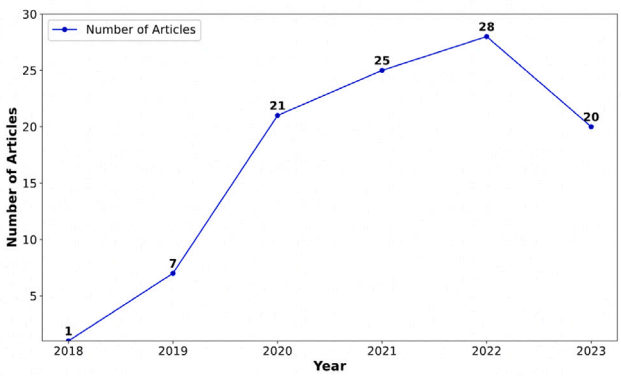


Fig. 14. Literature timeline.

The research direction explored in this paper can be regarded as a foundation, as we plan to continue our investigation by deep-diving into certain aspects only mentioned in the present work. For instance, a fascinating path can be the review of the paper exploiting existing security threats and countermeasures for BCFL systems to give the reader a larger spectrum of diverse problems in this domain. Moreover, an extensive and exhaustive technical description of all the implemented BCFL systems currently available is also a demanding task.

We sincerely aspire for this work to assist researchers and practitioners in comprehending the essential aspects of this field, capturing notable advancements, and highlighting future research progress.

CRediT authorship contribution statement

Sameera K.M.: Writing – original draft, Visualization, Validation, Methodology, Investigation, Conceptualization, Writing – review & editing. **Serena Nicolazzo:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Marco Arazzi:** Writing – review & editing, Writing – original draft, Visualization, Validation. **Antonino Nocera:** Methodology, Investigation, Conceptualization, Supervision, Validation, Writing – review & editing. **Rafidha Rehiman K.A.:** Methodology, Supervision, Validation, Writing – review & editing. **Vinod P.:** Conceptualization, Supervision, Validation, Writing – review & editing. **Mauro Conti:** Writing – review & editing, Conceptualization, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

Funded by the European Union

Finanziato dall'Unione europea NextGenerationEU

This work was supported in part by:

- the HORIZON Europe Framework Programme through the project "OPTIMA - Organization sPecific Threat Intelligence Mining and sharing" (101063107);
- the PRIN 2022 Project "HOMEY: a Human-centric IoE-based Framework for Supporting the Transition Towards Industry 5.0" (code: 2022NX7WKE, CUP: F53D23004340006) funded by the European Union - Next Generation EU;
- the SERICS project (grant number PE00000014) under the NRRP MUR program funded by the EU-NGEU.

References

[1] Y. Qu, M.P. Uddin, C. Gan, Y. Xiang, L. Gao, J. Yearwood, Blockchain-enabled federated learning: A survey, ACM Comput. Surv. 55 (4) (2022) 1–35.
[2] J. Zhu, J. Cao, D. Saxena, S. Jiang, H. Ferradi, Blockchain-empowered federated learning: Challenges, solutions, and future directions, ACM Comput. Surv. 55 (11) (2023) 1–31.
[3] N.S. Bitcoin, Bitcoin: A peer-to-peer electronic cash system, 2008.
[4] D. Li, D. Han, T.-H. Weng, Z. Zheng, H. Li, H. Liu, A. Castiglione, K.-C. Li, Blockchain for federated learning toward secure distributed machine learning systems: A systemic survey, Soft Comput. 26 (9) (2022) 4423–4440.
[5] M. Ali, H. Karimipour, M. Tariq, Integration of blockchain and federated learning for internet of things: Recent advances and future challenges, Comput. Secur. 108 (2021) 102355.
[6] D.C. Nguyen, M. Ding, Q.-V. Pham, P.N. Pathirana, L.B. Le, A. Seneviratne, J. Li, D. Niyato, H.V. Poor, Federated learning meets blockchain in edge computing: Opportunities and challenges, IEEE Internet Things J. 8 (16) (2021) 12806–12825.
[7] W. Issa, N. Moustafa, B. Turnbull, N. Sohrabi, Z. Tari, Blockchain-based federated learning for securing internet of things: A comprehensive survey, ACM Comput. Surv. 55 (9) (2023) 1–43.
[8] A. Qammar, A. Karim, H. Ning, J. Ding, Securing federated learning with blockchain: A systematic literature review, Artif. Intell. Rev. 56 (5) (2023) 3951–3985.
[9] J. Huang, L. Kong, G. Chen, Q. Xiang, X. Chen, X. Liu, Blockchain-based federated learning: A systematic survey, IEEE Network (2022).
[10] D. Li, Z. Luo, B. Cao, Blockchain-based federated learning methodologies in smart environments, Cluster Comput. 25 (4) (2022) 2585–2599.
[11] B. Chhetri, S. Gopali, R. Olapojoye, S. Dehbashi, A.S. Namin, A survey on blockchain-based federated learning and data privacy, in: 2023 IEEE 47th Annual Computers, Software, and Applications Conference, COMPSAC, IEEE, 2023, pp. 1311–1318.
[12] Google, Google scholar, 2023, https://scholar.google.com.
[13] Elsevier, Scopus, 2023, https://www.scopus.com.
[14] Scimago Lab, Scimago, 2023, https://www.scimagojr.com/.
[15] Computing Research & Education, CORE Conference Portal, 2023, https://portal.core.edu.au/conf-ranks.
[16] S. AbdulRahman, H. Tout, H. Ould-Slimane, A. Mourad, C. Talhi, M. Guizani, A survey on federated learning: The journey from centralized to distributed on-site learning and beyond, IEEE Internet Things J. 8 (7) (2020) 5476–5497.
[17] K.H. Lee, N. Verma, A low-power processor with configurable embedded machine-learning accelerators for high-order and adaptive analysis of medical-sensor signals, IEEE J. Solid-State Circuits 48 (7) (2013) 1625–1637.
[18] A. Pacheco, E. Flores, R. Sánchez, S. Almanza-García, Smart classrooms aided by deep neural networks inference on mobile devices, in: 2018 IEEE International Conference on Electro/Information Technology, EIT, IEEE, 2018, pp. 0605–0609.
[19] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, Y. Gao, A survey on federated learning, Knowl.-Based Syst. 216 (2021) 106775.
[20] Y. Cheng, Y. Liu, T. Chen, Q. Yang, Federated learning for privacy-preserving AI, Commun. ACM 63 (12) (2020) 33–36.
[21] H. Zhu, J. Xu, S. Liu, Y. Jin, Federated learning on non-IID data: A survey, Neurocomputing 465 (2021) 371–390.
[22] X. Ma, J. Zhu, Z. Lin, S. Chen, Y. Qin, A state-of-the-art survey on solving non-IID data in federated learning, Future Gener. Comput. Syst. 135 (2022) 244–258.
[23] P. Kairouz, H.B. McMahan, B. Avent, A. Bellet, M. Bennis, A.N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings, et al., Advances and open problems in federated learning, Found. Trends Mach. Learn. 14 (1–2) (2021) 1–210.
[24] Y. Zhao, J. Zhao, L. Jiang, R. Tan, D. Niyato, Z. Li, L. Lyu, Y. Liu, Privacy-preserving blockchain-based federated learning for IoT devices, IEEE Internet Things J. 8 (3) (2020) 1817–1829.
[25] T. Li, A.K. Sahu, A. Talwalkar, V. Smith, Federated learning: Challenges, methods, and future directions, IEEE Signal Process. Mag. 37 (3) (2020) 50–60.
[26] S. Niknam, H.S. Dhillon, J.H. Reed, Federated learning for wireless communications: Motivation, opportunities, and challenges, IEEE Commun. Mag. 58 (6) (2020) 46–51.
[27] A.N. Bhagoji, S. Chakraborty, P. Mittal, S. Calo, Analyzing federated learning through an adversarial lens, in: International Conference on Machine Learning, PMLR, 2019, pp. 634–643.
[28] S. Wang, T. Tuor, T. Salonidis, K.K. Leung, C. Makaya, T. He, K. Chan, Adaptive federated learning in resource constrained edge computing systems, IEEE J. Sel. Areas Commun. 37 (6) (2019) 1205–1221.
[29] J. Kang, Z. Xiong, D. Niyato, S. Xie, J. Zhang, Incentive mechanism for reliable federated learning: A joint optimization approach to combining reputation and contract theory, IEEE Internet Things J. 6 (6) (2019) 10700–10714.
[30] J. Zhang, Y. Wu, R. Pan, Incentive mechanism for horizontal federated learning based on reputation and reverse auction, in: Proceedings of the Web Conference 2021, 2021, pp. 947–956.
[31] M. Nofer, P. Gomber, O. Hinz, D. Schiereck, Blockchain, Bus. Inf. Syst. Eng. 59 (2017) 183–187.
[32] M. Belotti, N. Božić, G. Pujolle, S. Secci, A vademecum on blockchain technologies: When, which, and how, IEEE Commun. Surv. Tutor. 21 (4) (2019) 3796–3838.
[33] M. Jakobsson, A. Juels, Proofs of work and bread pudding protocols, in: Secure Information Networks: Communications and Multimedia Security IFIP TC6/TC11 Joint Working Conference on Communications and Multimedia Security, CMS'99, September 20–21, 1999, Leuven, Belgium, Springer, 1999, pp. 258–272.
[34] K. Košťál, T. Krupa, M. Gembec, I. Vereš, M. Ries, I. Kotuliak, On transition between PoW and PoS, in: 2018 International Symposium ELMAR, IEEE, 2018, pp. 207–210.
[35] V. Buterin, et al., A next-generation smart contract and decentralized application platform, White Paper 3 (37) (2014) 1–36.
[36] E. Androulaki, A. Barger, V. Bortnikov, C. Cachin, K. Christidis, A. De Caro, D. Enyeart, C. Ferris, G. Laventman, Y. Manevich, et al., Hyperledger fabric: A distributed operating system for permissioned blockchains, in: Proceedings of the Thirteenth EuroSys Conference, 2018, pp. 1–15.
[37] S.N. Khan, F. Loukil, C. Ghedira-Guegan, E. Benkhelifa, A. Bani-Hani, Blockchain smart contracts: Applications, challenges, and future trends, Peer-to-peer Network. Appl. 14 (2021) 2901–2925.
[38] R. Yang, R. Wakefield, S. Lyu, S. Jayasuriya, F. Han, X. Yi, X. Yang, G. Amarasinghe, S. Chen, Public and private blockchain in construction business process and information integration, Autom. Constr. 118 (2020) 103276.
[39] P. Grover, A.K. Kar, M. Janssen, Diffusion of blockchain technology: Insights from academic literature and social media analytics, J. Enterprise Inf. Manag. 32 (5) (2019) 735–757.
[40] C.V. Helliar, L. Crawford, L. Rocca, C. Teodori, M. Veneziani, Permissionless and permissioned blockchain diffusion, Int. J. Inf. Manage. 54 (2020) 102136.
[41] W. Ou, S. Huang, J. Zheng, Q. Zhang, G. Zeng, W. Han, An overview on cross-chain: Mechanism, platforms, challenges and advances, Comput. Netw. 218 (2022) 109378.
[42] P. Robinson, Survey of crosschain communications protocols, Comput. Netw. 200 (2021) 108488.
[43] A. Hope-Bailie, S. Thomas, Interledger: Creating a standard for payments, in: Proceedings of the 25th International Conference Companion on World Wide Web, 2016, pp. 281–282.
[44] Y. Lu, X. Huang, K. Zhang, S. Maharjan, Y. Zhang, Blockchain and federated learning for 5G beyond, Ieee Network 35 (1) (2020) 219–225.
[45] G. Xu, Z. Zhou, J. Dong, L. Zhang, X. Song, A blockchain-based federated learning scheme for data sharing in industrial internet of things, IEEE Internet Things J. (2023).
[46] S. Otoum, I. Al Ridhawi, H.T. Mouftah, Blockchain-supported federated learning for trustworthy vehicular networks, in: GLOBECOM 2020-2020 IEEE Global Communications Conference, IEEE, 2020, pp. 1–6.
[47] Y. Qu, S.R. Pokhrel, S. Garg, L. Gao, Y. Xiang, A blockchained federated learning framework for cognitive computing in industry 4.0 networks, IEEE Trans. Ind. Inform. 17 (4) (2020) 2964–2973.
[48] Y. Lu, X. Huang, Y. Dai, S. Maharjan, Y. Zhang, Blockchain and federated learning for privacy-preserved data sharing in industrial IoT, IEEE Trans. Ind. Inform. 16 (6) (2019) 4177–4186.
[49] S. Awan, F. Li, B. Luo, M. Liu, Poster: A reliable and accountable privacy-preserving federated learning framework using the blockchain, in: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, 2019, pp. 2561–2563.

[50] Y.J. Kim, C.S. Hong, Blockchain-based node-aware dynamic weighting methods for improving federated learning performance, in: 2019 20th Asia-Pacific Network Operations and Management Symposium, APNOMS, IEEE, 2019, pp. 1–4.

[51] Y. Li, C. Chen, N. Liu, H. Huang, Z. Zheng, Q. Yan, A blockchain-based decentralized federated learning framework with committee consensus, IEEE Network 35 (1) (2020) 234–241.

[52] T. Rückel, J. Sedlmeir, P. Hofmann, Fairness, integrity, and privacy in a scalable blockchain-based federated learning system, Comput. Netw. 202 (2022) 108621.

[53] M. Qi, Z. Wang, F. Wu, R. Hanson, S. Chen, Y. Xiang, L. Zhu, A blockchain-enabled federated learning model for privacy preservation: System design, in: Information Security and Privacy: 26th Australasian Conference, ACISP 2021, Virtual Event, December 1–3, 2021, Proceedings 26, Springer, 2021, pp. 473–489.

[54] A. Qammar, A. Naouri, J. Ding, H. Ning, Blockchain-based optimized edge node selection and privacy preserved framework for federated learning, Cluster Comput. (2023) 1–16.

[55] S.K. Lo, Y. Liu, Q. Lu, C. Wang, X. Xu, H.-Y. Paik, L. Zhu, Toward trustworthy ai: Blockchain-based architecture design for accountability and fairness of federated learning systems, IEEE Internet Things J. 10 (4) (2022) 3276–3284.

[56] B. Chen, H. Zeng, T. Xiang, S. Guo, T. Zhang, Y. Liu, ESB-FL: Efficient and secure blockchain-based federated learning with fair payment, IEEE Trans. Big Data (2022).

[57] D. Lamken, T. Wagner, T. Hoiss, K. Seidenfad, A. Hermann, M. Kus, U. Lechner, Design patterns and framework for blockchain integration in supply chains, in: 2021 IEEE International Conference on Blockchain and Cryptocurrency, ICBC, IEEE, 2021, pp. 1–3.

[58] S. Yuan, B. Cao, M. Peng, Y. Sun, Chainsfl: Blockchain-driven federated learning from design to realization, in: 2021 IEEE Wireless Communications and Networking Conference, WCNC, IEEE, 2021, pp. 1–6.

[59] Y. He, K. Huang, G. Zhang, F.R. Yu, J. Chen, J. Li, Bift: A blockchain-based federated learning system for connected and autonomous vehicles, IEEE Internet Things J. 9 (14) (2021) 12311–12322.

[60] S. Otoum, I. Al Ridhawi, H. Mouftah, Securing critical IoT infrastructures with blockchain-supported federated learning, IEEE Internet Things J. 9 (4) (2021) 2592–2601.

[61] C. Feng, B. Liu, K. Yu, S.K. Goudos, S. Wan, Blockchain-empowered decentralized horizontal federated learning for 5G-enabled UAVs, IEEE Trans. Ind. Inform. 18 (5) (2021) 3582–3592.

[62] N. Wang, W. Yang, X. Wang, L. Wu, Z. Guan, X. Du, M. Guizani, A blockchain based privacy-preserving federated learning scheme for internet of vehicles, Digit. Commun. Netw. (2022).

[63] Y. Lu, X. Huang, K. Zhang, S. Maharjan, Y. Zhang, Low-latency federated learning and blockchain for edge association in digital twin empowered 6G networks, IEEE Trans. Ind. Inform. 17 (7) (2020) 5098–5107.

[64] H. Kim, J. Park, M. Bennis, S.-L. Kim, Blockchained on-device federated learning, IEEE Commun. Lett. 24 (6) (2019) 1279–1283.

[65] S.R. Pokhrel, J. Choi, Federated learning with blockchain for autonomous vehicles: Analysis and design challenges, IEEE Trans. Commun. 68 (8) (2020) 4734–4746.

[66] Y. Zhan, J. Zhang, Z. Hong, L. Wu, P. Li, S. Guo, A survey of incentive mechanism design for federated learning, IEEE Trans. Emerg. Top. Comput. 10 (2) (2021) 1035–1044.

[67] S. Salim, B. Turnbull, N. Moustafa, A blockchain-enabled explainable federated learning for securing internet-of-things-based social media 3.0 networks, IEEE Trans. Comput. Soc. Syst. (2021).

[68] J. Chen, J. Xue, Y. Wang, L. Huang, T. Baker, Z. Zhou, Privacy-preserving and traceable federated learning for data sharing in industrial IoT applications, Expert Syst. Appl. 213 (2023) 119036.

[69] B. Jia, X. Zhang, J. Liu, Y. Zhang, K. Huang, Y. Liang, Blockchain-enabled federated learning data protection aggregation scheme with differential privacy and homomorphic encryption in IIoT, IEEE Trans. Ind. Inform. 18 (6) (2021) 4049–4058.

[70] R.N. Zaeem, K.S. Barber, The effect of the GDPR on privacy policies: Recent progress and future promise, ACM Trans. Manag. Inf. Syst. (TMIS) 12 (1) (2020) 1–20.

[71] A. Act, Health insurance portability and accountability act of 1996, Public Law 104 (1996) 191.

[72] N. Truong, K. Sun, S. Wang, F. Guitton, Y. Guo, Privacy preservation in federated learning: An insightful survey from the GDPR perspective, Comput. Secur. 110 (2021) 102402.

[73] L. Zhu, Z. Liu, S. Han, Deep leakage from gradients, in: Advances in Neural Information Processing Systems, vol. 32, 2019.

[74] L. Melis, C. Song, E. De Cristofaro, V. Shmatikov, Exploiting unintended feature leakage in collaborative learning, in: 2019 IEEE Symposium on Security and Privacy, SP, IEEE, 2019, pp. 691–706.

[75] Z. Wang, M. Song, Z. Zhang, Y. Song, Q. Wang, H. Qi, Beyond inferring class representatives: User-level privacy leakage from federated learning, in: IEEE INFOCOM 2019-IEEE Conference on Computer Communications, IEEE, 2019, pp. 2512–2520.

[76] M. Nasr, R. Shokri, A. Houmansadr, Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning, in: 2019 IEEE Symposium on Security and Privacy, SP, IEEE, 2019, pp. 739–753.

[77] H. Li, Y. Sun, Y. Yu, D. Li, Z. Guan, J. Liu, Privacy-preserving Cross-silo federated learning atop blockchain for IoT, IEEE Internet Things J. (2023).

[78] M. Shen, H. Wang, B. Zhang, L. Zhu, K. Xu, Q. Li, X. Du, Exploiting unintended property leakage in blockchain-assisted federated learning for intelligent edge computing, IEEE Internet Things J. 8 (4) (2020) 2265–2275.

[79] Y. Qu, L. Gao, T.H. Luan, Y. Xiang, S. Yu, B. Li, G. Zheng, Decentralized privacy using blockchain-enabled federated learning in fog computing, IEEE Internet Things J. 7 (6) (2020) 5171–5183.

[80] S. Singh, S. Rathore, O. Alfarraj, A. Tolba, B. Yoon, A framework for privacy-preservation of IoT healthcare data using federated learning and blockchain technology, Future Gener. Comput. Syst. 129 (2022) 380–388.

[81] J. Xu, J. Lin, W. Liang, K.-C. Li, Privacy preserving personalized blockchain reliability prediction via federated learning in IoT environments, Cluster Comput. 25 (4) (2022) 2515–2526.

[82] R. Kumar, A.A. Khan, J. Kumar, N.A. Golilarz, S. Zhang, Y. Ting, C. Zheng, W. Wang, et al., Blockchain-federated-learning and deep learning models for covid-19 detection using ct imaging, IEEE Sens. J. 21 (14) (2021) 16301–16314.

[83] M. Guduri, C. Chakraborty, M. Margala, et al., Blockchain-based federated learning technique for privacy preservation and security of smart electronic health records, IEEE Trans. Consum. Electron. (2023).

[84] Z. Li, H. Yu, T. Zhou, L. Luo, M. Fan, Z. Xu, G. Sun, Byzantine resistant secure blockchained federated learning at the edge, Ieee Network 35 (4) (2021) 295–301.

[85] H. Kasyap, S. Tripathy, Privacy-preserving and Byzantine-robust federated learning framework using permissioned blockchain, Expert Syst. Appl. (2023) 122210.

[86] S.M.H. Bamakan, A. Motavali, A.B. Bondarti, A survey of blockchain consensus algorithms performance evaluation criteria, Expert Syst. Appl. 154 (2020) 113385.

[87] C. Xu, Y. Qu, T.H. Luan, P.W. Eklund, Y. Xiang, L. Gao, An efficient and reliable asynchronous federated learning scheme for smart public transportation, IEEE Trans. Veh. Technol. (2022).

[88] L. Feng, Y. Zhao, S. Guo, X. Qiu, W. Li, P. Yu, BAFL: A blockchain-based asynchronous federated learning framework, IEEE Trans. Comput. 71 (5) (2021) 1092–1103.

[89] M. Sarhan, W.W. Lo, S. Layeghy, M. Portmann, HBFL: A hierarchical blockchain-based federated learning framework for collaborative IoT intrusion detection, Comput. Electr. Eng. 103 (2022) 108379.

[90] S. Guo, K. Zhang, B. Gong, L. Chen, Y. Ren, F. Qi, X. Qiu, Sandbox computing: A data privacy trusted sharing paradigm via blockchain and federated learning, IEEE Trans. Comput. 72 (3) (2022) 800–810.

[91] Z. Yang, Y. Shi, Y. Zhou, Z. Wang, K. Yang, Trustworthy federated learning via blockchain, IEEE Internet Things J. 10 (1) (2022) 92–109.

[92] P.C.M. Arachchige, P. Bertok, I. Khalil, D. Liu, S. Camtepe, M. Atiquzzaman, Local differential privacy for deep learning, IEEE Internet Things J. 7 (7) (2019) 5827–5842.

[93] K. Wei, J. Li, M. Ding, C. Ma, H.H. Yang, F. Farokhi, S. Jin, T.Q. Quek, H.V. Poor, Federated learning with differential privacy: Algorithms and performance analysis, IEEE Trans. Inf. Forensics Secur. 15 (2020) 3454–3469.

[94] Y. Qi, M.S. Hossain, J. Nie, X. Li, Privacy-preserving blockchain-based federated learning for traffic flow prediction, Future Gener. Comput. Syst. 117 (2021) 328–337.

[95] Y. Wang, Z. Su, N. Zhang, A. Benslimane, Learning in the air: Secure federated learning for UAV-assisted crowdsensing, IEEE Trans. Netw. Sci. Eng. 8 (2) (2020) 1055–1069.

[96] Y. Wan, Y. Qu, L. Gao, Y. Xiang, Privacy-preserving blockchain-enabled federated learning for B5G-Driven edge computing, Comput. Netw. 204 (2022) 108671.

[97] Y. Qu, L. Gao, Y. Xiang, S. Shen, S. Yu, Fedtwin: Blockchain-enabled adaptive asynchronous federated learning for digital twin networks, IEEE Network 36 (6) (2022) 183–190.

[98] L. Cui, Y. Qu, G. Xie, D. Zeng, R. Li, S. Shen, S. Yu, Security and privacy-enhanced federated learning for anomaly detection in IoT infrastructures, IEEE Trans. Ind. Inform. 18 (5) (2021) 3492–3500.

[99] H. Zhang, G. Li, Y. Zhang, K. Gai, M. Qiu, Blockchain-based privacy-preserving medical data sharing scheme using federated learning, in: Knowledge Science, Engineering and Management: 14th International Conference, KSEM 2021, Tokyo, Japan, August 14–16, 2021, Proceedings, Part III 14, Springer, 2021, pp. 634–646.

[100] L. Javed, A. Anjum, B.M. Yakubu, M. Iqbal, S.A. Moqurrab, G. Srivastava, ShareChain: Blockchain-enabled model for sharing patient data using federated learning and differential privacy, Expert Syst. 40 (5) (2023) e13131.

[101] Y. Liu, P. Liu, W. Jing, H.H. Song, PD2s: A privacy-preserving differentiated data sharing scheme based on blockchain and federated learning, IEEE Internet Things J. (2023).

[102] J. Li, Y. Shao, K. Wei, M. Ding, C. Ma, L. Shi, Z. Han, H.V. Poor, Blockchain assisted decentralized federated learning (BLADE-FL): Performance analysis and resource allocation, IEEE Trans. Parallel Distrib. Syst. 33 (10) (2021) 2401–2415.

[103] S. Ji, J. Zhang, Y. Zhang, Z. Han, C. Ma, LAFED: A lightweight authentication mechanism for blockchain-enabled federated learning system, Future Gener. Comput. Syst. 145 (2023) 56–67.

[104] Q. Miao, H. Lin, J. Hu, X. Wang, An intelligent and privacy-enhanced data sharing strategy for blockchain-empowered internet of things, Digit. Commun. Netw. 8 (5) (2022) 636–643.

[105] S. Zhang, J. Zhu, Privacy protection federated learning framework based on blockchain and committee consensus in IoT devices, in: 2023 IEEE 47th Annual Computers, Software, and Applications Conference, COMPSAC, IEEE, 2023, pp. 627–636.

[106] P.C.M. Arachchige, P. Bertok, I. Khalil, D. Liu, S. Camtepe, M. Atiquzzaman, A trustworthy privacy preserving framework for machine learning in industrial IoT systems, IEEE Trans. Ind. Inform. 16 (9) (2020) 6092–6102.

[107] E. Antwi-Boasiako, S. Zhou, Y. Liao, Q. Liu, Y. Wang, K. Owusu-Agyemang, Privacy preservation in distributed deep learning: A survey on distributed deep learning, privacy preservation techniques used and interesting research directions, J. Inf. Secur. Appl. 61 (2021) 102949.

[108] C. Zhang, S. Li, J. Xia, W. Wang, F. Yan, Y. Liu, {BatchCrypt}: Efficient homomorphic encryption for {Cross-Silo} federated learning, in: 2020 USENIX Annual Technical Conference, USENIX ATC 20, 2020, pp. 493–506.

[109] L. Zhang, J. Xu, P. Vijayakumar, P.K. Sharma, U. Ghosh, Homomorphic encryption-based privacy-preserving federated learning in iot-enabled healthcare system, IEEE Trans. Netw. Sci. Eng. (2022).

[110] Y. Chen, J. Li, F. Wang, K. Yue, Y. Li, B. Xing, L. Zhang, L. Chen, DS2PM: A data sharing privacy protection model based on blockchain and federated learning, IEEE Internet Things J. (2021).

[111] J. Weng, J. Weng, J. Zhang, M. Li, Y. Zhang, W. Luo, Deepchain: Auditable and privacy-preserving deep learning with blockchain-based incentive, IEEE Trans. Dependable Secure Comput. 18 (5) (2019) 2438–2455.

[112] Z. Sun, J. Wan, L. Yin, Z. Cao, T. Luo, B. Wang, A blockchain-based audit approach for encrypted data in federated learning, Digit. Commun. Netw. 8 (5) (2022) 614–624.

[113] Y. Miao, Z. Liu, H. Li, K.-K.R. Choo, R.H. Deng, Privacy-preserving Byzantine-robust federated learning via blockchain systems, IEEE Trans. Inf. Forensics Secur. 17 (2022) 2848–2861.

[114] B.B. Sezer, H. Turkmen, U. Nuriyev, Ppfchain: A novel framework privacy-preserving blockchain-based federated learning method for sensor networks, Internet Things 22 (2023) 100781.

[115] H. Zhu, R. Wang, Y. Jin, K. Liang, J. Ning, Distributed additive encryption and quantization for privacy preserving federated deep learning, Neurocomputing 463 (2021) 309–327.

[116] T. Muazu, M. Yingchi, A.U. Muhammad, M. Ibrahim, O. Samuel, P. Tiwari, IoMT: A medical resource management system using edge empowered blockchain federated learning, IEEE Trans. Netw. Serv. Manag. (2023).

[117] J.A. Alzubi, O.A. Alzubi, A. Singh, M. Ramachandran, Cloud-IIoT-based electronic health record privacy-preserving by CNN and blockchain-enabled federated learning, IEEE Trans. Ind. Inform. 19 (1) (2022) 1080–1087.

[118] M. Arazzi, S. Nicolazzo, A. Nocera, A fully privacy-preserving solution for anomaly detection in IoT using federated learning and homomorphic encryption, Information Systems Frontiers (2023) 1–24.

[119] Y. Li, X. Tao, X. Zhang, J. Liu, J. Xu, Privacy-preserved federated learning for autonomous driving, IEEE Trans. Intell. Transp. Syst. 23 (7) (2021) 8423–8434.

[120] Z. Zhou, Y. Tian, J. Xiong, J. Ma, C. Peng, Blockchain-enabled secure and trusted federated data sharing in IIoT, IEEE Trans. Ind. Inform. (2022).

[121] C. Zhao, S. Zhao, M. Zhao, Z. Chen, C.-Z. Gao, H. Li, Y.-a. Tan, Secure multi-party computation: Theory, practice and applications, Inform. Sci. 476 (2019) 357–372.

[122] C. Jiang, C. Xu, Y. Zhang, PFLM: Privacy-preserving federated learning with membership proof, Inform. Sci. 576 (2021) 288–311.

[123] C. Fang, Y. Guo, J. Ma, H. Xie, Y. Wang, A privacy-preserving and verifiable federated learning method based on blockchain, Comput. Commun. 186 (2022) 1–11.

[124] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H.B. McMahan, S. Patel, D. Ramage, A. Segal, K. Seth, Practical secure aggregation for privacy-preserving machine learning, in: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017, pp. 1175–1191.

[125] A.P. Kalapaaking, I. Khalil, X. Yi, Blockchain-based federated learning with SMPC model verification against poisoning attack for healthcare systems, IEEE Trans. Emerg. Top. Comput. (2023).

[126] Q. Wang, Y. Guo, X. Wang, T. Ji, L. Yu, P. Li, AI at the edge: Blockchain-empowered secure multiparty learning with heterogeneous models, IEEE Internet Things J. 7 (10) (2020) 9600–9610.

[127] J. Qi, F. Lin, Z. Chen, C. Tang, R. Jia, M. Li, High-quality model aggregation for blockchain-based federated learning via reputation-motivated task participation, IEEE Internet Things J. 9 (19) (2022) 18378–18391.

[128] M. Abdel-Basset, N. Moustafa, H. Hawash, Privacy-preserved cyberattack detection in industrial edge of things (IEOT): A blockchain-orchestrated federated learning approach, IEEE Trans. Ind. Inform. 18 (11) (2022) 7920–7934.

[129] K. Toyoda, A.N. Zhang, Mechanism design for an incentive-aware blockchain-enabled federated learning platform, in: 2019 IEEE International Conference on Big Data, Big Data, IEEE, 2019, pp. 395–403.

[130] D. Hamouda, M.A. Ferrag, N. Benhamida, H. Seridi, PPSS: A privacy-preserving secure framework using blockchain-enabled federated deep learning for industrial IoTs, Pervasive Mob. Comput. 88 (2022) 101738.

[131] H. Kasyap, A. Manna, S. Tripathy, An efficient blockchain assisted reputation aware decentralized federated learning framework, IEEE Trans. Netw. Serv. Manag. (2022).

[132] Z. Batool, K. Zhang, M. Toews, Fl-mab: client selection and monetization for blockchain-based federated learning, in: Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing, 2022, pp. 299–307.

[133] J. Kang, R. Yu, X. Huang, M. Wu, S. Maharjan, S. Xie, Y. Zhang, Blockchain for secure and efficient data sharing in vehicular edge computing and networks, IEEE Internet Things J. 6 (3) (2018) 4660–4670.

[134] J. Kang, Z. Xiong, D. Niyato, Y. Zou, Y. Zhang, M. Guizani, Reliable federated learning for mobile networks, IEEE Wirel. Commun. 27 (2) (2020) 72–80.

[135] Y. Qu, C. Xu, L. Gao, Y. Xiang, S. Yu, Fl-sec: Privacy-preserving decentralized federated learning using signsgd for the internet of artificially intelligent things, IEEE Internet Things Mag. 5 (1) (2022) 85–90.

[136] X. Qu, S. Wang, Q. Hu, X. Cheng, Proof of federated learning: A novel energy-recycling consensus algorithm, IEEE Trans. Parallel Distrib. Syst. 32 (8) (2021) 2074–2085.

[137] X. Zhu, H. Li, Privacy-preserving decentralized federated deep learning, in: Proceedings of the ACM Turing Award Celebration Conference-China, 2021, pp. 33–38.

[138] M. Shayan, C. Fung, C.J. Yoon, I. Beschastnikh, Biscotti: A blockchain system for private and secure federated learning, IEEE Trans. Parallel Distrib. Syst. 32 (7) (2020) 1513–1525.

[139] J. Sun, Y. Wu, S. Wang, Y. Fu, X. Chang, Permissioned blockchain frame for secure federated learning, IEEE Commun. Lett. 26 (1) (2021) 13–17.

[140] S. Bai, G. Yang, G. Liu, H. Dai, C. Rong, NttpFL: Privacy-preserving oriented no trusted third party federated learning system based on blockchain, IEEE Trans. Netw. Serv. Manag. 19 (4) (2022) 3750–3763.

[141] O. Samuel, A.B. Omojo, A.M. Onuja, Y. Sunday, P. Tiwari, D. Gupta, G. Hafeez, A.S. Yahaya, O.J. Fatoba, S. Shamshirband, IoMT: A COVID-19 healthcare system driven by federated learning and blockchain, IEEE J. Biomed. Health Inf. 27 (2) (2022) 823–834.

[142] A. Smahi, H. Li, Y. Yang, X. Yang, P. Lu, Y. Zhong, C. Liu, BV-ICVs: A privacy-preserving and verifiable federated learning framework for V2X environments using blockchain and zkSNARKs, J. King Saud Univ.-Comput. Inf. Sci. (2023) 101542.

[143] H. Liu, S. Zhang, P. Zhang, X. Zhou, X. Shao, G. Pu, Y. Zhang, Blockchain and federated learning for collaborative intrusion detection in vehicular edge computing, IEEE Trans. Veh. Technol. 70 (6) (2021) 6073–6084.

[144] F. Yang, Y. Qiao, M.Z. Abedin, C. Huang, Privacy-preserved credit data sharing integrating blockchain and federated learning for industrial 4.0, IEEE Trans. Ind. Inform. 18 (12) (2022) 8755–8764.

[145] J. Akram, M. Umair, R.H. Jhaveri, M.N. Riaz, H. Chi, S. Malebary, Chained-drones: Blockchain-based privacy-preserving framework for secure and intelligent service provisioning in internet of drone things, Comput. Electr. Eng. 110 (2023) 108772.

[146] W. Zhang, Q. Lu, Q. Yu, Z. Li, Y. Liu, S.K. Lo, S. Chen, X. Xu, L. Zhu, Blockchain-based federated learning for device failure detection in industrial IoT, IEEE Internet Things J. 8 (7) (2020) 5926–5937.

[147] S.K. Singh, L.T. Yang, J.H. Park, FusionFedBlock: Fusion of blockchain and federated learning to preserve privacy in industry 5.0, Inf. Fusion 90 (2023) 233–240.

[148] A. Yazdinejad, A. Dehghantanha, R.M. Parizi, M. Hammoudeh, H. Karimipour, G. Srivastava, Block hunter: Federated learning for cyber threat hunting in blockchain-based iiot networks, IEEE Trans. Ind. Inform. 18 (11) (2022) 8356–8366.

[149] H. Chai, S. Leng, Y. Chen, K. Zhang, A hierarchical blockchain-enabled federated learning algorithm for knowledge sharing in internet of vehicles, IEEE Trans. Intell. Transp. Syst. 22 (7) (2020) 3975–3986.

[150] D. Połap, G. Srivastava, K. Yu, Agent architecture of an intelligent medical system based on federated learning and blockchain technology, J. Inf. Secur. Appl. 58 (2021) 102748.

[151] T. Hai, J. Zhou, S. Srividhya, S.K. Jain, P. Young, S. Agrawal, BVFLEMR: An integrated federated learning and blockchain technology for cloud-based medical records recommendation system, J. Cloud Comput. 11 (1) (2022) 22.

[152] Y. Liu, W. Yu, Z. Ai, G. Xu, L. Zhao, Z. Tian, A blockchain-empowered federated learning in healthcare-based cyber physical systems, IEEE Trans. Netw. Sci. Eng. (2022).

[153] O. Friha, M.A. Ferrag, L. Shu, L. Maglaras, K.-K.R. Choo, M. Nafaa, FELIDS: Federated learning-based intrusion detection system for agricultural internet of things, J. Parallel Distrib. Comput. 165 (2022) 17–31.

[154] S.H. Alsamhi, F.A. Almalki, F. Afghah, A. Hawbani, A.V. Shvetsov, B. Lee, H. Song, Drones' edge intelligence over smart environments in B5G: Blockchain and federated learning synergy, IEEE Trans. Green Commun. Netw. 6 (1) (2021) 295–312.

[155] A. Lakhan, M.A. Mohammed, J. Nedoma, R. Martinek, P. Tiwari, A. Vidyarthi, A. Alkhayyat, W. Wang, Federated-learning based privacy preservation and fraud-enabled blockchain IoMT system for healthcare, IEEE J. Biomed. Health Inform. 27 (2) (2022) 664–672.

[156] T. Moulahi, R. Jabbar, A. Alabdulatif, S. Abbas, S. El Khediri, S. Zidi, M. Rizwan, Privacy-preserving federated learning cyber-threat detection for intelligent transport systems with blockchain-based security, Expert Syst. 40 (5) (2023) e13103.

[157] E.T.M. Beltrán, M.Q. Pérez, P.M.S. Sánchez, S.L. Bernal, G. Bovet, M.G. Pérez, G.M. Pérez, A.H. Celdrán, Decentralized federated learning: Fundamentals, state of the art, frameworks, trends, and challenges, IEEE Commun. Surv. Tutor. (2023).

[158] A. Imteaj, U. Thakker, S. Wang, J. Li, M.H. Amini, A survey on federated learning for resource-constrained IoT devices, IEEE Internet Things J. 9 (1) (2021) 1–24.

[159] U. Majeed, C.S. Hong, Flchain: Federated learning via MEC-enabled blockchain network, in: 2019 20th Asia-Pacific Network Operations and Management Symposium, APNOMS, IEEE, 2019, pp. 1–4.

[160] X. Jiang, F.R. Yu, T. Song, Z. Ma, Y. Song, D. Zhu, Blockchain-enabled cross-domain object detection for autonomous driving: A model sharing approach, IEEE Internet Things J. 7 (5) (2020) 3681–3692.

[161] J. Kang, X. Li, J. Nie, Y. Liu, M. Xu, Z. Xiong, D. Niyato, Q. Yan, Communication-efficient and cross-chain empowered federated learning for artificial intelligence of things, IEEE Trans. Netw. Sci. Eng. 9 (5) (2022) 2966–2977.

[162] H. Jin, X. Dai, J. Xiao, B. Li, H. Li, Y. Zhang, Cross-cluster federated learning and blockchain for internet of medical things, IEEE Internet Things J. 8 (21) (2021) 15776–15784.

[163] J. Kang, J. Wen, D. Ye, B. Lai, T. Wu, Z. Xiong, J. Nie, D. Niyato, Y. Zhang, S. Xie, Blockchain-empowered federated learning for healthcare metaverses: User-centric incentive mechanism with optimal data freshness, IEEE Trans. Cogn. Commun. Network. (2023).

[164] D. Kahneman, A. Tversky, Prospect theory: An analysis of decision under risk, in: Handbook of the Fundamentals of Financial Decision Making: Part I, World Scientific, 2013, pp. 99–127.

[165] R. Xu, Y. Chen, μDFL: A secure microchained decentralized federated learning fabric atop IoT networks, IEEE Trans. Netw. Serv. Manag. 19 (3) (2022) 2677–2688.

[166] R.S. Antunes, C. André da Costa, A. Küderle, I.A. Yari, B. Eskofier, Federated learning for healthcare: Systematic review and architecture proposal, ACM Trans. Intell. Syst. Technol. 13 (4) (2022) 1–23.

[167] R. Myrzashova, S.H. Alsamhi, A.V. Shvetsov, A. Hawbani, X. Wei, Blockchain meets federated learning in healthcare: A systematic review with challenges and opportunities, IEEE Internet Things J. (2023).

[168] J. Leng, W. Sha, B. Wang, P. Zheng, C. Zhuang, Q. Liu, T. Wuest, D. Mourtzis, L. Wang, Industry 5.0: Prospect and retrospect, J. Manuf. Syst. 65 (2022) 279–295.

[169] Y. Liu, X. Ma, L. Shu, G.P. Hancke, A.M. Abu-Mahfouz, From industry 4.0 to agriculture 4.0: Current status, enabling technologies, and research challenges, IEEE Trans. Ind. Inform. 17 (6) (2020) 4322–4334.

[170] F. Yang, S. Wang, J. Li, Z. Liu, Q. Sun, An overview of internet of vehicles, China Commun. 11 (10) (2014) 1–15.

[171] O. Kaiwartya, A.H. Abdullah, Y. Cao, A. Altameem, M. Prasad, C.-T. Lin, X. Liu, Internet of vehicles: Motivation, layered architecture, network model, challenges, and future aspects, IEEE Access 4 (2016) 5356–5373.

[172] S.H. Alsamhi, F.A. Almalki, H. Al-Dois, A.V. Shvetsov, M.S. Ansari, A. Hawbani, S.K. Gupta, B. Lee, Multi-drone edge intelligence and SAR smart wearable devices for emergency communication, Wirel. Commun. Mob. Comput. 2021 (2021) 1–12.

[173] M. Hao, H. Li, X. Luo, G. Xu, H. Yang, S. Liu, Efficient and privacy-enhanced federated learning for industrial artificial intelligence, IEEE Trans. Ind. Inform. 16 (10) (2019) 6532–6542.

[174] L. Cui, X. Su, Z. Ming, Z. Chen, S. Yang, Y. Zhou, W. Xiao, CREAT: Blockchain-assisted compression algorithm of federated learning for content caching in edge computing, IEEE Internet Things J. 9 (16) (2020) 14151–14161.