Applied Science & Engineering Journal for Advanced Research *Peer Reviewed and Refereed Journal*
ISSN (Online): 2583-2468
Volume-3 Issue-3 || May 2024 || PP. 34-39 DOI: 10.5281/zenodo.12643107

Blockchain based Federated Learning Models Methods and Applications

Rahul Sharma[1], Kritika Sharma[2] and Priya Patel[3]
[1]*Financial Risk, Indian School of Business (ISB), Hyderabad, India*
[2]*Business Administration, Indian Institute of Management (IIM), Bangalore, India*
[3]*Electronic Information Engineering, Indian Institute of Technology (IIT), Kanpur, India*

[1]*Corresponding Author: Gupta889@gmail.com*

Received: 21-04-2024	Revised: 15-05-2024	Accepted: 29-05-2024

ABSTRACT
This paper systematically discusses the application and development of federated learning in data privacy protection and data value sharing. With the rapid development of global information technology, especially the explosive growth of data from Internet of Things devices, data security and privacy protection are facing unprecedented challenges. This paper first analyzes the growth trend of global data volume and its importance to next generation technologies such as artificial intelligence technologies such as deep learning. Second, the paper provides an in-depth look at the impact of current data privacy regulations on data flows and value creation, particularly the EU's GDPR and China's Data Security and Personal Information Protection Law. Then, this paper introduces in detail federated learning, as a new distributed machine learning paradigm, which effectively solves the contradiction between existing data sharing and privacy protection by protecting individual data privacy and realizing global model collaborative construction. Finally, this paper discusses the combination of blockchain technology and federated learning, and proposes BeFL architecture as a new secure, decentralized and trusted federated learning system, which is expected to provide a comprehensive solution for large-scale data processing and value creation in multi-party scenarios. The research in this paper not only deepens the understanding of federation learning in theory, but also provides important reference and enlightenment for future research and application in related fields.

Keywords: *federated learning, data privacy protection, data circulation, blockchain technology*

I. INTRODUCTION

With the rapid development of information and intelligent technologies, data generated by the Internet and IoT devices is experiencing explosive growth. According to the International Data Corporation (IDC), global data is projected to reach 175 zettabytes (ZB) by 2025, with 90 ZB coming from IoT devices. Data interaction users are expected to increase from 5 billion in 2018 to 6 billion. Data has become a crucial and effective production factor, particularly for next-generation artificial intelligence technologies like deep learning. These require substantial data to train optimal models and enhance system performance and efficiency.

However, while data plays a crucial role, its collection and use also impact personal and national security. Consequently, regulations on data privacy protection, such as the EU's General Data Protection Regulation (GDPR) in 2018 and China's Data Security Law and Personal Information Protection Law in 2021, have become increasingly stringent. These regulations aim to protect data security and privacy but also restrict data flow and value creation to some extent, leading to data fragmentation across isolated data islands due to security, privacy, or geographical factors. Therefore, ensuring data circulation, sharing, and enhancing collaboration efficiency between institutions under privacy and security has become a pressing concern in academia and industry. Federated learning has emerged as a new paradigm in distributed machine learning in recent years. It allows institutions to collaborate on building a global model without sharing private data offsite or disclosing underlying raw data or its encrypted forms. Federated learning effectively reduces privacy leakage and data compliance risks by iteratively exchanging and updating encrypted parameters. Consequently, federated learning has garnered extensive and in-depth research attention.

The theoretical foundation of federated learning can be traced back to 1996 when Cheung et al. first implemented association rule mining in distributed databases, laying the groundwork for its theoretical and methodological development. In 2016, Google formally introduced federated learning, which was initially applied to optimize input methods. In 2017, Tan et al. proposed Distant Domain Transfer Learning, integrating transfer learning with federated learning, and later introduced federated transfer learning in 2020 to address the data island problem.

In recent years, federated learning has evolved into three service forms in industrial practice: horizontal federated learning (sample dimension), vertical federated learning (feature dimension), and federated transfer learning. Application

ISSN (Online): 2583-2468
Volume-3 Issue-3 || May 2024 || PP. 34-39 DOI: 10.5281/zenodo.12643107

architectures include client-central coordinators and peer-to-peer networks for horizontal federated learning. In contrast, vertical federated learning typically involves a third-party coordinator assuming mutual trust between two parties for privacy computation. Federated transfer learning combines traditional methods with privacy-protecting distributed machine learning architectures. This introduction sets the stage for discussing the development, methods, and applications of federated learning in various scenarios based on the distribution characteristics of participant data.

II. RELATED WORK

2.1 Blockchain-Enabled Federated Learning, BeFL

In recent years, combining blockchain technology with federated learning has emerged as a new trend to safeguard data security and privacy, and to construct novel infrastructures for data element circulation. The integration of these two technologies mutually benefits each other, providing a more comprehensive solution that incentivizes collaborative data training in federated learning while ensuring data privacy and security [15−17].

The potential advantages of Blockchain-enabled federated learning (BeFL) architecture can be summarized in three aspects:

Similar Cooperation Modes: Blockchain and federated learning both operate on distributed, multi-party collaborative network architectures. Blockchain can thus serve as the foundational topology for federated learning.

Trusted Features: Federated learning ensures trust by protecting privacy during data collaboration, while blockchain ensures trust through consensus and data validation mechanisms in its ledger process, making data tamper-proof and irrefutable.

Complementary Applications: Federated learning aims to "create value" by leveraging the complementary nature of data from various participants to enhance model effectiveness. Blockchain aims to "transfer value" by transparently recording all parties' contributions and providing rewards. Therefore, the integration of blockchain and federated learning is not only a development trend but also the primary research motivation of this paper [18−19].

Currently, research on Blockchain-enabled federated learning is still in its infancy. Existing studies mostly focus on application patterns combining blockchain with federated learning in typical scenarios such as edge computing, IoT, and vehicular networks. However, there is a lack of comprehensive review articles summarizing architectural models, theoretical frameworks, technological advancements, and applications specific to this field [20−28]. Recent domestic and international reviews differ from this paper as summarized in Table 1. Overall, based on comprehensive research in this field, this paper introduces the concept model of integrating blockchain with federated learning and systematically summarizes key issues, research methods, current advancements, application domains, and future research directions, aiming to provide valuable insights for the development of this field.

2.2 Federated Learning Model for Blockchain

To comprehensively outline the current research status and advancements in integrating blockchain and federated learning, this paper introduces the Blockchain-enabled Federated Learning (BeFL) architecture model. As depicted, the BeFL architecture is structured from bottom to top, encompassing the following layers: Infrastructure Layer, Consensus Mechanism Layer, Economic Incentive Layer, Smart Contract Layer, Privacy and Security Layer, and Application Domain Layer.

1. **Infrastructure Layer:** This layer adopts the blockchain's decentralized (or weakly centralized) distributed system architecture as the underlying framework for federated learning, ensuring system robustness and data credibility.
2. **Consensus Mechanism Layer:** Drawing from existing blockchain consensus algorithms and enhancements, this layer aims to prevent resource wastage and guarantee the trustworthiness of federated learning outcomes.
3. **Economic Incentive Layer:** This layer primarily designs reward mechanisms to maximize user node benefits and incentivize more nodes to join the collaborative learning process.
4. **Smart Contract Layer:** Smart contracts replace central servers for automated process management in federated learning, serving as the interface and carrier for integrating blockchain with artificial intelligence technologies.
5. **Privacy and Security Layer:** Leveraging encryption and privacy computing technologies to ensure the security of federated learning systems and user privacy.

The basic operational flow of the BeFL architecture comprises three main stages: Participant Node Authorization and Selection, Federated Modeling of Computational Processes, and Trusted Storage of Computational Results.

- Each participant holds their private data in the Participant Node Authorization and Selection stage. The task issuer broadcasts federated learning tasks via smart contracts deployed on the blockchain network and uses blockchain identity authentication mechanisms to screen and authorize participants. Authenticated participants become "miners" in the blockchain network, with a subset selected as validation nodes participating in federated training under the blockchain consensus mechanism.
- Based on data distribution characteristics, scenarios in the Federated Modeling of Computational Processes stage include Horizontal Federated Training, Vertical Federated Training, and Federated Transfer Training. Participants selected for

ISSN (Online): 2583-2468
Volume-3 Issue-3 || May 2024 || PP. 34-39 DOI: 10.5281/zenodo.12643107

training use their local private data samples and reference the training tasks broadcast by smart contracts to conduct local model training. Upon completion, model parameters are encrypted, and updates are broadcasted via P2P networks and corresponding processing information.

- Compared to Horizontal Federated Training, Vertical Federated Training requires pre-training entity alignment, executing alignment algorithms and rules defined in smart contracts before training begins with the system distributing public keys.

Training nodes are divided into source and target domains in Federated Transfer Training. According to the secret sharing protocol defined in the smart contract, they jointly calculate the loss function. Each participant node then receives encrypted model parameters trained locally from training nodes before the preset timestamp and aggregates local model parameters based on aggregation rules defined in the smart contract.

2.3 The Infrastructure of BeFL

BeFL is infrastructure significantly differs from the traditional centralized federated learning model. Traditional federated learning models typically employ a star network topology, with a central server coordinating communication rounds, broadcasting current models to participants, collecting gradient updates from local calculations, and aggregating to generate next-generation models. However, there may be several problems with this centralized network topology:

1. Single Point of Failure and Performance Bottleneck: Once the central server fails, the entire network will be paralyzed, and as the number of participants increases, the central server may become a bottleneck in performance and communication.

2. Lack of Trust: Participants need to trust a central server, even if it can ensure the security and privacy of model parameter updates. They may be reluctant to share parameters due to lack of trust.

3. Data Security and Privacy: While a central server can ensure the security of data aggregation, participants' data may be at risk of leakage and attack.

In contrast, BeFL has introduced blockchain technology as part of its infrastructure to solve these problems in the traditional model. Blockchain technology is based on P2P networks, where each node is topologically equal, without centralized special nodes and hierarchies. Key blockchain technology features include:

- **Decentralization:** there is no single centralized node in the blockchain network; each node has the same power and status and undertakes network routing, verification, data transmission, and other functions, thus avoiding a single point of failure.
- **Security and Credibility:** The consensus mechanism ensures the security and immutability of data, effectively preventing malicious attacks and data tampering.
- **Privacy Protection:** Use encryption technology and smart contracts to protect the privacy of participants and ensure security during data exchange and sharing.
- **Incentives:** Automated incentives through smart contracts that encourage participants to contribute data and computing resources to improve sustainability and engagement in the system.

As a result, BeFL's infrastructure leverages the characteristics of blockchain technology to build a more secure, decentralized and trusted federated learning system suitable for processing large-scale data and multi-party scenarios, facilitating the secure sharing of data and value creation.

III. SMART CONTRACT INTEGRATION IN BEFL ARCHITECTURE

3.1 Functions and Advantages of Smart Contracts

Smart contracts operate autonomously on the blockchain and are computer programs that execute predefined logic and operations. They can manage the publication of federated learning tasks, validation of participants, training of local models, and aggregation of global models, and can implement incentives to promote participant contributions. In the BeFL architecture, integrating smart contracts makes the federated learning process more efficient and automated. Each participant interacts with smart contracts to perform tasks through preset algorithms and logic, ensuring data privacy and model security. Smart contracts can also monitor and regulate participants' behavior through the immutability of the blockchain and automatically implement rewards and punishments to maintain the fairness and impartiality of the system. Compared with traditional centralized management, the federated learning model based on smart contracts reduces the operating cost of the system and attracts more users to join the joint training.

The integration of smart contracts with artificial intelligence technology is a key feature of the BeFL architecture. By using a variety of machine learning algorithms and deep learning models, smart contracts can handle and optimize a wide variety of complex tasks. For example, they can be used to detect fraud on the blockchain, process unbalanced data sets such as train operation data, or optimize automatic modulation classification tasks for radio signals. The integration of deep learning algorithms enables smart contracts to achieve higher levels of pattern recognition and data processing, although further research is needed for deeper parametric model integration. In addition, the application of reinforcement learning algorithms can help smart contracts optimize the efficiency of the system's computing and communication resources, thereby improving the overall system's operating effect and performance.

ISSN (Online): 2583-2468
Volume-3 Issue-3 || May 2024 || PP. 34-39 DOI: 10.5281/zenodo.12643107

3.2 Development and Prospects of Smart Contracts

Smart contracts are more than just programs that execute static rules; they are evolving into core components with intelligent agent capabilities. Smart contracts can have high-level capabilities such as perception, reasoning, learning, and autonomous decision-making, which allows them to run and coordinate multiple tasks and operations in more complex environments. In the BeFL architecture, smart contracts serve as the basis for decentralized autonomous organizations (DAOs), capable of coordinating and managing the execution and evaluation of multiple federated learning tasks, while guaranteeing participants' interests and data privacy. The development potential of smart contracts also lies in their ability to handle complex social and collaborative scenarios, through the interaction and communication of intelligent agents, to solve various business problems and conflicts in real systems. This integration will promote the further integration of blockchain and artificial intelligence technology, and promote the development and application of distributed intelligent systems.

The application of smart contracts on blockchain also provides a platform for further exploration of social intelligence and swarm intelligence. Through smart contracts, different organizations and communities can work together to solve complex public problems and challenges while protecting data privacy. This federal learning model based on smart contracts is not only a technological advance, but also an innovation in social organization. In the future, with the further development and popularization of smart contract technology, BeFL architecture is expected to become an integrated platform for diverse intelligent components, promoting the application and evolution of distributed intelligent systems in various fields.

3.3 Application Cases and Research Results of Smart Contracts

The specific applications and research results of smart contracts in several fields show their potential and applicability. From Ponzi scheme detection to train operation data processing to classification optimization of radio signals, smart contracts have demonstrated their wide range of application scenarios through different machine learning and deep learning algorithms. In recent years, researchers have been exploring ways to integrate smart contracts with more complex AI techniques to handle a variety of real-world data challenges and task demands. The development of smart contracts has also prompted the further evolution of blockchain systems, from simple data storage and transmission, to higher levels of data processing and intelligent decision support. With the advancement of technology and the expansion of application scenarios, smart contracts are expected to become the core component of distributed autonomous organizations and intelligent systems in the future, and promote the in-depth application and development of artificial intelligence technology in society and business.

IV. CONCLUSION

With the rapid development of information and smart technologies, the volume of global data has increased dramatically, especially from iot devices. The importance of data has gone beyond traditional production factors to become a key factor driving the next generation of AI technologies such as deep learning. However, the collection and use of data has significant implications for personal and national security. To protect data security and privacy, countries have introduced strict data privacy protection regulations, such as the European Union's GDPR and China's Data Security Law, and Personal Information Protection Law. While these regulations protect the security and privacy of data, they also limit the flow and value creation of data to a certain extent, resulting in data fragmentation for security, privacy or geographic reasons. Therefore, how to ensure the flow and sharing of data and enhance the efficiency of inter-agency collaboration under the premise of protecting privacy and security has become an urgent issue in academia and industry.

Federated learning emerged as a new paradigm for distributed machine learning that allows institutions to jointly build global models without sharing individual data or disclosing the underlying raw data or its encrypted form. By iteratively exchanging and updating encryption parameters, federated learning effectively reduces the risk of privacy disclosure and data compliance, which has received extensive and in-depth research attention. This paper reviews the development process, methods, and applications of federated learning and discusses its application in different scenarios based on the distribution characteristics of participant data. In the future, with the deep integration of blockchain technology and federated learning, such as the BeFL architecture, it is expected to become a more secure, decentralized and trusted federated learning system for handling large-scale data and securely sharing data and value creation in multi-party scenarios.

REFERENCES

1. Song, J., Cheng, Q., Bai, X., Jiang, W., & Su, G. (2024). LSTM-based deep learning model for financial market stock price prediction. *Journal of Economic Theory and Business Management*, *1*(2), 43-50.
2. Ni, C., Zhang, C., Lu, W., Wang, H., & Wu, J. (2024). *Enabling intelligent decision making and optimization in enterprises through data pipelines*.

3. Lu, W., Ni, C., Wang, H., Wu, J., & Zhang, C. (2024). *Machine learning-based automatic fault diagnosis method for operating systems*.
4. Liang, P., Song, B., Zhan, X., Chen, Z., & Yuan, J. (2024). Automating the training and deployment of models in MLOps by integrating systems with machine learning. *Appl. Comput. Eng., 67*, 1–7. https://doi.org/10.54254/2755-2721/67/20240690.
5. Zhou, Y., Zhan, T., Wu, Y., Song, B., & Shi, C. (2024). RNA secondary structure prediction using transformer-based deep learning models. *Appl. Comput. Eng., 64*, 95–101. https://doi.org/10.54254/2755-2721/64/20241362.
6. Liu, B., Cai, G., Ling, Z., Qian, J., & Zhang, Q. *Precise positioning and prediction system for autonomous driving based on generative artificial intelligence*.
7. Cui, Z., Lin, L., Zong, Y., Chen, Y., & Wang, S. *Precision gene editing using deep learning: a case study of the crispr-cas9 editor*.
8. Wang, B., He, Y., Shui, Z., Xin, Q., & Lei, H. (2024). Predictive optimization of DDoS attack mitigation in distributed systems using machine learning. *Applied and Computational Engineering, 64*, 95-100.
9. Xiao, J., Wang, J., Bao, W., Deng, T., & Bi, S. *Application progress of natural language processing technology in financial research*.
10. Wang, Yong, et al. (2024). Machine learning-based facial recognition for financial fraud prevention. *Journal of Computer Technology and Applied Mathematics 1*(1), 77-84.
11. Song, Jintong, et al. (2024). LSTM-based deep learning model for financial market stock price prediction. *Journal of Economic Theory and Business Management, 1*(2): 43-50.
12. Bai, Xinzhu, Wei Jiang, & Jiahao Xu. (2024). Development trends in AI-based financial risk monitoring technologies. *Journal of Economic Theory and Business Management 1*(2), 58-63.
13. Tian, J., Li, H., Qi, Y., Wang, X., & Feng, Y. (2024). Intelligent medical detection and diagnosis assisted by deep learning. *Appl. Comput. Eng., 64*, 116–121. https://doi.org/10.54254/2755-2721/64/20241356.
14. Shi, Y., Li, L., Li, H., Li, A., & Lin, Y. (2024). Aspect-level sentiment analysis of customer reviews based on neural multi-task learning. *Journal of Theory and Practice of Engineering Science, 4*(04), 1-8.
15. Li, Zihan, et al. (2024). *Robot navigation and map construction based on SLAM technology*.
16. Fan, C., Ding, W., Qian, K., Tan, H., & Li, Z. (2024). Cueing flight object trajectory and safety prediction based on SLAM technology. *Journal of Theory and Practice of Engineering Science, 4*(05), 1-8.
17. Ding, W., Tan, H., Zhou, H., Li, Z., & Fan, C. (2024). Immediate traffic flow monitoring and management based on multimodal data in cloud computing. *Appl. Comput. Eng., 71*, 1–6. https://doi.org/10.54254/2755-2721/71/2024ma0052.
18. Qian, K., Fan, C., Li, Z., Zhou, H., & Ding, W. (2024). Implementation of artificial intelligence in investment decision-making in the chinese a-share market. *Journal of Economic Theory and Business Management, 1*(2), 36-42.
19. Shi, Y., Yuan, J., Yang, P., Wang, Y., & Chen, Z. (2024). Implementing intelligent predictive models for patient disease risk in cloud data warehousing. *Appl. Comput. Eng., 67*, 34–40. https://doi.org/10.54254/2755-2721/67/2024ma0059.
20. Zhan, T., Shi, C., Shi, Y., Li, H., & Lin, Y. (2024). Optimization techniques for sentiment analysis based on LLM (GPT-3). *Appl. Comput. Eng., 67*, 41–47. https://doi.org/10.54254/2755-2721/67/2024ma0060.
21. Li, Huixiang, et al. (2024). AI face recognition and processing technology based on GPU computing. *Journal of Theory and Practice of Engineering Science, 4*(05), 9-16.
22. Bi, Shuochen, Wenqing Bao, Jue Xiao, Jiangshan Wang, & Tingting Deng. (2024). *Application and practice of AI technology in quantitative investment*. arXiv preprint arXiv:2404.18184(2024).
23. Yuan, J., Lin, Y., Shi, Y., Yang, T., & Li, A. (2024). Applications of artificial intelligence generative adversarial techniques in the financial sector. *Academic Journal of Sociology and Management, 2*(3), 59-66.
24. Yu, D., Xie, Y., An, W., Li, Z., & Yao, Y. (2023, December). Joint coordinate regression and association for multi-person pose estimation, A pure neural network approach. in *Proceedings of the 5th ACM International Conference on Multimedia in Asia*, pp. 1-8.
25. Lin, Y., Li, A., Li, H., Shi, Y., & Zhan, X. (2024). GPU-Optimized image processing and generation based on deep learning and computer vision. *Journal of Artificial Intelligence General science (JAIGS), 5*(1), 39-49.
26. Chen, Zhou, et al. (2024). Application of cloud-driven intelligent medical imaging analysis in disease detection. *Journal of Theory and Practice of Engineering Science 4*(05), 64-71.
27. Wang, B., Lei, H., Shui, Z., Chen, Z., & Yang, P. (2024). *Current state of autonomous driving applications based on distributed perception and decision-making*.
28. Zhan, X., Shi, C., Li, L., Xu, K., & Zheng, H. (2024). Aspect category sentiment analysis based on multiple attention mechanisms and pre-trained models. *Applied and Computational Engineering, 71*, 21-26.

ISSN (Online): 2583-2468
Volume-3 Issue-3 || May 2024 || PP. 34-39 DOI: 10.5281/zenodo.12643107

29. Wu, B., Xu, J., Zhang, Y., Liu, B., Gong, Y., & Huang, J. (2024). Integration of computer networks and artificial neural networks for an AI-based network operator. *Applied and Computational Engineering*, *64*, 115-120.
30. Liang, P., Song, B., Zhan, X., Chen, Z., & Yuan, J. (2024). Automating the training and deployment of models in MLOps by integrating systems with machine learning. *Appl. Comput. Eng.*, *67*, 1–7. Available at: https://doi.org/10.54254/2755-2721/67/20240690.
31. Li, A., Yang, T., Zhan, X., Shi, Y., & Li, H. (2024). Utilizing data science and AI for customer churn prediction in marketing. *Journal of Theory and Practice of Engineering Science*, *4*(05), 72-79.
32. Wu, B., Gong, Y., Zheng, H., Zhang, Y., Huang, J., & Xu, J. (2024). Enterprise cloud resource optimization and management based on cloud operations. *Applied and Computational Engineering*, *67*, 8-14.
33. Xu, J., Wu, B., Huang, J., Gong, Y., Zhang, Y., & Liu, B. (2024). Practical applications of advanced cloud services and generative AI systems in medical image analysis. *Appl. Comput. Eng.*, *64*, 83–88. https://doi.org/10.54254/2755-2721/64/20241361.
34. Zhang, Y., Liu, B., Gong, Y., Huang, J., Xu, J., & Wan, W. (2024). Application of machine learning optimization in cloud computing resource scheduling and management. *Appl. Comput. Eng.*, *64*, 17–22. https://doi.org/10.54254/2755-2721/64/20241359.
35. Huang, J., Zhang, Y., Xu, J., Wu, B., Liu, B., & Gong, Y. (2024). Implementation of seamless assistance with Google Assistant leveraging cloud computing. *Appl. Comput. Eng.*, *64*, 170–176. https://doi.org/10.54254/2755-2721/64/20241383.