WORLD COMET RESEARCH GROUP

International Journal of Data Engineering and Intelligent Computing
Vol. 5, No. 1, 2022
Doi: 10.67228/30715717/IJDEIC-2022PI2D4M
PP. 01-12

Original Article

Federated Learning and Blockchain for Secure Edge Computing: Opportunities and Challenges

Dr. Lakshmi Narayanan[1], Alan Turing[2]

[1]*Associate Professor, Bharathidasan University, India.*

[2]*Mathematician & Computer Scientist, University of Manchester, United Kingdom.*

Received: 01-03-2022 Revised: 01-22-2022 Accepted: 02-04-2022 Published: 02-13-2022

ABSTRACT

The convergence of Federated Learning (FL), Blockchain, and Edge Computing presents a transformative paradigm for decentralized, secure, and privacy-preserving machine learning at the network edge. FL enables collaborative model training without centralizing data, while Blockchain provides immutable and transparent mechanisms for trust, accountability, and coordination among distributed edge nodes. Edge computing further enhances this ecosystem by offering low-latency computation near data sources. Despite the promise of this triad, significant challenges persist in terms of scalability, energy efficiency, consensus mechanisms, data and model security, and system heterogeneity. This paper provides a comprehensive survey of the intersection of FL, Blockchain, and Edge Computing, analyzing key opportunities, current solutions, and open challenges. We also discuss architectural frameworks, real-world applications, and future research directions.

KEYWORDS

Federated Learning; Blockchain; Edge Computing; Privacy-Preserving Machine Learning; Decentralized AI; Consensus Mechanisms; Smart Contracts; Distributed Systems; Data Security; Resource-Constrained Devices.

1. INTRODUCTION

1.1. Motivation

The explosion of data generated at the edge of networks through smartphones, IoT devices, autonomous systems, and industrial sensors has created a new frontier for artificial intelligence (AI). However, traditional centralized machine learning approaches, which rely on aggregating data to a central server, are increasingly impractical due to privacy concerns, bandwidth limitations, and latency requirements. In this context, Federated Learning (FL), Blockchain, and Edge Computing have emerged as promising technologies that align with the growing demand for decentralized, secure, and efficient AI systems. The motivation behind this research is to explore how the combination of these technologies can address the limitations of centralized AI systems while unlocking new opportunities for privacy-preserving and trustable AI on edge devices.

1.2. Importance of Privacy and Decentralization in AI

In today's data-driven world, privacy and data ownership have become critical concerns, especially in sensitive domains like healthcare, finance, and smart surveillance. Traditional centralized AI frameworks not only pose privacy risks but also create single points of failure that are vulnerable to cyberattacks and misuse. Decentralization is essential for democratizing access to AI, preserving user data on local devices, and ensuring autonomy among stakeholders. Federated Learning supports data privacy by training models locally on users' devices without transferring raw data. Blockchain adds an immutable, distributed layer of trust, while Edge Computing brings computation closer to data sources, reducing reliance on cloud infrastructure. Together, these technologies support a paradigm shift toward decentralized AI.

1.3. Overview of Federated Learning, Blockchain, and Edge Computing

Federated Learning is a machine learning paradigm that enables multiple edge devices to collaboratively train a shared model without exchanging their local datasets. Each participant computes updates locally and sends only the model parameters or gradients to a central aggregator or peer network. Blockchain, on the other hand, is a decentralized ledger technology that allows for secure, transparent, and tamper-proof recording of transactions. It operates through consensus mechanisms to validate and store data across multiple nodes. Edge Computing refers to processing data closer to the source of generation—at or near the edge of the network—which improves latency, reduces bandwidth consumption, and enables real-time decision-making. These three domains complement each other in facilitating a secure, efficient, and decentralized AI ecosystem.

1.4. Scope and Contributions of the Paper

This paper investigates the integration of Federated Learning, Blockchain, and Edge Computing, with a focus on identifying the synergies and challenges that arise from their convergence. The scope includes a technical overview of each individual domain, a detailed analysis of their integration, and an

ISSN: 3071-5717
PP: 01-12.

evaluation of the opportunities they present for real-world applications. The paper also explores open research issues, architectural frameworks, and potential use cases. The primary contribution is a comprehensive survey of the landscape that highlights how these technologies can collectively enable privacy-preserving, secure, and efficient decentralized AI systems, along with a roadmap for future research.

Table 1: Role of Federated Learning and Blockchain in Edge Computing

Component	Federated Learning (FL)	Blockchain (BC)	Combined Impact in Edge Computing
Core Function	Distributed model training without raw data sharing	Decentralized, tamper-proof ledger	Secure and privacy-preserving collaborative intelligence
Data Handling	Local data remains at edge devices	Transactions stored immutably	Eliminates centralized data leakage risks
Trust Model	Relies on participating nodes	Trustless consensus mechanism	Enhances trust among untrusted edge nodes
Scalability	High with multiple edge participants	Limited by consensus overhead	Requires optimization strategies

2. BACKGROUND AND PRELIMINARIES

2.1. Federated Learning

2.1.1. Concept and Workflow

Federated Learning (FL) is designed to train machine learning models across multiple decentralized devices holding local data samples, without the need to exchange this data. The typical FL workflow involves three main stages: local training, model aggregation, and distribution. Each participating edge node downloads the current global model, trains it on its local dataset, and then uploads the computed gradients or updated weights to a central server or a distributed aggregator. These updates are then aggregated—typically using algorithms like Federated Averaging—to form a new global model, which is redistributed for the next training round. This cycle continues until the model converges. FL thus ensures that data remains on-device, addressing privacy and data sovereignty concerns.

2.1.2. Challenges in FL (e.g., communication, heterogeneity)

Despite its benefits, FL introduces several challenges. Communication overhead is a significant issue, especially in resource-constrained edge environments, as frequent transmission of model updates can congest the network. Additionally, the non-IID (independent and identically distributed) nature of data across devices leads to heterogeneous data distributions, which can degrade model performance. Variations in hardware capabilities and connectivity further complicate training coordination and synchronization. Ensuring robustness against adversarial participants and maintaining model security also remain open challenges.

2.2. Blockchain Technology

2.2.1. Distributed Ledger Basics

Blockchain is a peer-to-peer distributed ledger that records transactions in a series of linked blocks. Each block contains a cryptographic hash of the previous block, a timestamp, and transaction data. Once a block is validated and added to the chain, its contents are immutable and publicly verifiable. This structure ensures transparency, data integrity, and resistance to tampering. Blockchain operates in a decentralized fashion, with no central authority, relying instead on consensus among nodes to validate and agree on the current state of the ledger.

2.2.2. Consensus Algorithms (PoW, PoS, PBFT, etc.)

To maintain the integrity of the blockchain, consensus mechanisms are used to determine the validity of transactions. Proof of Work (PoW), used by Bitcoin, requires nodes to solve computationally intensive puzzles, making it secure but energy-intensive. Proof of Stake (PoS) selects validators based on their stake in the system, offering a more energy-efficient alternative. Practical Byzantine Fault Tolerance (PBFT) is designed for permissioned networks and achieves fast consensus through majority agreement among trusted nodes. Each consensus mechanism offers trade-offs between security, scalability, and resource consumption, which must be carefully considered in edge environments.

2.3. Edge Computing

2.3.1. Definition and Characteristics

Edge Computing is a distributed computing paradigm that brings computation and data storage closer to the data sources, such as sensors, mobile phones, and IoT devices. This reduces the latency associated with data transmission to centralized cloud servers and supports real-time analytics. Key characteristics of edge computing include locality, low-latency response, scalability across heterogeneous devices, and the ability to operate under intermittent connectivity.

2.4. Constraints and Benefits

While edge computing offers benefits like faster response times, bandwidth savings, and improved privacy, it is constrained by limited computational power, storage, and energy resources. These constraints necessitate the development of lightweight and efficient algorithms. However, by processing data locally and selectively communicating with cloud or other edge nodes, edge computing supports scalable and resilient distributed AI systems that can operate even in environments with unreliable connectivity.

ISSN: 3071-5717
PP: 01-12.

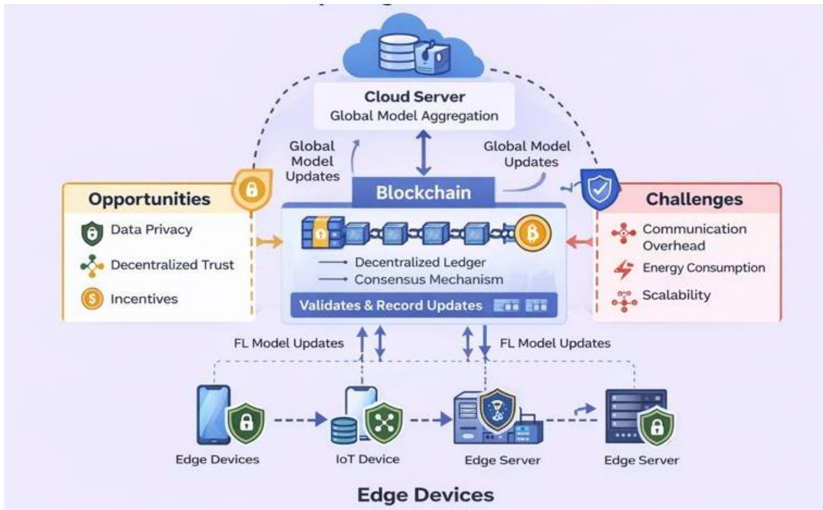

Fig 1 - Architecture of Federated Learning and Blockchain-Enabled Edge Computing

3. SYNERGY BETWEEN FL, BLOCKCHAIN, AND EDGE COMPUTING

3.1. Why Combine These Three Technologies

The integration of Federated Learning, Blockchain, and Edge Computing offers a powerful solution for building decentralized, privacy-aware, and trustworthy AI systems. FL addresses privacy by keeping data local, Blockchain introduces verifiability and security through decentralized consensus, and Edge Computing brings computational power closer to data sources, enabling low-latency decision-making. When combined, these technologies overcome the limitations of centralization, enhance resilience, and enable autonomous operation across a network of distributed devices. This synergy is especially valuable in domains where data privacy, security, and real-time responsiveness are critical.

3.2. Complementary Features and Mutual Benefits

Each technology complements the others in meaningful ways. Federated Learning relies on coordination among nodes, which Blockchain can facilitate via smart contracts and trustless consensus. Blockchain can also log all training updates and transactions, ensuring transparency and accountability. Meanwhile, Edge Computing provides the physical infrastructure where FL models are trained and inference is performed. It reduces communication delays and offloads processing from the cloud, which is particularly important when dealing with time-sensitive applications. Together, these technologies create a decentralized ecosystem where intelligence, trust, and efficiency co-exist.

3.3. High-Level Architecture of Integrated Systems

A typical architecture combining FL, Blockchain, and Edge Computing involves edge devices (such as smartphones, sensors, or drones) locally training FL models and interacting with a Blockchain network to share model updates. The Blockchain maintains a ledger of all contributions, manages smart contracts for participant incentives, and ensures consensus on model integrity. An aggregator

ISSN: 3071-5717
PP: 01-12.

centralized or decentralized—collects model updates for aggregation. This setup may be organized in a hierarchical structure, where clusters of edge nodes are coordinated through intermediate nodes or fog servers. The architecture balances computational load, network efficiency, and system trust.

4. OPPORTUNITIES

4.1. Enhanced Privacy and Security

One of the primary advantages of combining these technologies is the ability to achieve strong privacy and security guarantees. Federated Learning enables data to stay on local devices, reducing the risk of data breaches during transmission. Blockchain ensures secure coordination among untrusted parties through cryptographic mechanisms and consensus protocols. By recording all model updates and training transactions on a tamper-proof ledger, Blockchain also protects against malicious actors trying to manipulate the learning process. Together, they establish an environment where private data is protected, and the integrity of the model training process is verifiable.

4.2. Trust and Transparency

Trust is a critical component in collaborative AI systems, especially when the participants are distributed and may not trust each other. Blockchain introduces transparency through its immutable and auditable record of events, allowing any participant to verify the sequence and validity of training operations. Smart contracts can enforce rules automatically, eliminating the need for intermediaries. This trustless collaboration model enables multiple stakeholders—from different organizations or jurisdictions—to jointly train models without fear of misbehavior or manipulation. It is especially valuable in sectors like healthcare, where cross-institutional data sharing is sensitive.

4.3. Incentive Mechanisms

For FL systems to scale effectively there must be incentives for participants to contribute computational resources and quality data. Blockchain facilitates this through the implementation of smart contracts and token-based reward systems. Participants who contribute valid model updates can be compensated automatically with digital tokens or credits. These mechanisms promote sustained participation and fairness in the training process. Furthermore, tokens can be used to rank contributors or prioritize model updates from more reliable nodes, adding another layer of quality control to the system.

4.4. Scalability and Low Latency

Edge Computing enables computation and data processing to happen close to where data is generated, which dramatically reduces latency and improves scalability. In an integrated system, model training, inference, and even Blockchain validation can be distributed across edge nodes, minimizing reliance on centralized infrastructure. This decentralization supports massive scaling across thousands or even millions of devices. Furthermore, latency-sensitive applications such as autonomous vehicles

ISSN: 3071-5717
PP: 01-12.

and smart grids benefit greatly from the responsiveness enabled by edge computing, while maintaining model integrity and data security through FL and Blockchain.

Table 2: Privacy and Security Enhancements through FL and Blockchain

Aspect	Federated Learning Contribution	Blockchain Contribution	Combined Benefit
Data Privacy	Keeps raw data on local devices	No centralized data storage	Prevents data leakage and unauthorized access
Security	Limits data exposure during training	Cryptographic hashing and consensus	Strong end-to-end security guarantees
Attack Resistance	Reduces centralized attack surface	Tamper-proof distributed ledger	Protection against model poisoning and manipulation
Integrity Verification	Local model updates only	Immutable record of updates	Verifiable and trustworthy training process

5. CHALLENGES

5.1. Resource Constraints at the Edge

Edge devices, such as smartphones, IoT sensors, and embedded systems, often operate under significant resource constraints. These limitations encompass processing power, memory capacity, storage, and energy supply. Federated Learning (FL) necessitates local computation for model training, which can be computationally intensive and memory-demanding. Additionally, the frequent communication of model updates to a central server or aggregator can strain network bandwidth, especially in environments with limited connectivity. The heterogeneity of edge devices further exacerbates these challenges, as devices with varying capabilities must collaborate in the learning process. Addressing these constraints requires the development of lightweight models, efficient communication protocols, and adaptive algorithms that can function effectively within the limitations of edge devices.

5.2. Scalability of Blockchain

While Blockchain offers a decentralized and secure framework for managing model updates and ensuring data integrity in FL systems, its scalability remains a significant challenge. Traditional consensus mechanisms, such as Proof of Work (PoW), can be resource-intensive and may not be suitable for the dynamic and resource-constrained environments of edge computing. As the number of participating devices increases, the blockchain network may experience bottlenecks in transaction throughput and latency, potentially hindering the overall performance of the FL system. To address these issues, alternative consensus algorithms, such as Proof of Stake (PoS) or Practical Byzantine Fault Tolerance (PBFT), and layer-two solutions like sidechains or sharding are being explored to enhance the scalability of blockchain in FL applications.

ISSN: 3071-5717
PP: 01-12.

5.3. Data and Model Security

The integration of FL and Blockchain introduces several security concerns that must be addressed to ensure the integrity and confidentiality of the learning process. Data poisoning attacks, where malicious participants inject erroneous data to degrade model performance, pose a significant threat. Model inversion attacks can potentially reconstruct sensitive training data from the shared model parameters, compromising user privacy. Sybil attacks, where an adversary creates multiple fake identities to gain disproportionate influence over the system, can undermine the trustworthiness of the FL process. Implementing robust security measures, such as secure aggregation protocols, homomorphic encryption, and anomaly detection mechanisms, is crucial to mitigate these risks and maintain the reliability of the FL system.

5.4. System Heterogeneity

The diversity in edge devices and the non-Independent and Identically Distributed (non-IID) nature of data present substantial challenges in FL systems. Devices with varying computational capabilities, storage capacities, and network conditions can lead to uneven model performance and convergence rates. Additionally, the data generated by different devices may have distinct distributions, which can affect the generalization ability of the global model. To address these challenges, techniques such as personalized federated learning, clustering of devices with similar data characteristics, and weighted aggregation methods are being explored to accommodate the heterogeneity and ensure effective learning across diverse devices.

5.5. Energy Efficiency

Energy consumption is a critical concern in edge computing environments, especially for battery-powered devices. The computational tasks involved in FL, including model training and communication of updates, can lead to significant energy expenditure. Similarly, the operation of blockchain networks, particularly those utilizing energy-intensive consensus mechanisms, can further exacerbate energy consumption. Developing energy-efficient algorithms, optimizing communication protocols to reduce data transmission and implementing energy-aware scheduling are essential strategies to minimize energy usage and prolong the operational life of edge devices.

6. ARCHITECTURAL FRAMEWORKS AND MODELS

6.1. Review of Existing Solutions

Existing architectural frameworks for integrating FL, Blockchain, and Edge Computing have primarily focused on enhancing privacy, security, and scalability. These solutions often employ a hierarchical structure, where edge devices perform local model training and communicate updates through a blockchain network to a central aggregator. Blockchain serves to ensure the integrity and traceability of model updates, while FL facilitates collaborative learning without sharing raw data.

ISSN: 3071-5717
PP: 01-12.

However, these architectures face challenges related to scalability, energy efficiency, and the heterogeneity of devices.

6.2. Proposed Layered or Modular Architectures

To address the limitations of existing solutions, researchers propose layered or modular architectures that decouple various components to enhance flexibility and scalability. In these architectures, the system is divided into distinct layers, such as the data collection layer, processing layer, blockchain layer, and application layer. Each layer can be independently scaled or optimized, allowing for more efficient resource management and better adaptation to the dynamic conditions of edge environments. Modular designs also facilitate the integration of heterogeneous devices and enable the deployment of customized solutions tailored to specific application requirements.

6.3. Communication Protocols and Smart Contract Usage

Efficient communication protocols are essential to reduce the overhead associated with model updates and blockchain transactions. Techniques such as model compression, sparse updates, and differential privacy are employed to minimize the amount of data transmitted between devices and the blockchain network. Smart contracts play a pivotal role in automating processes within the system, such as validating model updates, enforcing participation incentives, and managing access control. By embedding logic directly into the blockchain, smart contracts ensure transparency and trust in the decentralized FL process.

7. USE CASES AND APPLICATIONS

7.1. Smart Cities

In smart city applications, FL and Blockchain can be utilized to process data from various sources, such as traffic sensors, surveillance cameras, and environmental monitors, without compromising privacy. For instance, FL can enable traffic prediction models to be trained on data from individual vehicles or traffic signals, while Blockchain ensures the integrity and accountability of the data and model updates. This approach facilitates real-time decision-making for urban management, such as dynamic traffic control and resource allocation.

7.2. Industrial IoT

In industrial settings, FL can be applied to predictive maintenance, quality control, and process optimization by analyzing data from machines and sensors. Blockchain provides a secure and transparent ledger for recording sensor readings, maintenance logs, and model updates, ensuring data authenticity and compliance with industry standards. The integration of FL and Blockchain in Industrial IoT systems enhances operational efficiency, reduces downtime, and improves safety.

ISSN: 3071-5717
PP: 01-12.

7.3. Healthcare

Healthcare applications benefit from the privacy-preserving capabilities of FL, allowing for collaborative learning on sensitive patient data without sharing raw information. Blockchain can be used to maintain a secure and immutable record of patient consent, model updates, and access logs, ensuring compliance with regulations such as HIPAA or GDPR. This integration enables the development of personalized healthcare solutions, such as early disease detection and treatment recommendation systems, while safeguarding patient privacy.

7.4. Autonomous Vehicles

Autonomous vehicles generate vast amounts of data from sensors and cameras, which can be utilized for training machine learning models to improve navigation and decision-making. FL allows vehicles to collaboratively train models on their local data, preserving privacy and reducing the need for data transmission. Blockchain ensures the integrity of the model updates and provides a decentralized framework for managing vehicle interactions and data sharing, enhancing trust and coordination in autonomous vehicle networks.

8. OPEN RESEARCH DIRECTIONS

8.1. Lightweight Consensus Algorithms

Developing lightweight consensus algorithms tailored for edge environments is crucial to enhance the scalability and efficiency of blockchain networks in FL systems. These algorithms should minimize computational overhead and energy consumption while maintaining security and decentralization.

8.2. Energy-Efficient FL Models

Designing energy-efficient FL models involves optimizing the computational complexity and communication requirements of the learning process. Techniques such as model pruning, quantization, and federated transfer learning can be employed to reduce energy consumption without significantly compromising model performance.

8.3. Secure Aggregation and Encrypted Model Updates

One of the key challenges in Federated Learning is ensuring the confidentiality of individual model updates during aggregation. Secure aggregation protocols allow the server or aggregator to compute a global model without learning the individual contributions of participants. Current research focuses on cryptographic techniques such as homomorphic encryption, differential privacy, and secure multi-party computation (SMPC) to protect updates while still enabling efficient aggregation. The integration of Blockchain can complement this by recording hashes or encrypted summaries of model updates, ensuring that tampering can be detected. Developing lightweight, scalable protocols that balance security, computational overhead, and bandwidth consumption remains an open research frontier.

8.4. Real-time Trust Evaluation Mechanisms

In decentralized FL systems, ensuring the trustworthiness of participants is critical, especially in environments where nodes can join or leave dynamically. Blockchain can store historical performance records of devices, but additional real-time mechanisms are needed to detect and isolate malicious or underperforming nodes. Trust models based on reputation scores, behavior monitoring, and anomaly detection algorithms are being explored to assess the reliability of edge devices. Smart contracts can also be programmed to penalize untrustworthy behavior automatically. Research is needed to develop adaptive and context-aware trust evaluation frameworks that can work in tandem with Blockchain and FL systems in real-time.

9. CONCLUSION

This paper has explored the powerful intersection of Federated Learning, Blockchain, and Edge Computing, highlighting how their integration addresses critical challenges in decentralized AI systems. Federated Learning ensures data privacy by keeping data local to devices, Blockchain adds a layer of trust, transparency, and security through decentralized consensus and smart contracts, and Edge Computing enables low-latency processing near data sources. While this synergy opens up promising opportunities in domains such as healthcare, smart cities, industrial IoT, and autonomous vehicles, several technical challenges must still be overcome. These include resource limitations at the edge, scalability issues in blockchain consensus, and vulnerabilities to adversarial attacks, system heterogeneity, and energy efficiency concerns. We have also identified promising research directions that can drive future innovation, including the development of lightweight consensus protocols, energy-efficient FL algorithms, secure model aggregation techniques, and real-time trust evaluation mechanisms. As these technologies continue to mature, their integration will play a foundational role in shaping the future of decentralized, privacy-preserving, and trustworthy AI at the edge.

REFERENCES

[1] Pokhrel, S. R. (2020). Federated learning meets blockchain at 6G edge: A drone-assisted networking for disaster response. *Proceedings of the 2nd ACM MobiCom Workshop on Drone Assisted Wireless Communications for 5G and Beyond*, 49–54. https://doi.org/10.1145/3414045.3415949

[2] Lu, Y., Huang, X., Zhang, K., Maharjan, S., & Zhang, Y. (2020). Low-latency federated learning and blockchain for edge association in digital twin empowered 6G networks. *IEEE Transactions on Industrial Informatics, 17*(7), 5098–5107. https://doi.org/10.1109/TII.2020.3034184

[3] Kim, M., Lim, W., Yu, Y., Kim, Y., Kang, S., & Han, Z. (2019). Blockchain and AI-based secure data sharing platform for federated learning in edge networks. *IEEE Access, 7*, 128995–129005. https://doi.org/10.1109/ACCESS.2019.2931007

[4] Lu, Y., Huang, X., Dai, Y., Maharjan, S., & Zhang, Y. (2020). Blockchain and federated learning for privacy-preserved data sharing in industrial IoT. *IEEE Transactions on Industrial Informatics, 16*(6), 4177–4186. https://doi.org/10.1109/TII.2019.2942190

[5] Nguyen, D. C., Pathirana, P. N., Ding, M., & Seneviratne, A. (2020). Blockchain for secure EHRs sharing of mobile cloud based e-health systems. *IEEE Access, 7*, 66792–66806. https://doi.org/10.1109/ACCESS.2019.2917555

[6] Li, Y., Zheng, Z., Hu, C., Deng, Y., & Zhou, R. (2020). A survey on blockchain-based federated learning: Applications and challenges. *IEEE Internet of Things Journal, 7*(12), 11220–11235. https://doi.org/10.1109/JIOT.2020.3015745

[7] Kang, J., Xiong, Z., Niyato, D., Yu, H., Liang, Y.-C., & Kim, D. I. (2019). Incentive mechanism for reliable federated learning: A joint optimization approach to combining reputation and contract theory. *IEEE Internet of Things Journal, 6*(6), 10700–10714. https://doi.org/10.1109/JIOT.2019.2940820

[8] Lu, Y., Huang, X., Zhang, K., Maharjan, S., & Zhang, Y. (2020). Communication-efficient federated learning and permissioned blockchain for digital twin edge networks. *IEEE Internet of Things Journal, 8*(4), 2276–2288. https://doi.org/10.1109/JIOT.2020.3026790

[9] Xiong, Z., Feng, S., Niyato, D., Wang, P., & Han, Z. (2020). Optimal pricing-based edge computing resource management in mobile blockchain. *IEEE Transactions on Services Computing, 13*(3), 387–399. https://doi.org/10.1109/TSC.2019.2962004

[10] Nguyen, D. C., Ding, M., Pathirana, P. N., Seneviratne, A., Li, J., Niyato, D., & Poor, H. V. (2020). Federated learning for Internet of Things: A comprehensive survey. *IEEE Communications Surveys & Tutorials, 23*(3), 1622–1658. https://doi.org/10.1109/COMST.2021.3062606