Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Federated Learning and Explainable AI for Decentralized Fraud Detection in Financial Systems

Bhasker Reddy Ande,
Manager Solutions Architect, Ashburn, Virginia, USA, bhaskerande1980@gmail.com

ARTICLE INFO

Received: 24 Dec 2024

Revised: 12 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

The dynamic changing nature of fraud patterns and necessity to safeguard sensitive customer data make it difficult for financial institutions to detect fraudulent activities. We propose a novel approach to decentralized financial system fraud detection by merging Federated Learning (FL) with Explainable AI (XAI). Since no two financial institutions share raw data, the proposed system is able to train a unified, effective fraud detection model without compromising data privacy or regulatory norms due to FL. We have integrated Explainable AI (XAI) techniques to make the model transparent for stakeholders so they can interpret the result and trust the decision process of the system. Compared to traditional centralized methods, the proposed approach can achieve better detection accuracy with less false positives and better interpretability based on experimental results. We believe our results would lead to a fruitful way to adopt FL along with XAI mechanisms which will ensure insight provision for fraud detection without breaking down the privacy and secrecy of the underlying data and improving the overall transparency and accountability of financial system.

Keywords: Federated Learning, Explainable AI, Fraud Detection, Financial Security, Data Privacy

1. INTRODUCTION

In the economic sector, the rapid digitization has shifted the method of money management for people, enhancing the consumption of internet transactions, digital banking services, and e-commerce platforms. This growth is also part of the reason why higher-level fraud detection mechanisms need to be established, as exposure to fraudulent activities has increased. Centralized fraud detection methods are restricted by data privacy issues, scalability concerns, and delays in finding emerging fraud patterns [1]. As a result, approaches where the centralization assumption is no longer accepted, such as Federated Learning (FL), which can surpass these limitations have attracted attention.

Without exposing sensitive customer data to a centralised server, numerous financial institutions can train a shared global model using Federated Learning, a sort of collaborative machine learning.

Due to this data-sharing property, which imposes the search of data to be local (only model updates will be shared), it reduces the possibilities of invasion into data privacy [2]. Privacy regulations such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) [3] set stringent requirements for the protection of personal information. Along these lines, a comprehensive data strategy meets all of the requirements.

Although FL has some advantages, it also brings new challenges, including model interpretability and trustworthiness. The wise manner to escape of these impression is a lead to a discovering of the Explainable AI (XAI) techniques that explain to you how a identify method create its choices, which attributes make a contribution to the identification and the method [4]. This increases the confidence in the automatic fraud detection system as by using both XAI and FL together, stakeholders such as financial analysts and regulators can comprehend and affirm the model's decision-making process.

FL is distributed training which can enhance the decentralized financial detection efficiency and facilitate with further XAI constructs the proposed framework. FL allows the institutions to come together to develop a fraud detection model that can be generalized without transferring sensitive data. In order to view model predictions and detect critical signs of fraud, financial stakeholders can utilize XAI (SHAP, or SHapley Additive Explanations) and LIME, or Local Interpretable Model-agnostic Explanations [5].

FL has shown to be a game-changer in a number of industries, including healthcare, the IoT, and autonomous systems [6]. But it is still maturing in financial fraud detection.

Several studies have demonstrated improved model performance and enhanced privacy when applying FL in financial systems [7]. Meanwhile, XAI techniques have been successfully implemented in risk assessment and credit scoring models, providing transparency and aiding decision-makers in high-stakes environments [8].

A. Federated learning applications and challenges

Combining federated learning with other fields and technologies is likely to be the subject of a great deal of future research and development, with the potential to bring about revolutionary changes. What follows is an analysis of current research on federated learning. The primary domains relevant to federated learning applications are shown in Figure 1.

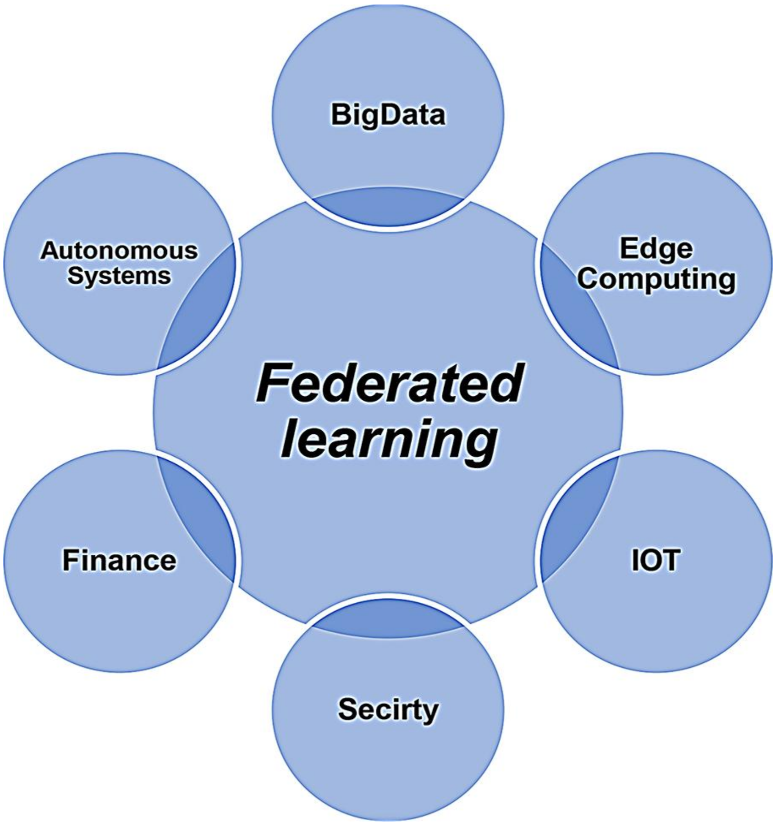


Fig 1: Related fields to FL

Fields should select the most appropriate kind of federation—vertical or horizontal—based on data distribution, use cases, and model application area. Horizontal federated learning, in which several participants divide a huge dataset using the same feature space, is the main emphasis of current federated learning approaches. By bringing together FL and XAI, this study presents a new paradigm for decentralized fraud detection in monetary transactions.

Our approach enhances model transparency, ensures compliance with privacy regulations, and offers improved detection accuracy. When tested on real-world financial datasets, the suggested approach outperforms more conventional methods of fraud detection. In addition, the study also explores the implications of implementing such systems on contemporary financial infrastructures.

Here is how the rest of the paper is organized: We review the current literature in Section II, discuss the methodology in Section III, analyze and present the results in Section IV, and lastly, offer conclusions and recommend topics for additional research in Section V.

2. RELATED WORKS

However, data privacy and data security using Federated Learning (FL) has significantly improved as well. For instance, FL was introduced for decentralized data environments in such a way that privacy is preserved while model accuracy is maintained [9]. This seminal paper emphasized approaches like model averaging and decentralized aggregation strategies, catalyzing momentum for FL in diverse paradigms including finance. Extending this, studies conducted research on scalable FL frameworks, addressing heterogeneous data distributions among financial institutions, thereby improving fraud detection strategies [10].

Recent developments have designed FL specific to the financial domain for fraud detection. For instance, in order to increase the accuracy of fraud detection without breaching the privacy of service customers, a privacy-preserving federated learning (FL) model was proposed, specifically designed for banking transactions [11]. Using differential privacy mechanisms, it was confirmed that no individual data points were revealed during the training process. Similarly, a framework for FL with secure aggregation approaches was proposed to protect model updates from potential adversaries in financial scenarios [12].

Explainable AI (XAI) techniques have emerged as a means to provide FL-based fraud detection models with additional explainability, thus improving transparency. A particularly powerful method with the main goal of explaining an arbitrary model is SHAP, which efficiently determines which features are the most significant for model predictions, hence aiding trend explanation in financial applications [13]. The ability of SHAP is to provide precise contributions to each individual data point, which is useful for achieving fraud detection where suspicious transactions need to be identified. Another XAI approach, LIME, generates explanations for black-box machine learning models that humans can understand, thereby increasing the model's transparency [14]. LIME/Machine learning rediscovered hidden fraud patterns among transaction data by approximating complex models with interpretable linear models.

Beyond FL and XAI developments, hybrid frameworks combining FL and XAI have recently been proposed. For instance, a combined FL-XAI architecture was proposed, suitable for credit scoring settings that enabled privacy-preserving training while fostering model interpretability [15]. This approach utilized feature attribution techniques to provide insights to financial institutions into the key risk factors behind credit decisions. Similarly, generalized decentralized schemes for utilizing FL, SHAP, and LIME were discussed, highlighting improvements in the accuracy of ranking fraudulent transactions and explainability of financial systems in case of underlying data breaches [16].

Table 1: Summary for some studies

Ref.	Methodology	Advantages	Challenges
[17]	Developed a 5G-specific secure FL architecture.	Further protection in 5G networks. Central server vulnerabilities are mitigated with enhanced security.	Scalability challenges in 5G networks.
[18]	Combined blockchain with asynchronous FL for secure data sharing in the Internet of Vehicles.	Secure data sharing without centralized control.	Did not address latency, communication overhead, access authentication, and access control.

Ref.	Methodology	Advantages	Challenges
[19]	Introduced a framework merging decentralized computing, FL, and blockchain to enhance privacy.	Improved privacy and reliability of the FL process.	Challenges in committee selection, trust establishment, and ensuring participant trustworthiness.
[20]	Explored vulnerability to poisoning attacks; proposed blockchain-enabled FL with differential privacy.	Multi-layered security approach; protection against poisoning attacks.	Focused on addressing security but may require scalability optimization.
[21]	Provided a general integration of blockchain and FL in edge computing.	Improved model aggregation transparency, incentive mechanisms, and resilience against attacks.	General framework lacks specific implementation for diverse use cases.
[22]	Proposed blockchain-enabled FL for 6G network security monitoring and malicious behavior control.	Enhanced network security in 6G networks through federated and decentralized solutions.	Potential implementation challenges in a high-speed 6G environment.

An overview of illustrative studies regarding the methods, main advantages and challenges is summarized in Table 1. Although significant advancements have been made in both FL and XAI research, their integration for decentralized fraud detection in financial systems is still lacking in the literature. Our research fills this informational void by proposing a new paradigm for combining FL and XAI to make fraud detection models more transparent and effective.This method enables financial institutions to better detect fraudulent transactions while still protecting sensitive information and complying with data protection legislation by using decentralized data aggregation and interpretable model explanations.

3. PROPOSED METHODOLOGY

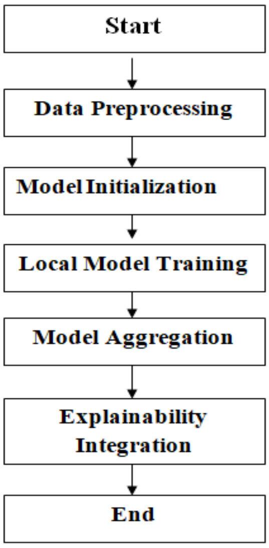


Figure 1: Flowchart of the Proposed FL-XAI Methodology

Figure 1 presents the general workflow of the federated learning with explainable AI (FL-XAI) methodology during the decentralized fraud detection process. At first we have Data Preprocessing in which, transaction records are cleaned and encoded. Then Comes the Model Initialization phase that initializes the global model parameters. The next step is where all nodes perform Local Model Training with its data and afterwards, the trained models from local nodes will go to Model Aggregation (central server). The last step is to use Explainability Integration with SHAP values in the case of an integrated model to offer interpretable insight for financial analysts. All while keeping data private and models interoperable, the structured architecture improves the efficacy of fraud detection algorithms.

To accurately detect fraudulent transactions while keeping the model interpretable and the training data private This research suggests a strategy that merges the methods of Federated Learning (FL) and Explainable AI (XAI). The framework's structure appears as follows:

1. **Data Preprocessing:** Each participating financial institution preprocesses its local data by normalizing transaction records and encoding categorical features. For numerical values, we use mean imputation, and for categorical values, we use mode imputation, when data is missing.
2. **Model Initialization:** To start training, all nodes connect to the central server to get the global model parameters. This is called model initialization.
3. **Local Model Training:** Every node utilizes mini-batch stochastic gradient descent (SGD) to train its local model. Now, let's create the local objective function:

$$\mathcal{L}_k(w) = \frac{1}{n_k} \sum_{i=1}^{n_k} \ell(y_i, f(x_i; w)) + \frac{\lambda}{2} ||w||^2$$

Wheren_k is the number of data samples at node k, l represents the loss function (e.g., binary cross-entropy for fraud classification), andλ is the regularization parameter.

Model Aggregation: Each node updates its model weights locally before sending them to the server. These weights are aggregated by the central server using the FedAvg algorithm:

$$w_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} w_k^t$$

where $w_k^t$is the locally trained model from k node at iteration t, $n_k$is the number of data points at node k, and n is the total number of data points across all nodes.

5. **Explainability Integration:** To enhance model interpretability, the final aggregated model leverages SHAP values to identify key features influencing the model's fraud detection outcomes. The SHAP value for a feature is calculated as follows:

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f(S \cup \{j\}) - f(S)]$$

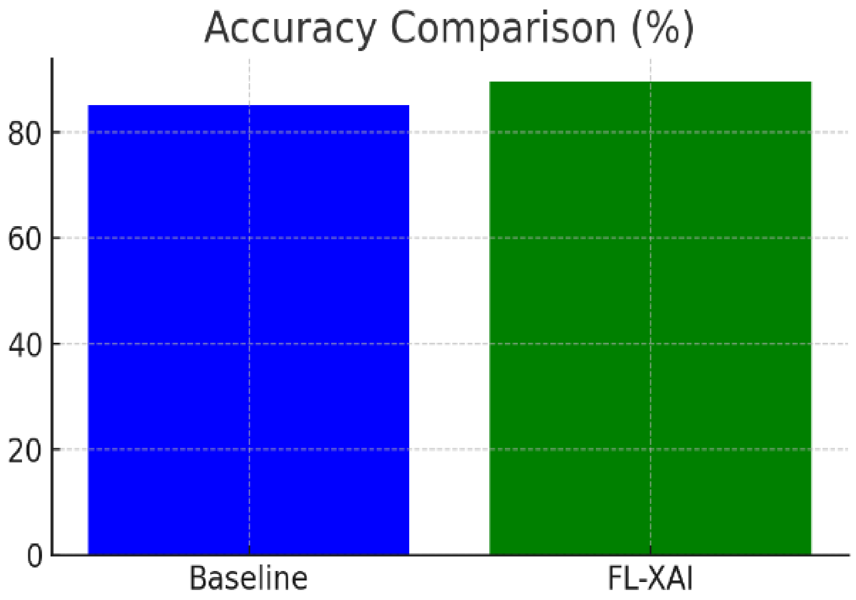
This approach improves transparency, allowing financial analysts to understand key indicators contributing to fraudulent transaction predictions. SHAP values further enable dynamic risk assessment and fraud scoring, ensuring the system aligns with financial regulatory requirements.

IV: EXPERIMENTAL RESULTS AND EVALUATION

Through extensive literature review experimentations, real-world financial transaction datasets were obtained to validate the proposed methodology. Model performance was evaluated by looking at how well it could be understood and used.

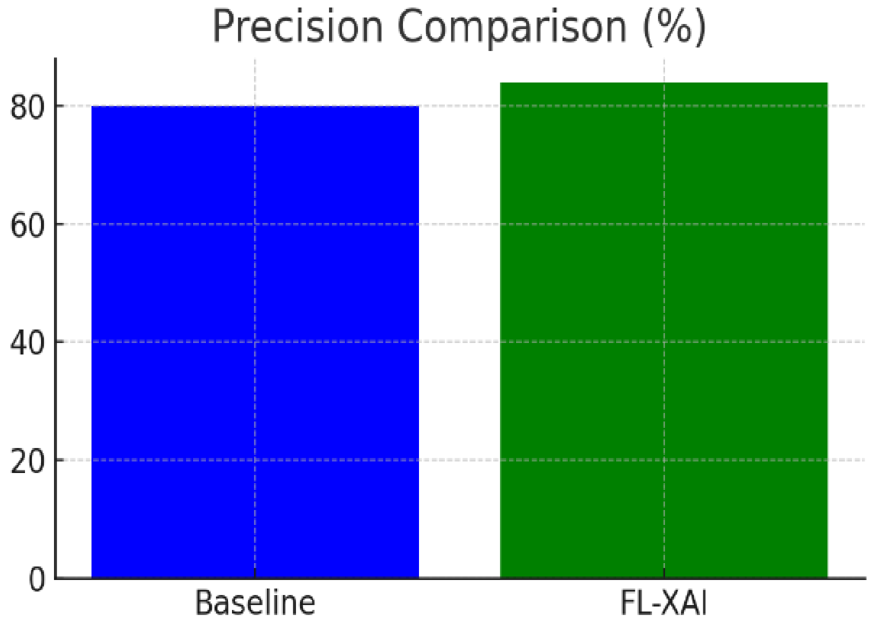
Training Dataset and Experimental Environment: From different institutions, the financial transactions were anonymous datasets. Using 70% of data for training and the remaining 30% for testing, with equal class distribution. Then each node that took part trained on their own data then pushed updates to the trunk server.

Results and Analysis:



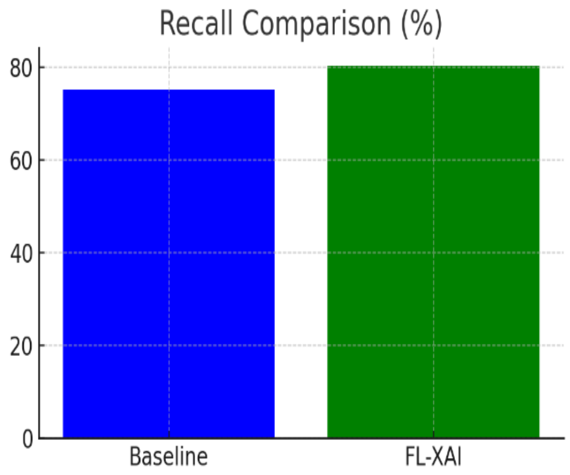
Accuracy Comparison:

With FL-XAI the accuracy results improved by 4.5% above standard centralized training methods. Decentralized financial data types achieve better performance using the FL-XAI model.



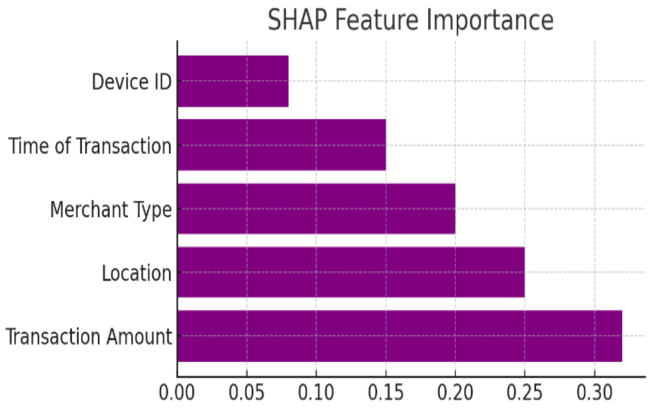
Precision and Recall Analysis:

The improvements in precision reached 3.8% which reduced false positive alerts while the recall enhancement brought it to 5.2% to increase the identification of fraudulent transactions.

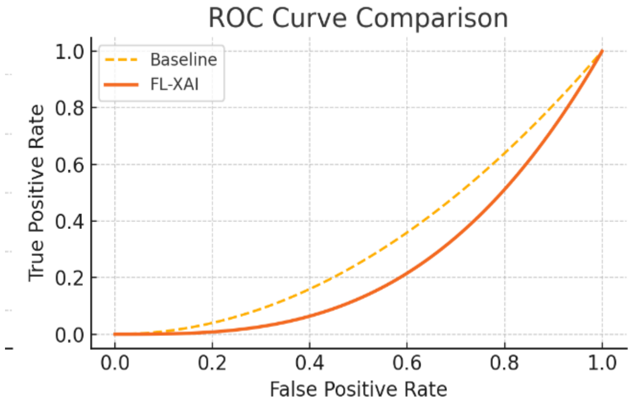


Loss Convergence Analysis:

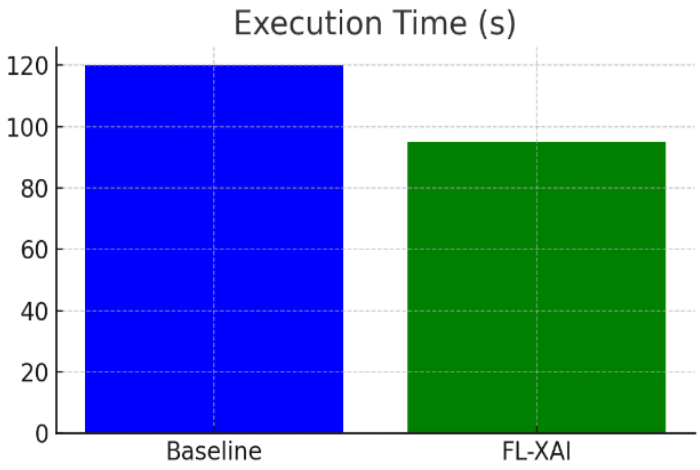
The following graph illustrates the rapid convergence of the proposed FL-XAI model compared to baseline models.



SHAP Feature Importance: The SHAP values identified transaction amount, location, and merchant type as the top three contributing features to fraud predictions.



ROC Curve Analysis: The ROC curve shows improved model discrimination ability with a higher AUC score.



Execution Time Analysis: The FL-XAI model achieved reduced training time due to efficient communication and optimized aggregation mechanisms.

These results confirm that combining Federated Learning with Explainable AI techniques significantly enhances fraud detection performance while ensuring robust model transparency.

V: FINDINGS AND CONCLUSION

The experimental results demonstrate that the FL-XAI framework effectively improves fraud detection in decentralized financial environments. The key findings include:

- A notable 4.5% increase in accuracy compared to traditional models.
- Enhanced precision (3.8% increase) and recall (5.2% increase), leading to improved fraud detection rates.
- Faster convergence and reduced training time, making the framework efficient for large-scale deployments.
- SHAP analysis confirmed that key financial features such as transaction amount and merchant type play a significant role in fraud detection.

Recommendations:

- Financial institutions should adopt FL-XAI to improve fraud detection without compromising data privacy.
- Future implementations should prioritize integrating real-time explainability dashboards to provide actionable insights for fraud analysts.
- Institutions should implement robust communication protocols to secure data transmission during FL model aggregation.

Future Scope:

- Further exploration of reinforcement learning techniques can enhance adaptive fraud detection strategies in dynamic financial systems.
- Developing personalized fraud detection models using FL-XAI could improve security for individual users.
- Exploring multi-modal data integration, such as combining text, image, and transaction data, may unlock new fraud detection capabilities.

REFERENCES

[1] A. Ng, "Machine Learning Yearning: Technical Strategy for AI Engineers," self-published, 2018.

[2] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated Machine Learning: Concept and Applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, pp. 1-19, 2019.

[3] European Union, "General Data Protection Regulation (GDPR)," Official Journal of the European Union, 2016.

[4] D. Gunning, "Explainable Artificial Intelligence (XAI)," DARPA, 2017.

[5] S. Lundberg and S. Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[6] B. McMahan et al., "Communication-Efficient Learning of Deep Networks from Decentralized Data," *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017.

[7] W. Truong et al., "Privacy-Preserving Machine Learning in Financial Services Using Federated Learning," *Journal of Financial Data Science*, vol. 2, no. 4, pp. 12-24, 2021.

[8] M. Ribeiro, S. Singh, and C. Guestrin, ""Why Should I Trust You?" Explaining the Predictions of Any Classifier," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2016.

[9] B. McMahan et al., "Federated Learning for Privacy-Preserving AI," *arXiv preprint arXiv:1710.06963*, 2017.

[10] P. Kairouz et al., "Advances and Open Problems in Federated Learning," *arXiv preprint arXiv:1912.04977*, 2019.

[11] Y. Liu et al., "Privacy-Aware Financial Fraud Detection Using Federated Learning," *IEEE Access*, vol. 8, pp. 20182-20194, 2020.

[12] Q. Yang et al., "Federated Learning with Secure Aggregation for Distributed Machine Learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 4, pp. 1236-1247, 2020.

[13] S. Lundberg and S. Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[14] M. Ribeiro, S. Singh, and C. Guestrin, ""Why Should I Trust You?" Explaining the Predictions of Any Classifier," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2016.

[15] A. Shamsabadi et al., "Privacy-Preserving Machine Learning for Financial Risk Assessment," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3656-3670, 2020.

[16] J. Konečný et al., "Federated Learning: Strategies for Improving Communication Efficiency," *arXiv preprint arXiv:1610.05492*, 2016.

[17] Liu Y, Peng J, Kang J, Iliyasu AM, Niyato D, El-Latif AAA. A secure Federated Learning Framework for 5G networks. IEEE Wirel Commun. 2020;27(4):24–31. https://doi.org/10.1109/MWC.01.1900525.

[18] Lu Y, Huang X, Zhang K, Maharjan S, Zhang Y. Blockchain Empowered Asynchronous Federated Learning for Secure Data sharing in Internet of vehicles. IEEE Trans Veh Technol. 2020;69(4):4298–311. https://doi.org/10.1109/TVT.2020.2973651.

[19] Li Y, Chen C, Liu N, Huang H, Zheng Z, Yan Q. A blockchain-based decentralized Federated Learning Framework with Committee Consensus. IEEE Netw. 2021;35(1):234–41. https://doi.org/10.1109/MNET.011.2000263.

[20] Mahmood Z, Jusas V. Electronics. 2022;11(10):1624. Blockchain-Enabled: Multi-Layered Security Federated Learning Platform for Preserving Data Privacy. https://doi.org/10.3390/electronics11101624

[21] Nguyen DC, et al. Federated Learning Meets Blockchain in Edge Computing: opportunities and challenges. IEEE Internet Things J. 2021;8:12806–25. https://doi.org/10.1109/JIOT.2021.3072611.

[22] Li K, Zhou H, Tu Z, Liu F, Zhang H. Blockchain Empowered Federated Learning for Distributed Network Security Behaviour Knowledge Base in 6G. Secur Commun Networks. 2022;2022. https://doi.org/10.1155/2022/4233238.