A Comprehensive Investigation of Fraud Detection Behavior in Federated Learning

Rui Sun

Software Engineering, Beijing University of Technology, 100124 Beijing, China

Abstract. This research delves into the application of Federated Learning (FL) models for detecting fraud across different financial bodies. FL facilitates decentralized training of models using local data, ensuring privacy, crucial for handling sensitive financial data. The comparison involves three machine learning models - Artificial Neural Networks (ANN), Random Forest (RF), and Convolutional Neural Networks (CNN) - to assess their efficacy in the FL context. While ANN and CNN demonstrate strong capacity in identifying complex fraud patterns, their communication efficiency and overfitting challenges are significant. In contrast, RF offers more robustness to Non-independent and Identically Distributed (non-IID) data and is less prone to overfitting, though it poses communication overhead issues. This paper also highlights the challenges of FL in fraud detection, including data heterogeneity, communication costs, and security risks. This paper proposed future research directions, emphasizing model personalization, communication optimization, and advanced privacy-preserving techniques. By addressing these challenges, FL can offer scalable, secure solutions for real-time fraud detection, ensuring the protection of sensitive financial data while enhancing detection accuracy across diverse data sources.

1 Introduction

Legally, fraud is defined as the unlawful holding of property for a specific reason, involving either false information or hiding the truth, leading to significant fraud in public or private property dealings. In 2023, Chinese police solved more than 437,000 fraud cases [1], but it may have been too long before the cases were solved, and the victims' financial losses could not be recovered, and these cases caused extremely serious economic losses. Governments all over the world have been trying to prevent and reduce the happening of fraud cases through various ways. The traditional way adopted by the government is to popularize various fraud methods to the public through the media. However, there are too many fraud methods and the means of cheating are constantly updated, so this method is challenging to achieve good results.

In recent years, with the development of computer technology, the Chinese government has updated a new means to resist fraud - an application installed on mobile phones, called the National Anti-Fraud Center. The application finds potential victims and cheaters through

Corresponding author: sunrui@emails.bjut.edu.cn

big data analysis, can intercept fraud short messages and fraud calls, and also has an early warning and dissuasion platform, which can remind the user by phone after the user receives a high-risk call [2], which has a very good performance on preventing fraud, but this application also has small shortcomings. The interception system cannot accurately identify the fraud Short Message Service (SMS), and it is likely to intercept other SMS by mistake, which brings inconvenience to the user's life. At the same time, the system also needs to review all the user's SMS and phone calls, which violates the user's privacy excessively. With this in mind, more advanced methods that can monitor bank cards can be considered such as federated learning.

In the past few years, many researchers have demonstrated the feasibility of this method. For example, Yang et al. proved that the average test results of Federated learning for Fraud Detection (FFD) are higher than those of traditional Fraud Detection System (FDS) [3]. And federated learning technology has also made great progress in recent years. For example, Salam et al. studied federated learning model for credit card fraud detection with data balancing techniques, and compared the different experimental results of logistic regression [4]. Reddy et al. designed a hybrid algorithmic optimization-based deep learning technique for fraudulent credit card transaction detection [5]. Therefore, it is necessary to conduct a comprehensive review of federated learning for fraud detection.

The remainder of the paper is organised as follows. In Section 2, this paper will discuss various methods for fraud detection via federated learning in the last two years. In Section 3, this paper will discuss the current advantages of federated learning for fraud detection, a horizontal comparison between different methods, and future challenges in this field. Section 4 summarizes the paper, and presents conclusions drawn from the contents discussed here.

2 Method

2.1 Introduction of federate learning

Federated learning (FL) is a collaborative method of machine learning in which several clients use their own data to train the model locally and then send the model updates to a central server for aggregation. This process ensures data privacy by keeping the original data on local devices [6]. Federated learning primarily operates on two key principles: local training on user devices and the aggregation of model updates by a central server. This makes FL highly suitable for privacy-sensitive applications, such as fraud detection, where sharing raw data is legally or ethically prohibited. It is distinct from traditional centralized machine learning due to its decentralized approach [7]. In contrast to centralized learning, which requires transferring all data to a central server for processing, FL enables learning across decentralized nodes without compromising user privacy. The system significantly reduces the risks of data exposure during transmission. This makes it an ideal solution for fraud detection, where data privacy is paramount, as sensitive financial data can be trained without leaving the devices.

Additionally, FL has proven to be highly beneficial for scenarios involving heterogeneous data. In fraud detection, data sources might be diverse across various devices and institutions. FL addresses this issue by supporting Non-independent and Identically Distributed (non-IID) data, allowing it to integrate diverse datasets while maintaining the accuracy of the global model. This ability to work with decentralized and diverse datasets makes FL particularly effective for fraud detection, improving model robustness and reducing the likelihood of overfitting or bias.

2.2 Federated learning workflow for fraud detection

Zhang et al. provides a detailed explanation of how FL can be applied to fraud detection in distributed environments. In an FL system, data remains localized at each institution, and models are trained on these local datasets. The workflow begins with the local model being trained on each bank's dataset. Model updates, rather than data, are sent to a central server, which aggregates these updates using techniques like Federated Averaging (McMahan et al.,), thus generating a global shared model. This approach ensures that significant data, such as bank transactions, never leaves its original location, safeguarding privacy.

Ye et al. highlight communication mechanisms in federated systems, particularly focusing on the importance of encrypted communication protocols [8]. These mechanisms are vital for ensuring data privacy during the transfer of model updates. Additionally, they emphasize the challenge of data heterogeneity, as banks may have different types of fraud detection data that are non-IID. Solutions such as model personalization and clustering help tackle this issue, improving the accuracy and robustness of fraud detection models.

Yang et al. further discuss the challenges of high communication costs and latency in FL, particularly for fraud detection. They propose several strategies to address these issues, such as decreasing the number of communication rounds and optimizing local computations. The system is designed to balance the balance between communication overhead and model accuracy. Moreover, their FL framework for fraud detection, known as FFD, shows significant improvements in accuracy (95.5% Area Under Curve) compared to traditional models, demonstrating the effectiveness of FL in addressing data insufficiency and skewed distribution challenges in fraud detection.

2.3 Federated learning-based artificial neural network (ANN) for fraud detection

Zhang et al. explain that the architecture of an ANN in the federated learning context follows a typical multi-layer structure, including input, hidden, and output layers. In a distributed setting, the ANN is designed to accommodate the distributed nature of data, with each client training the ANN locally using their dataset before sharing the model updates with the central server. The structure ensures that sensitive data, such as fraud detection features, remains on the local devices without compromising the model accuracy.

Yurochkin et al. further discuss how data preprocessing and feature selection are crucial in fraud detection [9]. Key fraud-related features, such as transaction patterns and user behaviors, must be effectively preprocessed and selected for training. In the federated learning setting, each node handles data preprocessing independently, ensuring that local variations are respected while standardizing feature extraction across the global model.

In terms of training, Zhao et al. emphasizes that federated learning with ANN relies on the Federated Averaging algorithm for training on distributed data [10]. Local models are trained using backpropagation, and the model parameters are shared with the central server, where they are averaged to renewal the global model. This reduces communication costs while ensuring that the model is optimized through techniques such as gradient descent.

The innovation in federated learning applied to ANN lies in privacy preservation and the ability to handle non-IID data. Yurochkin et al. highlights the Bayesian nonparametric approach for federated learning with neural networks, where model parameters are shared and matched across clients to build a global model. This method effectively decreases the impact of data heterogeneity and enhances model performance without exposing sensitive fraud detection data.

2.4 Federated learning-based random forest (RF) model

Haffar et al. discussed the foundational structure of the Random Forest model within the FL framework. In RF, multiple decision trees are independently constructed using different subsets of features and data. Each tree votes on a classification result, and the majority decision becomes the final output. This method provides robustness and reduces overfitting, which is crucial in fraud detection due to the high dimensionality and imbalance in fraud datasets. Within FL, each client node constructs its trees locally, using its own private data. This decentralized approach ensures data privacy, as raw data never leaves the node. Instead, only model parameters, such as the trained trees, are grouping centrally to form a global forest [11].

In terms of data characteristics, Yang et al. note that RF models are particularly effective at handling high-dimensional information with mixed feature types—common traits of fraud detection datasets. Fraud datasets often involve a combination of transactional details, user behavior, and historical data, which are non-IID across different organizations or devices. RF models are adept at managing such variability, making them suitable for fraud detection tasks where data is heterogeneously distributed [12].

When RF models are integrated with FL, the process involves constructing decision trees at each client node without sharing the raw data. The trees are then merged in a way that preserves privacy, utilizing encryption techniques such as differential privacy or secure multi-party computation. These methods guarantee that individual information points or sensitive features are not exposed during the model aggregation process. Yang et al. also highlights the use of homomorphic encryption to secure the communication of model parameters between the clients and the server, minimizing the risk of information leakage while maintaining high accuracy and efficiency in the RF model [11, 12].

2.5 Federated learning-based convolutional neural network (CNN) model

Poudyal et al. describe CNN's architecture within the FL framework, emphasizing its ability to extract critical features from high-dimensional fraud detection data. CNNs excel at detecting patterns through convolution and pooling layers, making them highly suitable for complex fraud scenarios where subtle transaction anomalies may be present. The convolution layers extract hierarchical features, while pooling layers downsample the data, reducing the model's complexity. This makes CNN particularly effective in fraud detection, as it handles large-scale input while maintaining computational efficiency [13].

In comparison with traditional Artificial Neural Networks, CNNs offer advantages in feature extraction. Zhu et al. explain that while ANNs rely on fully connected layers for feature extraction, CNNs can capture local dependencies within data through convolution, allowing for more efficient processing of spatial relationships, which are crucial in detecting fraud patterns. However, CNNs may require more computational resources, which can be a drawback in decentralized settings [14].

Federated learning significantly impacts CNN performance in terms of communication efficiency. As CNNs involve large models and numerous parameters, FL frameworks like Federated Averaging optimize communication by aggregating only model updates instead of full datasets, thus reducing the communication load. However, the convolution operations still generate substantial weight updates, which must be managed to ensure communication efficiency without degrading model accuracy. This balance is critical to effectively applying CNNs to FL in fraud detection [13, 14]**错误!未找到引用源。**.

3 Discussion

3.1 Comparisons of federated learning for fraud detection

Federated learning introduces a method for machine learning in scenarios where data privacy is of utmost importance, especially in the field of fraud detection. By enabling decentralized training across various nodes—such as financial institutions—FL ensures that sensitive data never leaves its local environment while contributing to a global model. In this context, different machine learning models have been adapted to function within the FL framework, each exhibiting unique strengths and limitations when applied to fraud detection.

Artificial Neural Networks are frequently employed within FL due to their capability to model complex, non-linear relationships, which is particularly useful in identifying subtle and sophisticated fraud patterns. Their flexibility allows them to be trained on distributed data while contributing to a global model that captures intricate behavioral patterns. However, a primary challenge with ANNs, particularly in the FL setting, is their tendency to overfit when local datasets are small or imbalanced. This issue is exacerbated by the non-IID nature of fraud detection information, where each financial institution may experience distinct types of fraud, making global generalization difficult. While the aggregation methods in FL, such as Federated Averaging, help in mitigating some of the overfitting concerns, ensuring consistent performance across diverse and unbalanced datasets remains an ongoing challenge.

In contrast, Random Forests offer a more robust and interpretable alternative for fraud detection within FL environments. RF models, by nature, are ensemble-based, consisting of multiple decision trees that operate on random subsets of the data. This structure allows RF to handle heterogeneous and high-dimensional fraud detection datasets more effectively than ANNs. In a federated setting, RF models benefit from their ability to reduce the impact of data imbalance or noise, as the independent decision trees are less prone to overfitting. However, integrating RF models into FL frameworks presents unique challenges. The decentralized nature of FL requires that only model parameters, rather than raw data, are exchanged between clients and the central server, and the transmission of decision trees—especially when the forest is large—can introduce significant communication overhead. This communication burden increases as the depth and number of trees expand, making it necessary to develop more efficient aggregation and compression techniques to minimize this cost.

Convolutional Neural Networks, though traditionally associated with image recognition tasks, have found relevance in fraud detection when transactional data can be represented in spatial or temporal formats. CNNs excel at capturing local dependencies through convolutional layers, making them particularly effective in identifying sequential anomalies or recurring behavior patterns in financial transactions. In the context of FL, CNNs provide powerful feature extraction capabilities that can be leveraged to detect fraud across time-series data or graph-based detection frameworks. However, CNNs, with their large number of parameters, tend to exacerbate the communication inefficiencies inherent in FL. The challenge, therefore, lies in reducing the communication costs associated with transferring model updates, while still maintaining the model's performance. Techniques such as gradient compression and sparse communication are being explored to address this issue, but the trade-off between accuracy and efficiency remains a critical consideration.

The performance of these models in FL for fraud detection also depends heavily on the choice of aggregation methods. Federated Averaging, though widely used, often struggles with non-IID data distributions, as local updates can vary significantly across clients. This divergence in updates can result in a global model that fails to generalize effectively across all participants. As a result, more advanced aggregation techniques, such as Federated Proximal (FedProx), have been introduced, introducing regularization terms that account for the heterogeneity in local data. These methods can help to enhance the robustness of the model and contribute to the broader challenge of balancing communication efficiency with model accuracy and privacy.

3.2 Challenges

The implementation of federated learning in fraud detection is not without significant challenges, despite its promise in preserving data privacy and enabling decentralized learning. One of the foremost issues is the heterogeneity of data across different clients, often referred to as the non-IID nature of the data. Financial institutions, for example, deal with distinct transaction patterns, types of fraud, and customer demographics, resulting in local datasets that may vary significantly. This variation interferes with the convergence process of the global model, and the reason is that the model needs to generalize on different datasets that may have local biases or skewed distributions. Non-IID data poses a challenge not only to model accuracy but also to the stability of the learning process, as updates from some clients may disproportionately influence the global model, leading to suboptimal performance for other clients. Addressing this requires novel approaches to model aggregation and client-specific adaptations to ensure the global model remains robust across varied environments.

Communication overhead presents another considerable challenge in FL, particularly in large-scale deployments where multiple clients are involved. Fraud detection systems rely on timely updates to maintain real-time effectiveness, but this also introduces a problem: frequent transfers of model parameters between clients and a central server substantially increase bandwidth consumption and may cause latency. This issue becomes even more pronounced when using models with a high number of parameters, such as deep neural networks, where the volume of data transferred is substantial. Optimizing the communication process is essential to ensure that FL systems remain practical for real-world applications. Techniques like gradient compression, sparse communication, or reducing the frequency of model updates have been explored to mitigate this challenge, but these solutions often introduce balance in terms of accuracy and timeliness. Striking the right balance between communication efficiency and model performance remains a persistent obstacle, particularly in environments where low latency is critical for preventing fraud in real-time.

Security vulnerabilities further complicate the deployment of FL in sensitive fields such as fraud detection. Although FL improves privacy by keeping data localized, the transmission of model updates still exposes the system to potential risks. Malicious clients may attempt to manipulate model updates, introducing poisoned data to corrupt the global model, which can result in increased false positives or negatives in fraud detection systems. Additionally, inference attacks may allow adversaries to glean sensitive information from the aggregated model updates, posing a direct threat to financial data security. Addressing these concerns requires the integration of robust security measures, such as secure aggregation, differential privacy, and adversarially resilient aggregation algorithms, to defend against potential attacks. However, implementing these techniques often comes with added computational and communication costs, further straining the already resource-intensive FL process.

Beyond technical challenges, regulatory and ethical concerns are equally significant in the application of FL for fraud detection. Financial data is subject to stringent privacy laws, such as the General Data Protection Regulation (GDPR) in Europe and the California Consumer Privacy Act (CCPA) in the United States, which place strict requirements on data collection, storage, and sharing. Although FL provides a framework that aligns with these regulations by keeping data decentralized, there are still concerns regarding the indirect exposure of sensitive information through model updates. Ensuring compliance with regulatory frameworks, while maintaining the transparency and fairness of the global model, is critical to the successful implementation of FL. Moreover, ethical considerations surrounding data ownership, consent, and bias in fraud detection models must be navigated carefully. Systems that disproportionately impact certain demographics or regions may inadvertently perpetuate inequalities or result in unfair treatment of minority groups. The ethical design of FL systems must therefore prioritize fairness and accountability, ensuring that the models are accurate and equitable.

The challenges of federated learning in fraud detection underscore the complexity of applying this technology in real-world financial systems. Addressing these issues will require continued innovation, not just in terms of algorithmic efficiency but also in building secure, compliant, and ethically sound frameworks that can operate effectively in diverse environments. Only through overcoming these barriers can federated learning realize its full potential as a scalable and privacy-preserving solution for combating financial fraud.

3.3 Future prospects

The future prospects for federated learning in fraud detection hold substantial promise, driven by both technological advancements and the growing need for privacy-preserving solutions. One of the most significant areas of future research is the development of model personalization techniques. Given the heterogeneous nature of fraud data across financial institutions, a one-size-fits-all global model may not sufficiently capture localized fraud patterns. Personalization techniques that allow the global model to adapt to the unique characteristics of each client's data without sacrificing the benefits of shared learning will be crucial. This could be achieved through approaches like meta-learning, which enables the model to learn how to fine-tune itself based on individual client needs, or through client-specific adjustments that allow each institution to retain some level of model autonomy. These techniques not only promise to improve fraud detection accuracy but also reduce false positives, which can be particularly costly in real-world financial systems.

Communication efficiency will continue to be a critical focus in the evolution of FL for fraud detection. As models become more complex and the number of clients participating in FL networks grows, the strain on communication infrastructure will intensify. Researchers are already exploring advanced gradient compression techniques, sparse updates, and adaptive communication schedules that aim to reduce the volume of data transmitted between clients and the central server without compromising the model's accuracy.As these methods develop further, it will become more practical to incorporate real-time fraud detection systems that rely on federated learning, particularly in environments where low-latency decision-making is critical. Furthermore, the future may see the emergence of hybrid approaches that combine FL with edge computing, where more computational tasks are distributed closer to the data sources, further reducing communication bottlenecks and improving responsiveness.

Security and privacy will remain at the forefront of FL's development, particularly as adversarial techniques evolve. Future research may focus on enhancing existing privacy protection ways, in order to build more robust defenses against increasingly complex attacks. The refinement of these techniques will be essential for maintaining the integrity of fraud detection systems, especially in the face of adversaries who might attempt to poison models or exploit vulnerabilities in model updates. Another exciting prospect is the integration of homomorphic encryption, which would allow computations to be performed on encrypted data, thereby further safeguarding sensitive financial information during model training. These innovations will need to strike a delicate balance between security and computational efficiency, guarantee that fraud detection models remain both robust and scalable.

In addition to technical advancements, regulatory compliance and ethical considerations will shape the future of FL in fraud detection. As data privacy laws continue to evolve, FL frameworks will need to adapt to meet stringent legal requirements across different jurisdictions. This may involve the development of more transparent aggregation methods that ensure data anonymity while still providing regulators with confidence that the system adheres to privacy standards. Ethical considerations, such as ensuring fairness and mitigating bias in fraud detection models, will also become more prominent. As FL systems are deployed more widely, it will be imperative to ensure that they do not disproportionately

impact certain demographics or regions. Future research may focus on creating fairness-aware algorithms that can detect and correct biases in real-time, ensuring that FL systems operate equitably across diverse populations.

Ultimately, the future of federated learning in fraud detection lies in its ability to evolve into a more adaptable, efficient, and secure system. As new techniques are developed to address the current challenges, FL is poised to become a cornerstone technology in the fight against financial fraud, offering both scalability and privacy protection without compromising on detection performance. The convergence of advanced personalization methods, optimized communication protocols, and enhanced security mechanisms will ensure that FL remains a cutting-edge solution in an increasingly data-driven financial landscape.

4 Conclusion

This research examines the use of federated learning in detecting financial fraud, emphasizing its capability to improve precision detection while ensuring the protection of data privacy. By evaluating various machine learning models within the federated learning framework, this research reveals their respective strengths and limitations in handling complex fraud patterns. Artificial Neural Networks and Convolutional Neural Networks demonstrate significant capabilities in capturing intricate fraud behaviors, particularly in dealing with high-dimensional and complex data. However, they face challenges in communication efficiency and data heterogeneity, where local datasets differ significantly from the global data, leading to potential overfitting or performance degradation. In contrast, Random Forest exhibits greater robustness when dealing with non-independent and identically distributed data, though its larger communication overhead limits its scalability.

The decentralized nature of federated learning provides an innovative solution for handling privacy-sensitive financial data, but it also introduces new challenges. This study identifies the main obstacles to being non-IID data distributions and high communication costs, which may impact on the real-time effectiveness and overall efficiency of the fraud detection models, especially in large-scale implementations. Although federated learning theoretically enhances data security and privacy, practical applications still necessitate sophisticated techniques like differential privacy and secure multi-party computation to better protect sensitive information.

Despite these challenges, the study maintains an optimistic outlook on the future of federated learning. With advances in model personalization and communication optimization, federated learning’ s performance in fraud detection can be significantly improved. Moreover, as privacy-preserving technologies continue to advance, federated learning could emerge as a cornerstone in fraud prevention strategies for financial institutions. Personalized models, which adapt to the unique data distributions of individual institutions, can enhance generalization and accuracy. Future research will likely focus on balancing communication efficiency, security, and real-time detection accuracy to refine the system’ s practicality and scalability.

References

1. J. Qu, K. Lin, Y. Wu, I. Y. Sun, Fear and perceived risk of cyber fraud victimization among Chinese University students. Crime, Law and Social Change, 1-20 (2024).
2. S. Shiyang, Study on Telecommunication Fraud from a Student's Perspective. International Journal of Frontiers in Sociology, 5(16) (2023).

3. W. Yang, Y. Zhang, K. Ye, et al., FFd: A federated learning based method for credit card fraud detection. In Proceedings of the Big Data–BigData 2019, Held as Part of the Services Conference Federation, San Diego, CA, USA, June 25–30, 18-32 (2019).
4. M. A. Salam, K. M. Fouad, D. L. Elbably, S. M. Elsayed, Federated learning model for credit card fraud detection with data balancing techniques. Neural Computing and Applications, 36, 6231-6256 (2024).
5. V. V. Krishna Reddy, R. V. Kumar Reddy, et al., Deep learning-based credit card fraud detection in federated learning. Expert Systems With Applications, 255, 124493 (2024).
6. L. Li, Y. Fan, M. Tse, et al., A review of applications in federated learning. Computers & Industrial Engineering, 149, 106854 (2020).
7. C. Zhang, Y. Xie, H. Bai, et al., A survey on federated learning. Knowledge-Based Systems, 216, 106775 (2021).
8. M. Ye, X. Fang, B. Du, et al., Heterogeneous federated learning: State-of-the-art and research challenges. ACM Computing Surveys, 56(3), 1-44 (2023).
9. M. Yurochkin, M. Agarwal, S. Ghosh, et al., Bayesian nonparametric federated learning of neural networks. In International Conference on Machine Learning. PMLR, 7252-7261 (2019).
10. Y. Zhao, M. Li, L. Lai, et al., Federated learning with non-iid data. arXiv preprint arXiv, 1806.00582 (2022).
11. R. Haffar, D. Sanchez, J. Domingo-Ferrer, Explaining predictions and attacks in federated learning via random forests. Applied Intelligence, 53(1), 169-185 (2023).
12. Y. Liu, Y. Liu, Z. Liu, et al., Federated forest. IEEE Transactions on Big Data, 8(3), 843-854 (2020).
13. A. Poudyal, U. Tamrakar, R. D. Trevizan, et al., Multiarea inertia estimation using convolutional neural networks and federated learning. IEEE Systems Journal, 16(4), 6401-6412 (2021).
14. H. Zhu, Y. Jin, Multi-objective evolutionary federated learning. IEEE Transactions on Neural Networks and Learning Systems, 31(4), 1310-1322 (2019).