Evaluating Deep Learning and Machine Learning Models in Federated Learning for Credit Card Fraud Detection. A comparative study

Șener ALI
Bucharest University of Economic Studies
Faculty of Cybernetics, Statistics and Economic Informatics
Bucharest, Romania
alisener20@stud.ase.ro

This study aims to evaluate the effectiveness of both machine learning and deep learning algorithms for credit card fraud detection in the context of a federated learning framework. The fast evolution of digital banking created better experiences for customers and facilitated their access to financial services, but it has simultaneously opened new pathways for cybercriminals, making real-time fraud detection essential. Both models used in this study, XGBoost and a neural network, were trained on a publicly available dataset containing highly imbalanced data, reflecting realistic fraud scenarios. Results demonstrate that both models achieved high accuracy, yet the neural network consistently outperformed XGBoost across critical metrics such as precision, recall, and F1 score. This indicates a superior ability of deep learning models to detect fraudulent transactions in federated learning environments, highlighting their potential to improve financial security through collaborative yet privacy-preserving approaches.

Keywords: *Federated Learning, Credit Card Fraud Detection, Privacy Protection, Machine Learning, Deep Learning*

1 Introduction

In the last years, digital banking has evolved significantly, but this progress has also given rise to more advanced fraud techniques. Cybercriminals can use advanced methods, such as stealing card information used in online transactions, to commit financial crimes. Besides the traditional ways, such as card theft, criminals have at this moment online methods to commit fraud [1].

Based on the central bank's reports, billions are lost every year due to fraudulent activities. Besides the financial losses of the clients, these frauds can have a negative impact on the trust of people in financial institutions. Undermining the credibility of the banking system can create crashes on financial markets and threaten the global economy [2].

To keep a high level of trust in financial institutions, we can employ two mechanisms: fraud prevention and fraud detection. In this study, we will focus on fraud detection, the mechanism that tries to detect fraudulent transactions in real time, through the use of federated learning (FL).

Introduced by Google in 2017 to improve their keyboard text prediction, federated learning became an approach used for training AI models without the need of sharing the data across institutions [3]. Unlike the traditional approaches where aggregating the data in a centralized dataset was necessary, the federated learning offers a decentralized approach. Each institution trains a model locally on its dataset and sends it to the server, where it is aggregated to the global model. This approach enables us to obtain a model trained on all the available data while the data remains confidential [4].

The key point in developing efficient AI models is to have diverse and high-quality data. The obligation to keep the data confidential presents a significant challenge for financial institutions because they can not share their data to create a centralized dataset and train an AI model on it [5]. Moreover, based on the dataset of

an institution, the AI model is trained to detect only the frauds that the bank has already experienced, but it can remain susceptible to the attacks that other banks encountered. This makes the entire financial system more vulnerable, allowing cybercriminals to repeatedly exploit the same type of vulnerability across multiple financial institutions [6].

To address the challenges of the centralized approach, FL solves these limitations by offering the financial institutions the possibility to train a global model without sharing their data. Each bank trains a local model on its own dataset. After training, the model updates are sent to a central server and aggregated to the global model [7]. Adopting this approach, the financial institutions solve both problems, they train a model on all their available data while keeping them confidential. By leveraging knowledge from multiple institutions, FL strengthens the entire financial system by protecting the companies even from fraud cases they have not previously experienced [8].

The application of FL in fraud detection has gained significant attention due to the growing sophistication of fraudulent activities. Traditional machine learning (ML) and deep learning (DL) models have been widely used in fraud detection, but their performance in a federated setting remains an area of active research. While deep learning models, such as convolutional neural networks and recurrent neural networks, can capture complex fraud patterns, traditional ML models, such as decision trees and support vector machines, may offer advantages in interpretability and computational efficiency [9].

Given the increasing adoption of FL and the diverse range of ML and DL models, it is crucial to evaluate their effectiveness in a federated setting. This study aims to investigate the comparative performance of these models in detecting credit card fraud. Specifically, the research seeks to answer to the following question:

RQ. Which performs better in a Federated Learning framework for credit card fraud detection: a machine learning model or a deep learning model?

By addressing this research question, this study aims to provide a comprehensive evaluation of ML and DL models in FL for fraud detection, offering insights into their real-world applicability and potential for improving financial security.

The rest of the paper is organized as follows. In Section 2, related work is discussed. Section 3 provides an analysis of the dataset. Section 4 gives the details of the federated learning fraud detection implementation. The results of the study and their interpretation are presented in section 5. Conclusions are discussed in section 6.

2 Literature review

The importance of credit card fraud detection is represented by the number of public available works. There are numerous articles that talk about this topic or researches that are trying to find the best method to stop fraudsters.

Given the new context of the rise of online financial transactions and the improvements of AI in the last years, the machine learning algorithms have gotten more attention to become a relevant solution for fraud detection [10]. One of the algorithms used for handling the complexity of credit card problems is XGBoost [11]. Being one of the gradient boosting frameworks, the algorithm has a focus on regularization and adopts an iterative approach, improving its accuracy over time.

Comparing XGBoost with other machine learning algorithms when it comes to detecting credit card fraud detection shows us that the XGBoost has better performance, obtaining better results for indicators such as accuracy, precision and recall [12].

Credit card fraud detection has two major challenges: the very limited time span in

which a transaction has to be accepted or rejected and the huge amount of transactions. Only VISA has millions of transactions every day. To address these challenges, the researchers thought that neural networks would be a great solution [13]. As shown in [14], using neural networks can be a great approach to detect credit card fraud.

Besides the two previous challenges we discussed, there is another challenge when it comes to fraud detection. The financial institutions can not share the data, it would be against the data privacy and safety policies [15]. As a result, each financial institution can train models that detect only the patterns of the attacks they experienced, not the patterns of the attacks experienced by other institutions. In this way, the fraudsters can attack each financial institution with the same pattern. What if we could prevent the other financial institutions from experiencing the same type of attack that other banks already experienced? To address this problem, we can implement federated learning.

Federated learning offers the possibility to train models on local devices and share just the model's updates. The main idea of federated learning is to train a global model on all the available data without sharing it. The local trained models are sent to the central server that aggregates the weights. After update, the global model is sent back to the local devices [16]. In this way, the global model's training takes advantage of all available data while keeping it confidential [17].

Given the three problems: the necessity for fast decisions, the need to handle large amount of transactions and the requirement to keep data confidential and the good results obtained in the other studies by both XGBoost and neural networks models, this study aims to compare the performance of these two algorithms in a federated learning architecture.

3 Data

For this study, we used a publicly available dataset that consists of 284.807 transactions [18]. Given the sensitivity of financial data, each transaction has 30 numerical features that have been transformed using Principal Component Analysis (PCA) to protect customers confidentiality.

Additionally, the transactions have two explicit features: Time, representing the seconds elapsed since the first transaction and Amount, denoting the transaction's value in monetary units. The dataset also contains a Class column that identifies whether a transaction is fraudulent (1) or legitimate (0).

As all financial frauds datasets, the dataset we used in this study is highly unbalanced. The fraudulent transactions make up only 0.1727% of the total data, with 492 fraud cases out of 284,807 transactions. To illustrate this imbalance, we created a 2D scatter plot where each point represents a transaction, colored by its class label. Legitimate transactions (blue) form the majority of the data points, while fraudulent transactions (red) are less frequent, reflecting the small proportion of fraud cases.

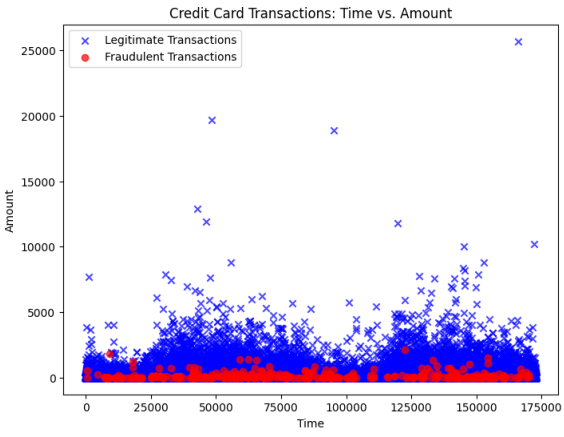

Fig. 1. Dataset visualization

For conducting this study, we needed to split the original dataset into five different smaller datasets. These subsets did not originate from actual banks, but were created to simulate local datasets. For

training each local model, 80% of the corresponding dataset was used as a training set, while the remaining 20% was reserved for testing. The initial dataset was used to test the global model.

4 Methodology

In this section, we explain the methodology employed in this research. We describe the technical approach, the algorithms implemented to address the research problem. Specifically, the methodology revolves around the implementation of XGBoost and neural networks models within the Federated Learning framework.

A. Federated Learning architecture

Fig. 2 illustrates the FL architecture used in this research. Each client node operates independently with its own dataset, representing a financial institution (e.g., a bank). Once a local model is trained on a bank's dataset, the learned parameters (weights) are transmitted to a central server.

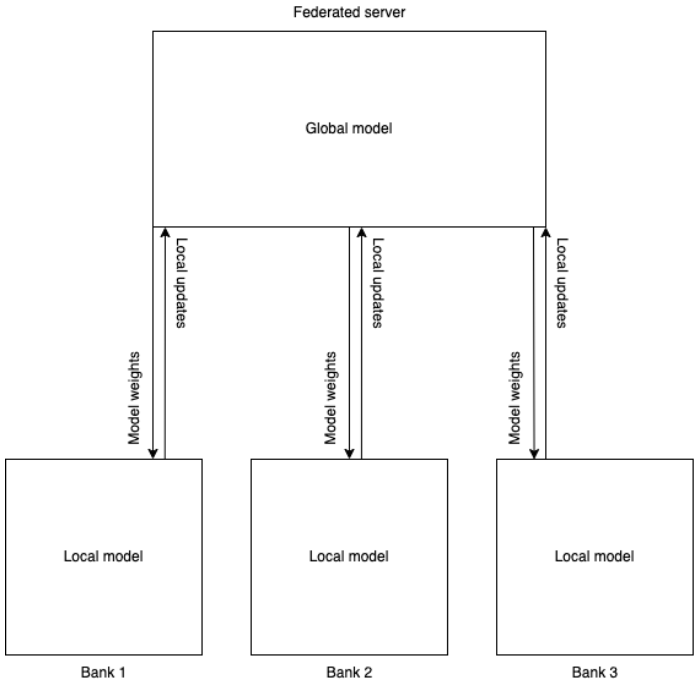

Fig. 2. Federated Learning Architecture

At the central server, the local updates are aggregated to improve the global model. The aggregation process involves computing an average of the local weights, where each bank's contribution is scaled by the number of samples it used during training. The aggregation is expressed by the formula [5]:

$$\omega_t = \sum_{k=1}^{K} \frac{n_k}{n} \omega_{t+1}^k \quad (1)$$

This approach ensures that the global model benefits from diverse data sources while preserving the privacy of each client's data. Before any updates are received, the server initializes the global model either with random parameters or, as in our approach, by pretraining on an available dataset. Once the initial global model is distributed to the clients, each institution trains locally and sends back only its updated model parameters.

On the client side, each institution is responsible to train a local model on its own dataset and send the model weights to the server. This process is important because only the weights are shared, the data remains private.

B. XGBoost Model

The machine learning model chosen for this study was XGBoost. Each financial institution trains locally a model using the

XGBoost classifier, a gradient boosting framework effective with tabular data and robust in handling class imbalance. To ensure the optimal selection of hyperparameters, we integrated GridSearchCV. The grid included variations in tree depth, learning rates and number of estimators. In this way, different combinations of these parameters were evaluated based on their performance, measured by F1-score. This exhaustive search ensured that the selected hyperparameters were the best suited for handling the class imbalance challenge common in financial datasets.

C. Neural Network Model

Our research employs a neural network model, a model well-suited for handling the imbalanced nature of financial frauds datasets.

The network consists of multiple layers:

a. **Input Layer**: The input layer has two major responsibilities: to receive the data and to properly format its features, ensuring that the next layers receive standardized inputs. The job of the input layer is critical because it builds the foundation for accurate pattern recognition throughout the network.

b. **Hidden Layers**: The hidden layers are the core of the network. The learning happens here. Our neural network implements three hidden layers, each with a progressively reduced number of neurons, from 128 neurons in the first layer down to 32 neurons in the final layer. Each hidden layer is followed by batch normalization, which improves the network's training stability, and ReLU, for improving the catch of complex patterns in the data. Moreover, we apply dropout regularization at a rate of 50% after each hidden layer to avoid the risk of overfitting.

c. **Output Layer**: The output layer is responsible for preparing the data for getting out of the network. The output layer's job is to synthetize the information learned by the hidden layers into a final decision. The final input produces a logit, the raw output value, corresponding to the log-odds of the positive class. To get a probability score ranging between 0 and 1, the logit is passed through a sigmoid function. The probability score decides if the transaction is classified as fraudulent or legitimate.

To handle one of the biggest challenges of financial fraud detection, the highly imbalanced nature of datasets, we integrated a focal loss function. By doing this, the model can concentrate more on the minority class, the fraudulent cases, rather than being overly influenced by the majority class. In this way, the overall performance of the model is improved by learning more effectively from the minority class. Mathematically, the focal loss function is expressed as [19]:

$$FL(p_t) = -\propto_t (1 - p_t)^{\gamma} \log(p_t) \quad (2)$$

5 Results and discussions

This section presents the results of our study. The discussion about each model's performance revolves around five key metrics: accuracy, precision, recall, F1 score and AUC-PR [20]. Moreover, for a better evaluation, we interpreted the confusion matrices.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F1\ score = \frac{2*Precision*Recall}{Precision+Recall} \quad (6)$$

Method	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	AUC-PR (%)
XGBoost	99.91	73.88	73.58	73.73	71.88
Neural Networks	99.94	85.13	80.28	82.64	76.08

Fig. 3. Results

Fig. 3 presents the results obtained by each model in the federated learning framework. The results and its interpretations for the XGBoost and neural network models are discussed individually below.

A. XGBoost model

The high accuracy of 0.9991 achieved by the XGBoost model indicates that almost all predictions were correct. The percent of positive predictions is around 74% indicated by the precision of 0.7388. This also shows us that the percent of false positives is pretty high, around 26%. The recall value of 0.7358 obtained by this study shows us that 74% of the true positive instances were detected, while 26% of them went undetected. The F1 score of 0.7373 shows us the trade-off between precision and recall, highlighting both the model's strength in overall prediction accuracy and its limitations in precisely capturing all relevant cases. The model obtained an AUC-PR score of 0.7188. This suggests that it has a moderate balance between precision and recall across different thresholds.

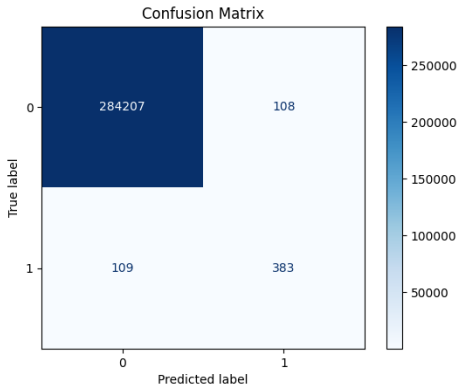

Fig. 4. XGBoost - confusion matrix

The confusion matrix presented in Fig. 4 illustrates the number of instances classified by the XGBoost model as true negatives, false negatives, true positives and false positives. The model performed well in detecting non fraudulent transactions, with 284.207 instances classified as true negatives and only 108 false negatives. For positive cases, the model demonstrated worse performance, with 383 true positives, but 109 false positives. This indicates that around 22% of fraudulent transactions were not correctly identified by the model.

B. Neural Network Model

As shown in Fig. 3, the Neural Network model's accuracy has a value of 0.9994, meaning that nearly all predictions were correct. The precision of 0.8513 indicates that approximately 85% of positive predictions were correct, while the the rest of 15% were false positives. Similarly, the recall of 0.8028 means that the model correctly identified about 80% of the true positive instances, indicating that a portion of actual positives went undetected. The F1 score of 0.8264, which balances precision and recall, illustrates this trade-off, underscoring the model's effectiveness in overall prediction while also highlighting its limitations in accurately identifying all relevant instances. The AUC-PR of 0.7608 highlights that the Neural Network model achieves a stronger level of precision across various recall thresholds than the XGBoost model.

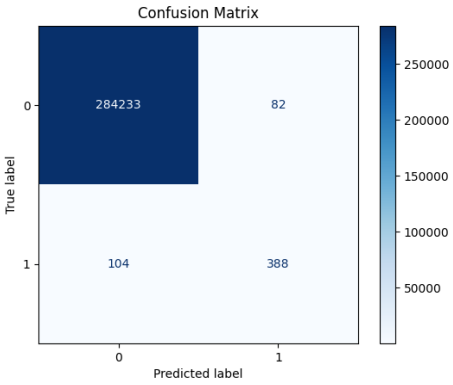

Fig. 5. Neural Network - confusion matrix

The confusion matrix represented by Fig. 5 shows the results obtained by the neural

network model. This matrix shows 284.233 correctly classified negative cases and 388 correctly identified positive cases. Compared to the XGBoost model, the false positives have decreased to 82, suggesting the model is more precise in avoiding incorrect positive classifications. Similarly, false negatives have reduced slightly to 104, enhancing the model's sensitivity by detecting a higher proportion of actual positives.

The results of this comparative study can serve as a starting point for developing a real-world fraud detection system based on the federated learning framework. Given its better performance, the neural network can be the primary candidate model for building such a system. The main limitation encountered during this study was the lack of sufficient data, as the confidentiality requirements restrict the availability of public financial datasets. Having multiple distinct datasets, rather than dividing one into several subsets, would likely have led to improved results. Future research can focus on extending the federated learning approach to other types of financial fraud, including insurance fraud, loan fraud or money laundering. Such studies would validate the applicability and effectiveness of federated learning framework across a broader spectrum of financial crime detection scenarios.

6. Conclusions

In this study we evaluated the performance of both XGBoost and neural network models within a federated learning model. In terms of accuracy, both models performed exceptionally. In key performance metrics such as precision, accuracy, F1 score, and AUC-PR score the neural network model outperformed the XGBoost model. While in previous studies, XGBoost had the best performance when compared to other machine learning models, the deep learning model demonstrated a better ability in detecting fraudulent transactions, making it a promising choice for real-world cases. Overall, this study highlights the potential of federated learning to improve collaboration among financial institutions to develop better fraud detection systems.

References

[1] K. S. Amit, "An Overview of Digital Payment Frauds: Causes, Consequences, and Countermeasures," *Journal of Informatics Education and Research*, vol. 5, no. 1, pp. 2297-2311, 2025.

[2] Yukun, "The Impact of Financial Fraud on Economic Stability: An Extensive Economic Analysis," *Journal of Internet Banking and Commerce*, vol. 28, no. 4, 2023.

[3] Y. Wensi, Z. Yuhang, Y. Kejiang, L. Li and X. Cheng-Zhong, "FFD: A Federated Learning Based Method for Credit Card Fraud Detection," *Big Data – BigData 2019*, vol. 11514, pp. 21-31, 2019.

[4] M. M. Priyanka, "Federated Learning: Opportunities and Challenges," in *Proceedings of ACM Conference (Conference'17)*, New York, 2021.

[5] M. Brendan, M. Eider, R. Daniel, H. Seth and A. y. A. Balise, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, Florida, 2017.

[6] M. Prakash, "Federated Learning for Cross-Bank Fraud Defense," *Intenational Journal of Computer Engineering & Technology*, vol. 16, no. 2, pp. 221-241, 2025.

[7] G. Poonam, H. Jayesh and K. Prakash, "Federated Learning in Banking: A Secure and Intelligent Approach to Financial Fraud Detection," *Journal of Emerging Technologies and Innovative Research*, vol. 12, no. 6, pp. 629-636, 2025.

[8] T. Nguyen, S. Kai, W. Siyao, G. Florian and G. Yike, "Privacy

preservation in federated learning: An insightful survey from the GDPR perspective," *Computers & Security,* vol. 110, pp. 1-23, 2021.

[9] Z. Han, "Federated Learning-Based Credit Card Fraud Detection: A Comparative Analysis of Advanced Machine Learning Models," *ITM Web of Conferences,* vol. 70, pp. 1-6, 2025.

[10] A.Abdulalem, A. R. Shukor, H. O. Siti, A. E. E. Taiseer, A.-D. Arafat, N. Maged, E. Tusneem, E. Hashim and S. Abdu, "Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review," *Machine Learning for Cybersecurity Threats, Challenges, and Opportunities II* , pp. 1-24, 26 September 2022.

[11] R. N. Teuku, M. I. Ghalieb, M. Aga, H. Irsan, S. R. Edi and I. Rinaldi, "Credit Card Fraud Detection for Contemporary Financial Management Using XGBoost-Driven Machine Learning and Data Augmentation Techniques," *Indatu Journal of Management,* vol. 1, no. 1, pp. 29-35, 2023.

[12] K. M. Krishna, Z. K. Mohammad and I. Ajay, "Credit Card Fraud Prediction Using XGBoost: An Ensemble Learning Approach," *International Journal of Information Retrieval Research (IJIRR),* vol. 12, no. 2, pp. 1-17, 2022.

[13] P. Raghavendra and S. Lokesh, "Credit Card Fraud Detection Using Neural Network," *International Journal of Students' Research in Technology & Management,* vol. 2, no. 2, pp. 84-88, 2015.

[14] R. Asha and K. K. Suresh, "Credit card fraud detection using artificial neural network," *Global Transitions Proceedings,* vol. 2, no. 1, pp. 35-41, 2021.

[15] K. A. Saif, J. A. Saif, D. Hussain and A. K. Muhammad, "Secure and Transparent Banking: Explainable AI-Driven Federated Learning Model for Financial Fraud Detection," *Journal of Risk and Financial Management,* vol. 18, no. 4, pp. 1-26, 2025.

[16] Y. Qiang, L. Yang, C. Tianjian and T. Yongxin, "Federated Machine Learning: Concept and Applications," *ACM Transactions on Intelligent Systems and Technology* , vol. 10, no. 2, pp. 1-19, 2019.

[17] Tomisin, M. S. Raj and P. Bernardi, "Transparency and Privacy: The Role of Explainable AI and Federated Learning in Financial Fraud Detection," *IEEE Access,* vol. 12, pp. 64551-64560, 2024.

[18] M. L. G. -. ULB, "Credit Card Fraud Detection Dataset," Kaggle, 2018. [Online]. Available: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud. [Accessed 4 February 2025].

[19] L. Tsung-Yi, G. Priya, G. Ross, H. Kaiming and D. Piotr, "Focal Loss for Dense Object Detection," *IEEE International Conference on Computer Vision (ICCV)* , pp. 2999-3007, 2017.

[20] N. M.Z and A. Amir, "Insights into Performance Fitness and Error Metrics for Machine Learning," *Architecture, Structures and Construction,* pp. 1-19, 2020.

[21] M. Samaneh, B. Ali, S. Sima and F. Francesco, "Balancing privacy and performance in federated learning: A systematic literature review on methods and metrics," *Journal of Parallel and Distributed Computing,* vol. 192, 2024.

Șener Ali is a master's student at the Academy of Economic Studies in Bucharest, where he is expected to graduate in 2025. He currently works as a web developer, and his main areas of interest include web development, artificial intelligence, and blockchain.