Advancing Credit Risk Management in Open Banking with Enhanced Federated Averaging Algorithm

Adil Oualid
MISI Laboratory, Hassan 1 University, FST Settat, Morocco
a.oualid@uhp.ac.ma (corresponding author)

Youssef Qasmaoui
Hassan 1 University, FST Settat, Morocco
qasmaoui@gmail.com

Youssef Balouki
MISI Laboratory, Hassan 1 University, FST Settat, Morocco
y.balouk@uhp.ac.ma

Bouzgarne Itri
2IACS, ENSET Mohammedia, Hassan II University of Casablanca, Casablanca, Morocco
bouzgarne.itri@gmail.com

Lahcen Moumoun
MISI Laboratory, Hassan 1 University, FST Settat, Morocco
l.moumoun@uhp.ac.ma

Received: 16 January 2025 | Revised: 9 February 2025 and 12 February 2025 | Accepted: 14 February 2025

 DOI: https://doi.org/10.48084/etasr.10266

ABSTRACT

This paper addresses the pressing challenge of credit risk management in contemporary banking by integrating Federated Learning (FL) and Open Banking, employing an Enhanced Federated Averaging (FedEn) algorithm. Against Open Banking's transformative impact on financial services, the current research responds to the critical need for improved credit risk assessment in Non-Independently and Identically Distributed (Non- IID) data landscapes. The integration of FL and Open Banking is showcased by applying the Federated Averaging (FedAvg) algorithm, which offers a novel framework for credit risk management. The proposed methodology, grounded in theoretical foundations and validated through practical case studies, underscores the effectiveness of this integrated approach. The main contribution of the present work lies in demonstrating that the synergy of FL and Open Banking, facilitated by FedAvg, significantly enhances credit risk prediction accuracy while ensuring robust data privacy. Despite data security and regulatory compliance challenges, this integration presents a promising direction for financial institutions. The current research contributes through a comprehensive understanding of these technologies' confluence, providing valuable insights for banks, policymakers, and researchers navigating the dynamic landscape of credit risk management in the era of Open Banking.

Keywords-federated learning; credit risk management; open banking; privacy preservation; non-IID data; model aggregation

I. INTRODUCTION

The advent of Open Banking, marked by the transparent exchange of financial data through Application Programming Interfaces (APIs), has reshaped the landscape of financial services [1]. This transformative paradigm offers unparallel transparency and customer-centric solutions. Concurrently, FL, putting forward a decentralized model training approach, presents a novel avenue for collaborative machine learning across disparate devices while preserving data privacy.

A wide range of critical challenges in dealing with non-IID data in FL [2] need to be addressed. Strategies for handling different types of data heterogeneity, including Group-level and Client-level personalization, are crucial. Additionally, the lack of standardized benchmarks representing various non-IID scenarios hampers the practical evaluation of the proposed methods. Furthermore, efforts to categorize works based on their approach to non-IID data and the specific type of heterogeneity they address are essential for advancing this field.

This research seeks to bridge Open Banking and FL, incorporating the FedAvg algorithm [3]. The unique synergy of Open Banking and FL, highlighted by FedAvg, proposes a groundbreaking framework for enhanced credit risk management. Amidst the evolving financial landscape characterized by diverse and non-uniform data distributions, traditional credit risk models face challenges in accuracy and robustness [4]. The present study addresses this critical gap by exploring how the collaborative power of FL, guided by the FedAvg algorithm, can augment the predictive capabilities of credit risk assessments in non-IID settings. The FL process unfolds in two distinct components: the server (coordinator) and the nodes (participants), interconnected through a carefully designed mechanism. Each participant, denoted as i, engages in training a local model L_i using its individual dataset $D_i = \{(x_i, y_i)\}$. This local model commences with an initialization based on a globally shared model parameter W and undergoes fine-tuning with the data from node i, resulting in localized parameters W_i.

The coordinator in FL orchestrates the learning of a global model controlled by w , intended for sharing across all participants distributed on nodes. Through iterative communication rounds, the global model undergoes gradual refinement to better align with the diverse participant datasets. The ultimate global model represents an optimal solution for each participant in further tasks. Specifically, the optimal global model aims to minimize the cumulative loss across all participants, expressed as:

$$\sum_{i=1}^{n} p_i \cdot L\,(D_i, w) = \sum_{i=1}^{n} p_i \cdot L_i \quad (1)$$

where $L(.)$ denotes the loss function for each participant's learning task, w signifies the model parameters, and p_i represents the weight reflecting each node's importance. Determining p_i typically considers the size of the node's dataset $|D_i|$, ensuring that each instance, irrespective of its location or data owner, contributes equally to the overall loss. At times, L_i is employed as a concise representation of $L(D_i, w)$.

The integration of Open Banking and FL represents a significant advancement in credit risk management, facilitated by the utilization of transparency and data privacy to improve financial decision-making. Open banking facilitates secure financial data exchange via APIs, while FL permits collaborative machine learning among various decentralized institutions without jeopardizing sensitive information.

Authors in [5] introduced Federated Meta-Learning for the detection of fraudulent credit card transactions, enhancing conventional FL techniques with meta-learning capabilities. Unlike conventional FedAvg models, this design employs a Meta-Learning Classifier, incorporating supplementary layers for feature extraction and relational modeling. The impact of computational overhead from meta-learning on real-world implementation in resource-constrained financial systems remains unclear, despite its intriguing novelty in fraud detection. Authors in [6] examine a centralized FL model that facilitates collaboration among credit bureaus and financial institutions. It was founded on FedAvg; however, due to its constraints concerning data and system heterogeneity, some researchers have implemented various techniques, such as the proximal term in FedProx [7], to mitigate the divergence among client updates. Per-FedAvg [8] enhances the methodology by employing a multi-task meta-learning framework, known as MAML, to optimize model customization for individual clients. Although these techniques enhance personalization, they may incur higher computational costs and necessitate meticulous hyperparameter tuning to ensure fairness among institutions. Q-FedAvg [9] promotes equity by assigning a greater aggregation weight to inferior models, thereby mitigating performance disparities. Nonetheless, due to the significant variability in the quality and quantity of data collected by various financial institutions, its practical applicability remains ambiguous. Ultimately, in addition to the aforementioned work, SimFL [10] introduced FL for gradient boosting decision trees, while authors in [11] formulated Fed2Codl, a co-distillation-based FL method that aligns local models with a global model. While Fed2Codl facilitates knowledge sharing, its dependence on probability harmonization presents significant concerns about model consistency among institutions with markedly imbalanced datasets.

Notwithstanding these advancements, the literature predominantly emphasizes algorithmic enhancements while neglecting a comprehensive examination of the implications for Open Banking. The convergence of these two fields is still fully examined, particularly concerning regulatory compliance, implementation obstacles, and financial data variability. Moreover, the majority of studies prioritize enhancements to technical models, overlooking practical deployment considerations, such as interoperability among banking APIs and federated systems. The present research addresses this gap by examining the interplay between Open Banking and FL, specifically concentrating on FedAvg. A horizontal FL model is proposed that aligns with the principles of Open Banking to ensure practical applicability. This work offers insights regarding its highly practical impact for financial institutions, policymakers, and researchers as they traverse the changing landscape through an empirical evaluation of model efficacy and financial data integrity [12].

The present study makes a significant contribution by integrating FL and Open Banking, employing FedEn algorithm, to address the crucial challenge of credit risk management in contemporary banking. Its main contributions are:

- A novel framework by seamlessly integrating FL and Open Banking is proposed.
- The challenge of Non-IID data landscapes in credit risk assessment is adequately addressed.

- The challenges in Non-IID data landscapes, leading to significantly improved credit risk prediction accuracy, are also addressed.

II. METHODOLOGY

A. The Proposed Model

The conventional FedAvg algorithm, while effective in aggregating model updates from distributed clients, may face limitations when dealing with non-IID data distribution, where data across clients exhibit varying statistical properties. For this reason, an enhanced FedAvg algorithm tailored for credit risk management applications in Open Banking environments is proposed to tackle this issue. The introduced enhanced algorithm integrates privacy-preserving mechanisms to ensure data confidentiality while facilitating collaborative model training across multiple financial institutions. By incorporating differential privacy and secure multi-party computation techniques into the aggregation process, this study aims to mitigate privacy concerns associated with sharing sensitive financial data.

Additionally, personalized model updates are introduced to adapt to the heterogeneity of client data distributions commonly observed in credit risk management scenarios. Leveraging techniques, such as multi-task meta-learning, the proposed algorithm allows for personalized model training on individual client datasets, thereby enhancing the robustness and accuracy of the federated model [13, 14]. The proposed architecture for the FedAvg-based Enhanced Credit Risk Management in Non-IID (ECRM-N) system consists of the following components:

Client Nodes: Each client node represents a financial institution participating in the FL process. Client nodes hold proprietary credit risk data collected from their respective customer bases.

Server Node: The central server node coordinates the FL process and aggregates model updates from client nodes. It also manages the distribution of global model parameters to client nodes for training.

Privacy-Preserving Mechanism: This component ensures the confidentiality of sensitive credit risk data during the aggregation process. Techniques, such as differential privacy or secure multi-party computation, aggregate model updates while preserving data privacy [15].

Personalized Model Training Module: This module enables personalized model updates tailored to the unique data distributions of individual client nodes. Techniques, such as multi-task and meta-learning, adapt the FL to heterogeneous client datasets.

Evaluation and Monitoring Module: This module evaluates the performance of the federated model on test data and monitors model convergence during the training process. It provides insights into the effectiveness of the FL approach for credit risk management in non-IID data settings.

The Enhanced Federated Averaging for Open Banking (EFAB) algorithm introduces novel enhancements to the conventional FedAvg approach to address the specific challenges encountered in credit risk management within open banking environments. At each FL round, the server initializes the global model parameters and randomly selects a subset of clients to participate in training. Concurrently, each client updates its local model using its data through the ClientUpdate function, where the data are divided into batches for local training over multiple epochs. The aggregation of model updates is performed on the server using a privacy-preserving mechanism, such as differential privacy or secure multi-party computation, ensuring the confidentiality of sensitive financial data. This aggregated model update is then returned to the server for further iterations, enabling collaborative model training while safeguarding data privacy in open banking.

```
Algorithm 1. Enhanced Federated Averaging
for Open Banking
Server executes:
initialize w_0
for each round t = 1, 2, ... do
      m ← max(C, K, 1)
      S_t ← random subset of m clients
      for each client k ∈ S_t in parallel
do
      w_k_t+1 ← ClientUpdate(k, w_t)
      w_t+1 ← AggregateAndUpdate(w_k_t+1)
// Aggregation with privacy-preserving
mechanism
      // Run on client k
ClientUpdate(k, w):
      B ← split client data into batches
of size B
      for each local epoch i from 1 to E
do
      for batch b ∈ B do
            w ← w - η∇(l(w; b)) // Local
training
      return w to server
AggregateAndUpdate(w_k_t+1):
      // Apply differential privacy or
secure multi-party computation for
aggregation
      w_t+1 ←
PrivacyPreservingAggregation(w_k_t+1)
      return w_t+1 to server
```

B. Data Sharing

In the context of the proposed data-sharing strategy for EFA within the realm of Open Banking, a respective framework is shown in Figure 2. Initially, a globally shared dataset (G) is utilized, encompassing data from Taiwan Credit Dataset [16], Give Me Some Credit (GMSC) [17], and Home Credit (HC) [18], which is shared publicly. This dataset is centralized in the cloud and contains credit-related information. During the initialization phase of EFA, a preliminary model trained on G, along with a random fraction α of G, is distributed to each client. Each client's local model is subsequently trained on these shared data from G in combination with their private data. The cloud then aggregates the local models from all clients to train a global model using EFA. This strategy involves balancing two key trade-offs: first, between test accuracy and the size of G (β), defined as the ratio of ‖G‖ to ‖D‖, where D represents the total client data; and second, between test accuracy and α, the fraction of G distributed to each client. This study's experiments, conducted on the Give Me Some Credit dataset, explore these trade-offs by dividing the training set into client data (D) and a holdout set (H), using different subsets of G to assess their impact on test accuracy. The findings indicate significant improvements in test accuracy with increased β, and suggest that distributing only a portion of G to clients can achieve comparable accuracy, providing valuable insights for optimizing data distribution strategies in EFA for Open Banking. Figure 1 depicts the data sharing strategy.

C. Experimental Design

This section outlines the experimental design, detailing the datasets used and the performance measures employed to assess the efficacy of FL algorithms.

1) Datasets and Model

To ensure a comprehensive evaluation, three credit datasets varying in size and imbalance ratios, are utilized as follows: Taiwan Credit Dataset (TCD), Give Me Some Credit (GMSC), and Home Credit (HC). The characteristics of these datasets, including the number of samples, features, and imbalance ratios, are summarized in Table II. The Taiwan dataset from the UCI machine learning repository includes 23,364 non-default samples and 6636 default samples, with 23 features per sample. The GMSC and HC datasets, sourced from Kaggle competitions, consist of varying numbers of non-default and default samples, with different feature compositions. To standardize the datasets for analysis, preprocessing techniques, such as normalization, one-hot encoding, and correlation analysis were deployed. Continuous features were normalized to a range of [0, 1]. Categorical features were converted into binary features using one-hot encoding. Features with high correlation coefficients (> 0.97) were removed to reduce multicollinearity. Post-preprocessing, the feature sets' dimensions were adjusted accordingly.

2) Performance Measures

In assessing the introduced FL model for credit scoring, it is imperative to employ robust evaluation metrics that effectively capture the model's performance across various dimensions. The chosen metrics provide insights into the model's predictive accuracy, its ability to handle imbalanced datasets, and the impact of privacy-preserving techniques on model performance [19]. The following key evaluation metrics were employed:

Accuracy: Accuracy remains a fundamental metric, representing the proportion of correctly classified instances. In the context of credit scoring, accuracy indicates the model's correctness in predicting creditworthiness.

Precision and Recall: Precision and recall are crucial in assessing the model's performance concerning false positives and false negatives. Precision measures the accuracy of positive predictions, while recall gauges the ability to capture all positive instances. In credit scoring, precision is valuable for minimizing false approvals, while recall is critical for identifying as many creditworthy individuals as possible [19].

F1 Score: The F1 score, which harmonizes precision and recall, is a composite metric that balances the trade-off between false positives and false negatives. It provides a comprehensive assessment of the model's overall performance, especially in scenarios where imbalanced classes are prevalent.

KS (Kolmogorov-Smirnov) Metric: The KS metric is a non-parametric test that computes the maximal distance between the cumulative distribution functions of two distinct groups—in this case, the predicted scores for bad (events) and good (non-events) credit outcomes. By computing the maximal absolute difference between CDFs, the KS statistic provides an intuitive measure of a model's discriminative power. A higher KS value means that the two groups are better separated, and hence is indicative of better model performance in discriminating between defaulters and non-defaulters.

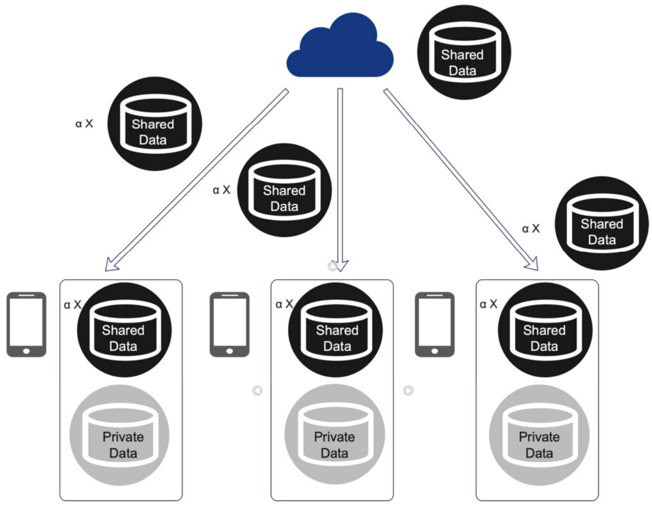

Fig. 1. Data sharing strategy

3) Evaluation Approach

In evaluating the models, the study primarily focused on comparing the performance of local models against global models optimized implementing the FedAvg algorithm. The key metric used for this comparison was the F1-score, a balanced measure of a model's precision and recall. This metric was chosen due to its relevance in assessing the accuracy of models in classification tasks [19].

Throughout 100 simulations, both the local and global models were evaluated on a test dataset to ascertain their respective performance levels. The results were visually represented through histograms, with the global models' performance being depicted in orange and the local models' in blue. Notably, a statistical analysis, specifically a t-test, was conducted to determine the significance of the performance differences observed. The analysis revealed that, on average, the global models outperformed the local models with statistical significance, reinforcing the efficacy of the FedAvg algorithm in enhancing model performance.

Furthermore, a more nuanced analysis was undertaken, focusing on the top-performing models from each simulation. This examination indicated that the best-performing global models were on par with the best local models, suggesting that while global models generally outperform, the top-tier local models still hold substantial value [20].

In a separate but related experiment, the study thoroughly investigated the performance of the FedAvg algorithm in non-IID settings. This involved evaluating the F1-score performance of FedAvg across various levels of non-IID data skew, ranging from 0 to 0.99. The experiment aimed to assess both the best and average F1-scores for locally trained models under these conditions. This part of the study was critical in understanding the robustness and adaptability of the FedAvg algorithm when dealing with heterogeneous data distributions, a common challenge in FL environments.

III. RESULTS AND DISCUSSION

The experimental framework involved the utilization of PyTorch Version 2.2.1 to emulate a federated scenario with a single server and K clients, specifically focusing on the Non-IID data distribution setting. To achieve this, the present study randomly disrupted and partitioned the original dataset into K equal segments, representing the clients. This partitioning may lead to an incomplete label space for each client, particularly in datasets with significant class imbalances. In all experiments, 30% of the clients were randomly chosen to participate across all rounds, with the selected clients remaining consistent for fair comparison among methods. Each round employed 10 passes for local updating, using cross-entropy as the local loss. A Stochastic Gradient Descent (SGD) solver with a constant learning rate of 0.5 and momentum of 0.5 for all FL methods was employed. A threshold of 0.5 was set, in line with common practice in credit scoring. The proposed FedEnh utilized Mean-Squared Error (MSE) as the distillation loss, with a temperature parameter of 1 due to the binary labels in credit datasets. The relative weight parameter λ in Fedenh was set to 0.3 to prioritize the distillation term over the loss term. The performance results were averaged over 10 runs with different random seeds, representing the average performance across all client test sets.

Figures 2 and 3 provide a comprehensive analysis of the performance and stability of various FL methods across three credit datasets: TCD, GMSC, and HC. In Figure 2, which illustrates the accuracy comparison across these datasets, it is evident that the proposed method, FedEnh, consistently achieves higher accuracy levels compared to the benchmark methods (FedAvg, FedProx, and FedCodl). Notably, FedEnh demonstrates superior performance, particularly on the GMSC dataset, where it achieves the highest accuracy.

Figure 3 showcases the standard deviation of accuracy for each method across the same datasets under Non-IID conditions. The results reveal that FedEnh maintains the lowest standard deviation across all datasets, indicating a more stable performance compared to the other methods. This highlights the robustness of FedEnh in mitigating accuracy fluctuations, which is crucial in real-world scenarios where data are often heterogeneous and imbalanced. The reduced variance in accuracy achieved by FedEnh underscores its effectiveness in balancing the transfer of knowledge between global and local models, thereby enhancing the generalization capability of the trained model. These findings affirm the efficacy of FedEnh in improving discrimination performance while addressing the challenges posed by heterogeneous credit data.

To thoroughly assess the efficacy of the proposed approach, a comparative analysis of its performance was conducted against benchmark federated methods (FedAvg, FedProx, and FedCodl), denoted as FedEnh, across three credit datasets over 50 communication rounds. The comparison results depicted in Figures 2 and 3 reveal the superior performance of FedEnh across both data distribution settings. Table I provides a comparative analysis of FL methods in Non-IID settings across different datasets. Performance measures, including Accuracy, Recall, F1-score, and KS, are presented for four FL methods: FedAvg, FedProx, FedCodl, and FedEnh. For the Taiwan dataset, FedEnh exhibits the highest Accuracy (82.34%), Recall (94.95%), F1-score (90.02%), and KS (44.56%). Similarly, for the GMSC dataset, FedEnh achieves superior performance in Accuracy (94.15%), Recall (98.85%), F1-score (96.51%), and KS (59.38%). Moreover, for the HC dataset, FedEnh demonstrates notable improvements in Accuracy (91.65%), Recall (98.94%), F1-score (95.52%), and KS (33.13%) compared to other methods. These findings underscore the efficacy of FedEnh in enhancing discrimination performance across diverse datasets in Non-IID settings.

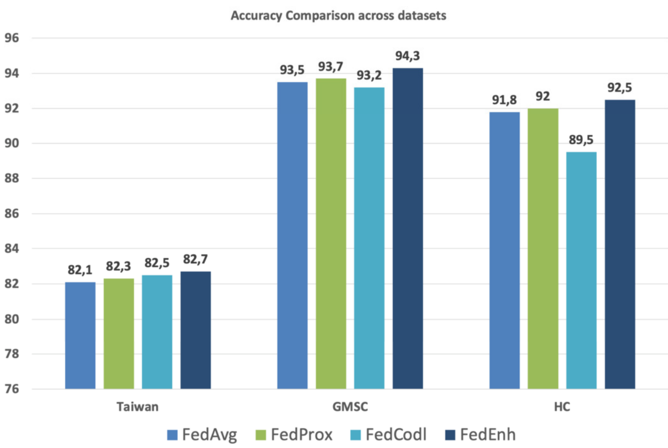

Fig. 2. Accuracy comparison across datasets.

Table II compares various machine learning methods across three distinct datasets: Taiwan, GMSC, and HC. Each method's performance is evaluated based on four key measures: Accuracy, Recall, F1-score, and KS. The methods compared include Logistic Regression (LR), Random Forest (RF), Extreme Gradient Boosting (XGB), and the proposed FedEnh method. FedEnh consistently demonstrates superior performance across all measures compared to LR, RF, and XGB for the Taiwan dataset. Similar trends are observed for the GMSC and HC datasets, highlighting FedEnh's efficacy in enhancing classification Accuracy, Recall, F1-score, and KS statistics values across diverse datasets and evaluation metrics.

FedEnh consistently outperforms LR, RF, and XGB across all datasets, showcasing significant enhancements in classification accuracy, recall rates, F1-scores, and KS statistics. Notably, FedEnh exhibits remarkable improvements ranging from +2.73% to +11.82% in Accuracy, +1.74% to +5.31% in Recall, +2.20% to +7.70% in F1-score, and +12.82% to +26.43% in KS compared to the baseline algorithms. These findings underscore the effectiveness of FedEnh in bolstering credit scoring performance in non-IID environments.

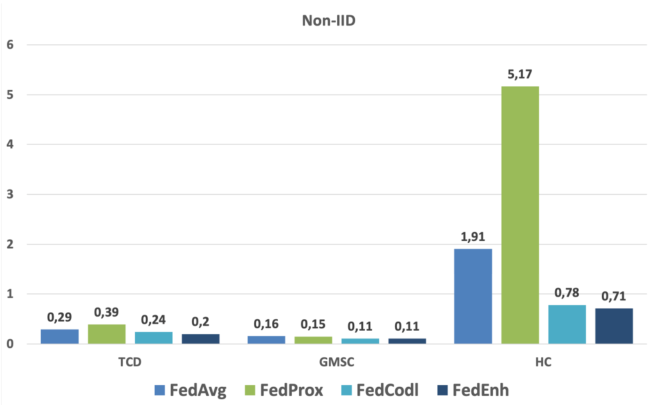

Fig. 3. Standard deviation of accuracy on three credit datasets.

TABLE I. COMPARATIVE ANALYSIS OF FL METHODS IN NON-IID SETTINGS

Datasets	Measures	Methods			
		FedAvg	FedProx	FedCodl	FedEnh
Taiwan	Accuracy	81.75	81.49	82.27	82.34
	Recall	94.06	94.43	94.74	94.95
	F1-score	88.99	88.88	89.33	90.02
	KS	42.26	41.95	44.15	44.56
GMSC	Accuracy	93.32	93.17	93.34	94.15
	Recall	98.43	98.73	98.78	98.85
	F1-score	96.51	96.41	96.50	96.51
	KS	59.34	58.89	59.12	59.38
HC	Accuracy	89.62	91.42	89.61	91.65
	Recall	96.27	98.94	96.84	98.94
	F1-score	94.44	95.50	94.37	95.52
	KS	26.18	29.91	30.67	33.13

TABLE II. COMPARATIVE ANALYSIS OF NON-FL METHODS AND FEDENH IN NON-IID ENVIRONMENTS

Datasets	Measures	Methods			
		LR	RF	XGB	FedEnh
Taiwan Dataset	Accuracy	79.61	79.43	78.43	82.34
	Recall	93.21	93.52	91.14	94.95
	F1-score	85.98	85.27	86.82	90.02
	KS	32.74	31.52	37.06	44.56
GMSC Dataset	Accuracy	92.56	92.38	92.00	94.15
	Recall	96.99	96.76	97.26	98.85
	F1-score	94.63	94.53	94.32	96.51
	KS	52.63	54.19	51.07	59.38
HC Dataset	Accuracy	87.74	87.96	87.56	91.65
	Recall	94.54	95.00	94.26	98.94
	F1-score	92.68	93.81	92.58	95.52
	KS	33.03	33.95	33.80	33.13

IV. CONCLUSION

This study presents a privacy-preserving framework utilizing horizontal Federated Learning (FL) for credit scoring. It is specifically engineered to address the complexities of Non-Independently and Identically Distributed (Non- IID) environments. The proposed approach extends prior works that rely solely on conventional federated averaging methods by incorporating knowledge transfer mechanisms, such as fine-tuning and knowledge distillation, to improve learning efficiency among distributed financial entities.

A significant contribution in the present research is the dual-focus optimization, which enhances the discrimination performance of federated models by addressing the frequently overlooked class imbalance issue in FL for credit risk assessment. Comparative assessments against non-federated techniques, namely Logistic Regression (LR), Random Forest (RF), Extreme Gradient Boosting (XGB), and leading federated algorithms indicate that the introduced method surpasses current approaches in predictive accuracy and robustness, particularly under conditions of highly skewed data distributions.

In contrast to Federated Averaging (FedAvg) and FedProx, which merely aggregate global updates without addressing heterogeneity, the proposed method utilizes knowledge transfer to enhance generalization among local models, making it more effective in the context of diverse financial data. Moreover, in contrast to FedCodl, which employs a complex co-distillation framework, the presented method yields superior outcomes while incorporating a remarkably straightforward yet effective knowledge-sharing mechanism.

The proposed framework additionally allows financial institutions to engage in data sharing that enables collaborative credit score modeling while safeguarding individual user privacy, albeit at a significant risk cost. This paves the way for the advancement of sophisticated data partitioning methods in this context, while seeking more adaptive heterogeneous distributions and necessitating comprehensive and meticulous communication during the theoretical overhead analysis, which facilitates broader FL opportunities in credit-risk management.

REFERENCES

[1] B. Alshawi, "Utilizing GANs for Credit Card Fraud Detection: A Comparison of Supervised Learning Algorithms," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12264–12270, Dec. 2023, https://doi.org/10.48084/etasr.6434.

[2] G. Luo, N. Chen, J. He, B. Jin, Z. Zhang, and Y. Li, "Privacy-preserving clustering federated learning for non-IID data," *Future Generation Computer Systems*, vol. 154, pp. 384–395, May 2024, https://doi.org/10.1016/j.future.2024.01.005.

[3] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, vol. 54, pp. 1273–1282, Apr. 2017, https://doi.org/10.48550/arXiv.1602.05629.

[4] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, "On the Convergence of FedAvg on Non-IID Data." *2020 International Conference on Learning Representations*, Jun. 25, 2020, https://doi.org/10.48550/arXiv.1907.02189.

[5] W. Zheng, L. Yan, C. Gou, and F.-Y. Wang, "Federated meta-learning for fraudulent credit card detection," in *Proceedings of the Twenty-Ninth*

International Joint Conference on Artificial Intelligence, pp. 4654–4660, Jan. 2021, https://doi.org/10.24963/ijcai.2020/642.

[6] A. Mora, A. Bujari, and P. Bellavista, "Enhancing generalization in Federated Learning with heterogeneous data: A comparative literature review," Future Gener. Comput. Syst., vol. 157, no. C, pp. 1–15, Aug. 2024, https://doi.org/10.1016/j.future.2024.03.027.

[7] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated Optimization in Heterogeneous Networks." *Proceedings of Machine learning and systems*, vol. 2, pp. 429–450, Apr. 2020, https://doi.org/10.48550/arXiv.1812.06127.

[8] A. Fallah, A. Mokhtari, and A. Ozdaglar, "Personalized Federated Learning: A Meta-Learning Approach." *34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, Oct. 2020, https://doi.org/10.48550/arXiv.2002.07948.

[9] T. Li, M. Sanjabi, A. Beirami, and V. Smith, "Fair Resource Allocation in Federated Learning." *The Eighth International Conference on Learning Representations*, Feb. 2020, https://doi.org/10.48550/arXiv.1905.10497.

[10] Q. Li, Z. Wen, and B. He, "Practical Federated Gradient Boosting Decision Trees," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 04, pp. 4642–4649, Apr. 2020, https://doi.org/10.1609/aaai.v34i04.5895.

[11] Z. Wang, J. Xiao, L. Wang, and J. Yao, "A novel federated learning approach with knowledge transfer for credit scoring," *Decision Support Systems*, vol. 177, Feb. 2024, Art. no. 114084, https://doi.org/10.1016/j.dss.2023.114084.

[12] G. Long, Y. Tan, J. Jiang, and C. Zhang, "Federated Learning for Open Banking." *Distributed, Parallel, and Cluster Computing*, Aug. 2021, https://doi.org/10.48550/arXiv.2108.10749.

[13] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. McMahan, S. Patel, and D. Ramage, "Practical Secure Aggregation for Privacy-Preserving Machine Learning," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, New York, NY, USA, pp. 1175–1191, Oct. 2017, https://doi.org/10.1145/3133956.3133982.

[14] L. Yang, J. Huang, W. Lin, and J. Cao, "Personalized Federated Learning on Non-IID Data via Group-based Meta-learning, " *ACM Trans. Knowl. Discov. Data*, vol. 17, no. 4, Mar. 2023, Art. no. 49:1-49:20, https://doi.org/10.1145/3558005.

[15] H. He, Z. Wang, H. Jain, C. Jiang, and S. Yang, "A privacy-preserving decentralized credit scoring method based on multi-party information, " *Decision Support Systems*, vol. 166, Mar. 2023, Art. no. 113910, https://doi.org/10.1016/j.dss.2022.113910.

[16] D. D. Nguyen, "Taiwan Credit Scoring. " Github Repository, Aug. 2021, [Online]. Available: https://github.com/doandongnguyen/TaiwanCreditScoring

[17] Credit Fusion and W. Cukierski, "Give Me Some Credit", Kaggle competition, Sep. 2011, [Online]. Available: https://kaggle.com/competitions/GiveMeSomeCredit.

[18] A. Montoya, Inversion, K. Odintsov, and M. Kotek, "Home Credit Default Risk", Kaggle competition, May 2018, [Online]. Available: https://www.kaggle.com/competitions/home-credit-default-risk.

[19] L. van der Maaten and G. Hinton, "Visualizing Data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579–2605, Nov. 2008.

[20] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning." *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, pp. 265–283, May 2016, https://doi.org/10.48550/arXiv.1605.08695.