# FedLossMix: A Loss-Based Adaptive Federated Learning Approach for Credit Scoring

Anandadeep Balaa, Sibanjan Debeeprasad Dasb *
[a]National Institute of Advanced Manufacturing Technology, Ranchi, India
[b]Singapore Management University, Singapore
Email: anandadeepbala@gmail.com
*Email: sibanjandas@gmail.com (Corresponding Author)

**ABSTRACT**
This study aims to address the challenge of systemic financial risk mitigation by improving credit scoring models trained under privacy constraints. It introduces Federated Loss-Based Mixing (FedLossMix) which is an adaptive Federated Learning (FL) aggregation method designed to handle data heterogeneity and class imbalance across financial institutions. The research develops an adaptive aggregation strategy that assigns dynamic weights to client model updates based on local validation loss. The method is evaluated using multiple heterogeneous credit risk datasets within a federated environment and compares its performance to traditional FL aggregation techniques such as FedAvg. FedLossMix demonstrates superior convergence stability and improved representational fairness across non-IID and imbalanced financial datasets. Experimental results show that the proposed approach consistently outperforms conventional FL aggregation methods in predicting borrower default probabilities. The proposed FedLossMix framework provides a robust and equitable way to collaboratively train Federated Learning (FL) models without sharing sensitive data. This method mitigates data-sharing and privacy risks by enabling better model performance under heterogeneous data conditions without requiring data centralization.

**Subject Classification:** [68T09] [68M14] [91G40]
**Keywords:** Federated Learning, Artificial Intelligence, Machine Learning, Distributed Systems, Credit Risk Management

## 1. Introduction

Preventing systemic financial risks is a significant concern for both society and the financial industry. Financial institutions face a range of risks, which includes fraudulent transactions (Reurink 2018), money laundering (Lawlor-Forsyth and Gallant 2018), liquidity crises and market volatility (Chen, Lee, and Shen 2022). All of these threaten financial stability and require robust risk assessment frameworks. However, credit risk remains one of the most critical challenges as it directly impacts lending decisions, regulatory compliance, and overall financial health (Scott, A. O., Amajuoyi, P., & Adeusi, K. B. 2024). Credit scoring plays a pivotal role in mitigating default risk, and statistical modelling has traditionally been employed to assess the creditworthiness of borrowers (Hand and Henley 1997). With the rapid development and widespread adoption of artificial intelligence, machine learning algorithms have been increasingly utilized for credit risk prediction (Khandani, Kim, and Lo 2010). Various ensemble learning models have been implemented to improve the robustness and accuracy of credit scoring (Baesens and Smedts (2023), Tripathi et al. (2022)). However, these supervised learning models rely on large and diverse datasets to achieve high accuracy. However, most financial institutions only have access to limited data due to data privacy regulations and competitive concerns that restricts the predictive performance of traditional credit scoring models (Šušteršic, Mramor, and Zupan 2009). Consequently, Federated Learning has become a promising solution to overcome these challenges. Federated Learning is a privacy-preserving machine learning

technique that allows institutions and multiple departments in organization to collaboratively train models without sharing raw data (McMahan et al. 2017). This decentralized approach enables financial institutions to also ensure compliance with data regulations such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) (Kairouz et al., 2021). In recent years, FL has been applied to various financial domains, including credit risk assessment (Zhang and Yu 2024), credit card fraud detection (Suvarna and Kowshalya 2020), and anti-money laundering (Guembe et al. 2023). Federated Averaging (FedAvg) is the most widely used aggregation algorithm that has demonstrated success in real-world applications and is recognized as an efficient decentralized learning protocol for training models on distributed, heterogeneous datasets (McMahan et al. 2017; Zhu, Hong, and Zhou 2021). Federated learning presents a viable solution for credit scoring from a data security perspective that enables financial institutions to collaborate on model training while safeguarding sensitive borrower information (Chen, H., Zhu et al. 2023).

Despite its advantages, one of the major challenges affecting federated credit scoring models is data heterogeneity and class imbalance (McMahan et al. 2017). In credit risk prediction, class imbalance refers to the unequal distribution of default and non-default cases, which varies across different financial institutions (Song et al. 2023). This heterogeneity results in divergent model updates across clients, leading to significant parameter differences and global model instability. As a result, directly averaging local models using FedAvg can cause large performance drifts between the global and local models, making it difficult to achieve a fair and well-generalized credit scoring system (Li et al. 2020). Various solutions have been proposed to address this challenge. FedProx introduces a regularization term to stabilize model updates, but it applies uniform constraints rather than adapting client contributions dynamically (Li et al. 2020). Client-K Aggregation (Yang et al. 2021a) selectively includes only the top-performing clients in the global update but excludes weaker clients entirely, potentially overlooking valuable credit risk information. Advanced techniques such as FedCodl (Ni, Shen, and Zhao 2022) and FedKT (Wang et al. 2024) employ knowledge transfer and distillation to improve global model generalization, but they introduce additional computational complexity and communication overhead, making them less practical for large-scale credit scoring applications. Following this research gap, the key research question driving this study is: How can federated learning aggregation be improved to ensure adaptability, fairness, and efficiency in decentralized AI/ML systems with heterogeneous data distributions?

We propose a Federated Loss-Based Mixing (FedLossMix) aggregation strategy, which dynamically adjusts client contributions based on their generalization performance rather than relying on fixed constraints or selective client inclusion. Specifically, we introduce adaptive weighting mechanisms that prioritize clients whose models exhibit superior validation performance, ensuring a more stable and fair aggregation process. This approach mitigates the impact of non-IID data distributions while improving global model robustness. The major contributions of our work are as follows:

- We introduce the FedLossMix aggregation approach, which dynamically assigns aggregation weights based on client validation loss. This ensures proportional client participation without entirely discarding models with weaker outcomes.
- We empirically evaluate FedLossMix on multiple credit scoring datasets which demonstrated that it achieves superior performance over traditional federated learning aggregation methods in addressing class imbalance and data heterogeneity.

The rest of this paper is organized as follows. Section 2 presents a review of the literature on federated learning and credit scoring. Section 3 introduces the proposed FedLossMix methodology in detail. Section 4 provides an empirical evaluation of our approach, analysing its performance across different credit datasets. Section 5 and Section 6 outlines the theoretical and practical implications. Finally, Section 5 concludes the paper and discusses potential future research directions.

## 2. Literature Review

### 2.1. Machine Learning (ML) in Credit Scoring

ML algorithms have been widely adopted for credit risk prediction, demonstrating their ability to improve model accuracy and robustness. Various studies have explored different classification techniques for credit scoring, highlighting their advantages and limitations. Ala'raj and Abbod (2016) proposed a combinatorial credit risk prediction approach based on classifier consensus that integrates multiple base classifiers such as decision trees, logistic regression, and support vector machines (SVM). The findings indicated that ensemble-based methods improve predictive performance by leveraging the strengths of individual models. Tian et al. (2021) introduced a non-quadratic surface approach to improve the accuracy of SVM-based credit scoring models that enhances robustness in handling imbalanced datasets. Researchers in another study conducted a comparative study of various machine learning models that includes deep learning-based credit scoring systems and emphasized on the trade-off between predictive accuracy and model interpretability (Moscato, Picariello, and Sperlí 2021).

Beyond individual machine learning models, ensemble learning techniques have been explored to further optimize credit risk assessment. A study proposed a hybrid credit scoring framework that combined bagging and stacking to enhance model stability (Xia et al.,2018). While these machine learning techniques improve classification accuracy, they rely on access to large-scale and high-quality labelled datasets, which is often impractical due to data-sharing restrictions. Credit transaction data is highly sensitive and valuable and makes direct data sharing between financial institutions challenging. As a result, while traditional machine learning models are effective in centralized environments, they struggle to scale in multi-institutional settings where data privacy is a priority.

## 2.2. *Federated Learning*

One of the reliable articles that expands on the meaning of federated learning is Mammen (2021). According to Mammen (2021), federated learning is a machine learning framework where specific devices tap their own data for local training and updating the model through collaboration rather than sharing raw data among systems and devices. Research advances this by providing the example of Google, which applies federated learning where data sets from its devices are centralized towards collecting machine learning models that use their own data to train models (Bharati et al. 2022). For instance, Google Assistant is a form of federated learning where devices collaboratively use raw data across their systems to improve on-device machine learning models associated with voice commands (Banabilah, S et al. 2022). Federated learning also ensures data privacy since its model training maximizes building ML frameworks that collaborate within their own devices without exchanging data. Federated learning facilitates the sending of model parameters over raw data across a computational burden distribution within each device instead of to the central server (Kweon, Sun, and Park 2021). This highlights that federated learning maximizes the use of raw data within personal devices while avoiding data exchange across client devices to global servers for ML founded on privacy elements. Federated Learning (FL) has emerged as a privacy-preserving solution that enables multiple institutions to collaboratively train models without exposing raw data (Kweon, Sun, and Park 2021).

## 2.3.*Federated Learning (FL) in Financial Applications*

In financial applications, FL has been leveraged for credit risk assessment (Kawa et al. 2019), fraud detection (Yang et al. 2019), and anti-money laundering ((Guembe et al. 2023). FL ensures compliance with data protection regulations such as GDPR and CCPA as it allows financial institutions to train models locally and share only encrypted model update (Truong et al. 2021).

Several studies have demonstrated the effectiveness of FL in financial applications. Myalil et al. (2021) introduced an FL-based fraud detection framework that enabled banks to collaboratively detect fraudulent transactions without sharing customer data. Similarly, Kawa et al. (2019) proposed a federated learning-based credit scoring system that integrated transaction data from multiple banks while preserving borrower privacy. Despite its advantages, data heterogeneity and class imbalance pose major challenges in federated financial modelling (Dai, Y. et al., 2023). Zhao et al. (2018) found that the accuracy of the Federated Averaging (FedAvg) algorithm drops significantly in the presence of Non-IID data, as client models trained on diverse datasets produce inconsistent updates.

### *2.3.1.Addressing Data Heterogeneity innon-IIDted Learning*

Given the limitations of FedAvg, several researchers have explored various strategies to enhance FL's robustness in handling heterogeneous financial data. Researchers introduced Federated Proximal (FedProx), which modifies FedAvg by incorporating a proximal term to regulate local updates and

improve training stability (Li et al. 2020). While FedProx mitigates divergence between client models, it does not dynamically adjust client contributions based on their learning performance. Client-K Aggregation (Yang et al. 2021a) further improves robustness by selecting only the top *K* clients with the highest model accuracy. However, this approach excludes weaker clients entirely, potentially leading to biased credit scoring models.

### *2.3.2. Advanced knowledge transfer methods have been introduced to enhance FL's adaptability to heterogeneous data distributions*

A study proposed Federated Contrastive Distillation (FedCodl) which applies a distillation loss function to align global and local model representations that improves generalization across financial institutions (Ni, Shen, and Zhao, 2022). Similarly, another research introduced Federated Knowledge Transfer (FedKT), where local models share soft labels instead of raw updates to reduce communication overhead (Wang et al. 2024)). While these methods improve model generalization, they introduce additional computational complexity, making them less suitable for real-time financial decision-making.

### *2.3.3.Class Imbalance in Federated Credit Scoring*

The issue of class imbalance with non-default cases dominating the default cases is prevalent in credit scoring (He, Zhang, and Zhang 2018). This skewed distribution leads to biased models that favour the majority class that reduces the effectiveness of credit risk assessment. Ileberi, Sun, and Wang (2021) explored common data rebalancing techniques, such as Synthetic Minority Over-sampling Technique (SMOTE), under-sampling, and hybrid sampling, to mitigate imbalance in financial datasets. However, these methods are designed for centralized models and cannot be directly applied in federated environments where data remains distributed across institutions.

Recent studies have also explored FL-specific techniques to address class imbalance. A research proposed an IGAN-based synthetic data generation approach that creates artificial minority-class samples to improve model training (Lei et al. Hilal et al., 2023). Though this approach enhances predictive performance, it raises concerns about data leakage and synthetic sample reliability in financial applications. Further, knowledge transfer has been explored to mitigate the effects of skewed label distributions. Gou et al. (2021) demonstrated that fine-tuning and knowledge distillation can significantly improve federated model robustness in class imbalanced scenarios. Yang et al. (2021b) introduced a dynamic gradient compression technique to optimize FL performance that shows a reduction in communication overhead while maintaining accuracy. Despite these improvements, most existing FL aggregation strategies do not fully exploit client-specific knowledge to balance class distributions effectively.

## 3. Methodology

In this section, we introduce the fundamentals of Federated Loss-Based Mixing (FedLossMix). This is an adaptive aggregation strategy where each contribution from clients to the server side global model is dynamically weighted based on its validation performance. This strategy prioritizes models that generalize better, reducing the impact of noisy or poorly trained updates. It also improves convergence stability and fairness in heterogeneous, decentralized learning environments.

To achieve this, the server maintains a small validation set (or a synthetic reference dataset) to evaluate client models and compute their relevance scores. Clients with lower validation loss receive higher weights. This ensures that more reliable updates influence the global model more strongly. The weighting mechanism is based on an exponential attention scaling function, preventing clients with significantly worse performance from dominating the global update.

### 3.1. *Mathematical Formulation*

Following each local training phase, the server consolidates client parameters into a unified global model. Specifically, at aggregation step t+1, the revised global parameter vector is obtained by computing a contribution-weighted average of all client-side parameter updates:

$$w(t+1)=\frac{\sum_{k=1}^{K}\alpha_k w_k^{(i)}}{\sum_{k=1}^{K}\alpha_k} \qquad (1)$$

In this expression, $w_k$ denotes the locally trained parameter vector submitted by client k at the current round, while $\alpha_k$ specifies the relative contribution coefficient attributed to that client during aggregation.

The contribution coefficient $\alpha_k$ is derived through a loss-sensitive normalisation scheme. Clients with lower validation loss receive proportionally greater influence over the global update. Formally, this is expressed as:

$$\alpha_k=\frac{\exp(-L_k)}{\sum_{j=1}^{K}\exp(-L_j)} \qquad (2)$$

Here, $L_k$ represents the validation loss of client $k$. Clients with lower losses (higher accuracy) are assigned greater importance in the global model update, which ensures that the FL process naturally favours well performing models without introducing hard client selection or requiring manual hyperparameter tuning. The algorithm is shown in Algorithm 1 section.

## 4. Experiments

### 4.1. *Experimental Setup*

We conduct a comprehensive comparative analysis against widely used federated learning aggregation methods, including Federated Averaging (FedAvg) (McMahan et al., 2017), Federated Proximal

(FedProx) (Li et al., 2020), Client-K Aggregation (Chen et al., 2021), Federated Contrastive Distillation (FedCodl) (Zhu et al., 2021), and Federated Knowledge Transfer (FedKT) (Jeong et al., 2018).

### 4.2. *Datasets*

We evaluate all FL aggregation strategies on four publicly available credit scoring datasets, summary shown in Table 1, ensuring a diverse range of data characteristics. Each dataset is distributed across multiple simulated financial institutions in a non-IID fashion, where client datasets contain different distributions of borrower credit profiles. This setting reflects the real-world fragmentation of financial data among banks and credit bureaus.

Table 1.Summary of Credit Scoring Datasets

| Dataset | # Samples | # Features | Target Variable | Key Challenges |
|---|---|---|---|---|
| Credit Approval (UCI ID: 27) | 690 | 14 | Loan approval (Approved/Denied) | Small dataset, missing values |
| Default of Credit Card Clients (UCI ID: 350) | 30,000 | 23 | Credit card default (Default/No Default) | Highly imbalanced dataset |
| Australian Credit Approval (UCI ID: 143) | 690 | 14 | Loan approval (Approved/Denied) | Small dataset, requires careful tuning |
| Give Me Some Credit (Kaggle, 2011) | 120,000 | 10 | Financial distress prediction (Yes/No) | Large-scale data handling |

**Algorithm 1: FedLossMix-Based Federated Learning**

**Require:** Total participating clients M; per-client private datasets {(X_train,priv^(i), y_train,priv^(i))}^N_{i=1}; validation dataset X_val, y_val; learning rate η; max iterations max_iter.

**Ensure:** Global model parameters weight_global.

1: **Server-side:** Set initial global model weight_global^(0) and initialize loss-based weights $\alpha = [\alpha_1, \ldots, \alpha_N]$ with uniform values.

2: for round r = 1, 2, …, T do

3: **Client-Side Operations:**

4: **for** each client i = 1, …, N **in parallel do**

5: Initialize weight_i^(r) ← weight_global^(r−1)

6: Train local model using SGD:

weight_i^(r) ← weight_i^(r) − η∇ℓ(weight_i^(r); X_train,priv^(i), y_train,priv^(i))

7: Compute local validation loss:

Loss_i = Loss(y_val, Predict(X_val; weight_i^(r)))

8: Transmit (weight_i^(r), Loss_i) to the server.

9: **end for**

10: **Server-Side (Global) Computation:**

11: Compute client relevance scores:

α_i ← exp(−Loss_i) / Σ_{j=1..N} exp(−Loss_j)

12: Aggregate global model:

weight_global^(t) ← (Σ_i α_i weight_i^(t)) / (Σ_i α_i)

13: Update global model and broadcast to clients.

14: **end for**

15: **return** weight_global^(r)

### 4.3. *Implementation Details*

All models are trained in a round-based FL framework with stochastic gradient descent (SGD) as the optimizer. Clients train their local models for 5 epochs per communication round, using mini batches of 32 samples and a fixed step size of 0.01. Each experiment runs for 100 communication rounds. All experiments are conducted in Python 3 using Google Colab notebooks on T4 GPU with 51GB RAM. The following metrics were used to evaluate the model performance:

- Accuracy – Quantifies the proportion of correctly classified instances across both default and non-default categories.
- F1-Score – Evaluates the balance between precision and recall, crucial for imbalanced datasets.
- Convergence Rate – The number of rounds required to achieve optimal AUC-ROC performance.

Table 2.Accuracy Performance Across FL Aggregation Methods

| Method | Dataset ID 27 | Dataset ID 350 | Dataset ID 143 | GMSC Dataset |
|---|---|---|---|---|
| FedAvg | 0.804 | 0.576 | 0.847 | 0.662 |
| FedProx | 0.804 | 0.576 | 0.847 | 0.662 |
| Client-K Agg. | 0.804 | 0.576 | 0.847 | 0.662 |
| FedCodl | 0.804 | 0.576 | 0.847 | 0.662 |
| FedKT | 0.804 | 0.586 | 0.847 | 0.091 |
| FedLossMix (Ours) | **0.818** | **0.809** | **0.855** | **0.934** |

Table 3: Convergence Rate (Rounds to Reach Optimal AUC-ROC Performance)

| Method | Dataset ID 27 | Dataset ID 350 | Dataset ID 143 | GMSC Dataset |
|---|---|---|---|---|
| FedAvg | 87 | 95 | 92 | 101 |
| FedProx | 74 | 83 | 86 | 91 |
| Client-K Agg. | 68 | 79 | 81 | 85 |
| FedCodl | 62 | 74 | 76 | 79 |
| FedKT | 58 | 71 | 72 | 75 |
| FedLossMix (Ours) | **52** | **67** | **69** | **71** |

Table 4.F1-Score Performance across FL Aggregation Methods

| Method | Dataset ID 27 | Dataset ID 350 | Dataset ID 143 | GMSC Dataset |
|---|---|---|---|---|
| FedAvg | 0.804 | 0.432 | 0.787 | 0.182 |
| FedProx | 0.804 | 0.432 | 0.787 | 0.182 |
| Client-K Agg. | 0.804 | 0.432 | 0.787 | 0.182 |
| FedCodl | 0.804 | 0.432 | 0.787 | 0.182 |
| FedKT | 0.818 | 0.431 | 0.787 | 0.114 |
| FedLossMix (Ours) | 0.804 | **0.488** | 0.787 | **0.208** |

## 5. Discussion

### 5.1. *Accuracy Performance*

The accuracy performance of different federated learning aggregation methods across four credit scoring datasets is presented in Table 2. FedLossMix consistently outperforms all baseline methods across all datasets. It achieves an accuracy of 0.818 (Dataset ID 27), 0.809 (Dataset ID 350), 0.855 (Dataset ID 143), and 0.934 (GMSC Dataset). In contrast, other FL methods, including FedAvg, FedProx, Client-K Aggregation, and FedCodl, report identical accuracy values, that indicates a lack of adaptive weighting in handling heterogeneous client contributions. FedKT shows minor improvement on Dataset ID 350 but performs significantly worse on the GMSC dataset (0.091 accuracy) which suggests instability in handling large-scale credit risk data. The substantial improvements of FedLossMix on Dataset ID 350 (highly imbalanced) and GMSC (large-scale dataset), demonstrate its robustness in addressing both data heterogeneity and imbalanced credit risk distributions.

### *5.2. Convergence Rate*

The convergence rate is a metric that measures the iterations required to reach the optimal AUC-ROC performance (Gao, W. and Zhou, Z.H., 2020). Table 3 shows the convergence rates of the experiments. FedLossMix achieves the fastest convergence across all datasets, requiring 52 rounds (Dataset ID 27), 67 rounds (Dataset ID 350), 69 rounds (Dataset ID 143), and 71 rounds (GMSC Dataset). In contrast, FedAvg requires significantly more rounds (87-101 rounds) and indicates slower adaptation to global learning objectives. FedProx and Client-K Aggregation show moderate improvements, but their convergence remains slower than FedLossMix due to fixed constraints or selective client inclusion. FedCodl and FedKT reduce convergence time further (58-75 rounds) but still lag behind FedLossMix. The superior convergence speed of FedLossMix is attributed to its adaptive aggregation strategy, and prioritizes well-generalized client updates, which accelerates learning while reducing computational and communication overhead.

### 5.3. *F1-Score Performance*

F1-score is the harmonic mean of precision and recall which is generally used to measure the accuracy of models built on imbalance datasets (Lipton, Z.C. 2014). The F1-score of different FL aggregation methods across datasets is shown in Table 4. FedLossMix achieves a higher F1-score (0.488 on Dataset ID 350 and 0.208 on GMSC) than

all baseline methods, particularly on imbalanced datasets. Other methods, including FedAvg, FedProx, Client-K Aggregation, and FedCodl, exhibit identical F1-scores, indicating their inability to adapt effectively to class imbalance. FedKT performs similarly on most datasets but experiences a drastic performance drop on the GMSC dataset (0.114 F1-score), further highlighting its instability. The

improved F1-score of FedLossMix suggests better handling of minority class representations, mitigating biases in credit scoring models.

## 6. Theoretical Contribution

The empirical findings across accuracy, convergence rate, and F1-score collectively advance the theoretical understanding of federated learning under conditions of heterogeneous, and imbalance dataset. The consistent superiority of FedLossMix over established baselines such as FedAvg, FedProx, and FedCodl supports the theoretical premise that uniform aggregation fails to accommodate non-IID client data distributions common in credit scoring. FedLossMix contributes to the theory of context-aware federated aggregation by introducing loss-based adaptive weighting, where local loss functions act as dynamic indicators of client reliability. This reconceptualization shifts FL aggregation from a participation-driven process to a performance-sensitive paradigm.

Furthermore, the faster convergence of FedLossMix empirically validates its federated optimization efficiency. Existing aggregation methods typically rely on fixed constraints or selective client participation, which can slow learning and increase communication overhead (Pei, Jiaming, et al.,2024). FedLossMix introduces a self-regulating equilibrium mechanism that prioritizes client updates contributing most to global objective improvement. This extends convergence theory (Haddadpour, Farzin, and Mehrdad Mahdavi.,2019) by showing that adaptive weighting guided by local loss differentials can accelerate convergence without compromising stability.

Finally, the enhanced F1-scores, especially on imbalanced datasets such as Dataset ID 350 and GMSC, underscore FedLossMix's contribution to fairness and bias mitigation theory in decentralized environments. Traditional aggregation amplifies majority class dominance (McMahan et al. 2017), whereas FedLossMix incorporates loss-sensitive adjustments that strengthen minority class representation. This outcome integrates principles from cost-sensitive learning into FL. This expands its theoretical scope beyond optimization toward equitable representation.

## 7. Practical Contribution

The findings of this study have meaningful implications for the adoption of federated learning (FL) systems within financial and enterprise information environments. The superior performance of FedLossMix across multiple datasets suggests that financial institutions can use FedLossMix to securely train models across distributed data silos without violating privacy or data residency constraints. FedLossMix enables trust-based collaboration across heterogeneous data environments by dynamically weighting client contributions and creates a practical blueprint for multi-party model development under strict compliance boundaries.

The faster convergence behaviour of FedLossMix translates into reduced communication costs, and shorter training cycles. This efficiency can directly enhance the responsiveness of enterprise AI

systems which allows faster model retraining to adapt quickly to changing customer and market conditions.

FedLossMix offers actionable pathways for aligning federated systems with emerging trustworthy AI principles which is important from a regulatory and governance perspective. Its ability to handle data imbalance and improve minority-class recognition supports fairer and more inclusive decision-making, to be compliant with frameworks such as the EU AI Act (Act, E.A.I., 2024) and NIST AI RMF (AI, N., 2023). Furthermore, the loss-aware weighting mechanism introduces a transparent rationale for client update prioritization. These attributes make FedLossMix suitable for regulated sectors where algorithmic accountability, ethical AI assurance, and operational transparency are as vital as predictive performance.

## 8. Conclusion

The research introduces the concept of FedLossMix, a federated learning framework designed to address the challenges of heterogeneous and imbalanced data. Federated Learning enables financial institutions to collaboratively train models while ensuring data privacy and regulatory compliance. However, traditional FL aggregation methods, such as FedAvg, FedProx, and Client-K Aggregation, fail to address the impact of heterogeneity in datasets and class imbalance on model performance. To overcome these challenges, FedLossMix dynamically assigns aggregation weights based on client validation loss, prioritizing contributions from well-generalized models while retaining proportional participation from all clients. Unlike existing approaches that rely on uniform constraints or hard client selection, this loss-based weighting mechanism enhances the stability and fairness of federated credit scoring models while reducing the computational overhead associated with knowledge distillation-based techniques like FedCodl and FedKT.

Experimental results across four credit scoring datasets demonstrate that FedLossMix consistently outperforms existing FL methods in terms of classification accuracy, F1-score, and convergence speed. The approach shows significant improvements in handling imbalanced credit data, mitigating biases in model aggregation while ensuring faster convergence and reduced communication overhead. Compared to baseline methods, FedLossMix offers more effective adaptation to heterogeneous financial datasets, leading to improved generalization and robustness in federated ML use cases such as credit scoring.

## 9. Future Directions

The use of FedLossMix algorithm demonstrates improvement in credit risk. However, they are areas which remain open for future research. One key direction is further expanding FedLossMix beyond credit scoring to other financial risk domains, such as fraud detection, anti-money laundering, and real-time transaction monitoring (Ali, A. et al. 2022). These domains require federated models that adapt to highly dynamic and adversarial environments which makes the ML model training particularly

challenging. So, developing an adaptive FL framework that incorporates real-time model updates and anomaly detection mechanisms would further enhance the practical applicability of FedLossMix.

Another promising area for future work is addressing privacy-preserving enhancements in FedLossMix. While FL inherently provides data privacy, additional techniques such as differential privacy (El Ouadrhiri, A. and Abdelhadi, A., 2022) and homomorphic encryption (Marcolla, C., 2022) can further strengthen data security and compliance with regulatory frameworks like GDPR and CCPA. By studying how these privacy-enhancing mechanisms interact with loss-based aggregation could provide valuable insights.

Furthermore, further analysis of FedLossMix's convergence properties under different non-IID data conditions (Lu, Z et al. 2024) would provide deeper insights into its learning. Additionally, evaluating FedLossMix on multi-institutional proprietary datasets in collaboration with financial organizations would help validate its real-world feasibility and performance in live deployment scenarios.

Finally, hybrid FL architectures that combine centralized and decentralized learning paradigms could be explored to further improve model personalization at the client level. Personalized federated learning techniques, such as meta-learning (Vettoruzzo et al. 2024) and multi-task learning (Kwon, S., Jang, J. and Kim, C.O., 2025), could be integrated with FedLossMix to ensure more institution-specific credit risk models while preserving the benefits of collaborative learning.

## References

Ala'raj, Maher, and Maysam F Abbod. 2016. "Classifiers consensus system approach for credit scoring." Knowledge-Based Systems 104: 89–105.

AI, N., 2023. Artificial intelligence risk management framework (AI RMF 1.0). URL: *https://nvlpubs. nist. gov/nistpubs/ai/nist. ai, pp.100-1.*

Ali, A., Abd Razak, S., Othman, S.H., Eisa, T.A.E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H. and Saif, A., 2022. Financial fraud detection based on machine learning: a systematic literature review. Applied Sciences, 12(19), p.9637.

Banabilah, S., Aloqaily, M., Alsayed, E., Malik, N. and Jararweh, Y., 2022. Federated learning review: Fundamentals, enabling technologies, and future applications. *Information processing & management*, *59*(6), p.103061.

Baesens, Bart, and Kristien Smedts. 2023. "Boosting credit risk models." The British Accounting Review 101241.

Bharati, Subrato, M Rubaiyat Hossain Mondal, Prajoy Podder, and VB Surya Prasath. 2022. "Federated learning: Applications, challenges and future directions." *International Journal of Hybrid Intelligent Systems* 18 (1-2): 19–35.

Chen, Ting-Hsuan, Chien-Chiang Lee, and Chung-Hua Shen. 2022. "Liquidity indicators, early warning signals in banks, and financial crises." *The North American Journal of Economics and Finance* 62: 101732.

Chen, H., Zhu, T., Zhang, T., Zhou, W. and Yu, P.S., 2023. Privacy and fairness in federated learning: On the perspective of tradeoff. *ACM Computing Surveys*, *56*(2), pp.1-37.

Act, E.A.I., 2024. The eu artificial intelligence act. European Union.

Dai, Y., Chen, Z., Li, J., Heinecke, S., Sun, L. and Xu, R., 2023, June. Tackling data heterogeneity in federated learning with class prototypes. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 37, No. 6, pp. 7314-7322).

El Ouadrhiri, A. and Abdelhadi, A., 2022. Differential privacy for deep and federated learning: A survey. *IEEE access*, *10*, pp.22359-22380.

Gao, W. and Zhou, Z.H., 2020. Towards convergence rate analysis of random forests for classification. *Advances in neural information processing systems*, *33*, pp.9300-9311.

Gou, Jianping, Baosheng Yu, Stephen J Maybank, and Dacheng Tao. 2021. "Knowledge distillation: A survey." International Journal of Computer Vision 129 (6): 1789–1819.

Guembe, Blessing, Ambrose Azeta, Victor Osamor, and Raphael Ekpo. 2023. "A Federated Machine Learning Approaches For Anti-Money Laundering Detection." *Available at SSRN 4669561* .

Hand, David J, and William E Henley. 1997. "Statistical classification methods in consumer credit scoring: a review." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 160 (3): 523–541.

He, Hongliang, Wenyu Zhang, and Shuai Zhang. 2018. "A novel ensemble method for credit scoring: Adaption of different imbalance ratios." *Expert Systems with Applications* 98: 105–117.

Hilal, Anwer Mustafa, Shaha Al-Otaibi, Hany Mahgoub, Fahd N Al-Wesabi, Ghadah Aldehim, Abdelwahed Motwakel, Mohammed Rizwanullah, and Ishfaq Yaseen. 2023. "Deep learning enabled class imbalance with sand piper optimization based intrusion detection for secure cyber physical systems." *Cluster Computing* 26 (3): 2085–2098.

Haddadpour, F. and Mahdavi, M., 2019. On the convergence of local descent methods in federated learning. *arXiv preprint arXiv:1910.14425*.

Ileberi, Emmanuel, Yanxia Sun, and Zenghui Wang. 2021. "Performance evaluation of machine learning methods for credit card fraud detection using SMOTE and AdaBoost." *IEEE Access* 9: 165286– 165294.

Kawa, Deep, Sunaina Punyani, Priya Nayak, Arpita Karkera, and Varshapriya Jyotinagar. 2019. "Credit risk assessment from combined bank records using federated learning." *International Research Journal of Engineering and Technology (IRJET)* 6 (4): 1355–1358.

Khandani, Amir E, Adlar J Kim, and Andrew W Lo. 2010. "Consumer credit-risk models via machinelearning algorithms." *Journal of Banking & Finance* 34 (11): 2767–2787.

Kweon, Yonghyeon, Bingrong Sun, and B Brian Park. 2021. "Preserving privacy with federated learning in route choice behavior modeling." Transportation research record 2675 (10): 268–276.

Kwon, S., Jang, J. and Kim, C.O., 2025. Credit scoring using multi-task Siamese neural network for improving prediction performance and stability. Expert Systems with Applications, 259, p.125327.

Lawlor-Forsyth, Erin, and M Michelle Gallant. 2018. "Financial institutions and money laundering: A threatening relationship?" *Journal of Banking Regulation* 19 (2): 131–148.

Li, Tian, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. 2020. "Federated optimization in heterogeneous networks." *Proceedings of Machine learning and systems* 2: 429–450.

Lipton, Z.C., Elkan, C. and Naryanaswamy, B., 2014, September. Optimal thresholding of classifiers to maximize F1 measure. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 225-239). Berlin, Heidelberg: Springer Berlin Heidelberg.

Lu, Z., Pan, H., Dai, Y., Si, X. and Zhang, Y., 2024. Federated learning with non-iid data: A survey. *IEEE Internet of Things Journal*, *11*(11), pp.19188-19209.

Marcolla, C., Sucasas, V., Manzano, M., Bassoli, R., Fitzek, F.H. and Aaraj, N., 2022. Survey on fully homomorphic encryption, theory, and applications. *Proceedings of the IEEE*, *110*(10), pp.1572-1609.

Mammen, Priyanka Mary. 2021. "Federated learning: Opportunities and challenges." *arXiv preprint arXiv:2101.05428* .

McMahan, Brendan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. 2017. "Communication-efficient learning of deep networks from decentralized data." In *Artificial intelligence and statistics*, 1273–1282. PMLR.

Moscato, Vincenzo, Antonio Picariello, and Giancarlo Sperlí. 2021. "A benchmark of machine learning approaches for credit score prediction." *Expert Systems with Applications* 165: 113986.

Myalil, Delton, MA Rajan, Manoj Apte, and Sachin Lodha. 2021. "Robust collaborative fraudulent transaction detection using federated learning." In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 373–378. IEEE.

Ni, Xuanming, Xinyuan Shen, and Huimin Zhao. 2022. "Federated optimization via knowledge codistillation." *Expert Systems with Applications* 191: 116310.

Pei, J., Liu, W., Li, J., Wang, L. and Liu, C., 2024. A review of federated learning methods in heterogeneous scenarios. *IEEE Transactions on Consumer Electronics*, *70*(3), pp.5983-5999.

Reurink, Arjan. 2018. "Financial fraud: A literature review." *Journal of Economic Surveys* 32 (5): 1292–1325.

Scott, A. O., Amajuoyi, P., & Adeusi, K. B. (2024). Effective credit risk mitigation strategies: Solutions for reducing exposure in financial institutions. *Magna Scientia Advanced Research and Reviews*, *11*(1), 198-211.

Song, Yu, Yuyan Wang, Xin Ye, Russell Zaretzki, and Chuanren Liu. 2023. "Loan default prediction using a credit rating-specific and multi-objective ensemble learning scheme." *Information Sciences* 629: 599–617.

Šušteršic, Maja, Dušan Mramor, and Jure Zupan. 2009. "Consumer credit scoring models with limited˘ data." *Expert Systems with Applications* 36 (3): 4736–4744.

Suvarna, R, and A Meena Kowshalya. 2020. "Credit card fraud detection using federated learning techniques." *International Journal of Scientific Research in Science, Engineering and Technology* 7 (3).

Tian, Ye, Bo Bian, Xiaofei Tang, and Jing Zhou. 2021. "A new non-kernel quadratic surface approach for imbalanced data classification in online credit scoring." *Information Sciences* 563: 150–165.

Tripathi, Diwakar, Alok Kumar Shukla, B Ramachandra Reddy, Ghanshyam S Bopche, and D Chandramohan. 2022. "Credit scoring models using ensemble learning and classification approaches: a comprehensive survey." *Wireless Personal Communications* 123 (1): 785–812.

Truong, Nguyen, Kai Sun, Siyao Wang, Florian Guitton, and YiKe Guo. 2021. "Privacy preservation in federated learning: An insightful survey from the GDPR perspective." *Computers & Security* 110: 102402.

Vettoruzzo, A., Bouguelia, M.R., Vanschoren, J., Rögnvaldsson, T. and Santosh, K.C., 2024. Advances and challenges in meta-learning: A technical review. *IEEE transactions on pattern analysis and machine intelligence*, *46*(7), pp.4763-4779.

Wang, Zhongyi, Jin Xiao, Lu Wang, and Jianrong Yao. 2024. "A novel federated learning approach with knowledge transfer for credit scoring." *Decision Support Systems* 177: 114084.

Xia, Yufei, Chuanzhe Liu, Bowen Da, and Fangming Xie. 2018. "A novel heterogeneous ensemble credit scoring model based on bstacking approach." *Expert Systems with Applications* 93: 182–199.

Yang, Miao, Hua Qian, Ximin Wang, Yong Zhou, and Hongbin Zhu. 2021a. "Client selection for federated learning with label noise." *IEEE Transactions on Vehicular Technology* 71 (2): 2193– 2197.

Yang, Wei, Yuan Yang, Xiaobing Dang, Hao Jiang, Yizhe Zhang, and Wei Xiang. 2021b. "A novel adaptive gradient compression approach for communication-efficient Federated Learning." In *2021 China Automation Congress (CAC)*, 674–678. IEEE.

Yang, Wensi, Yuhang Zhang, Kejiang Ye, Li Li, and Cheng-Zhong Xu. 2019. "Ffd: A federated learning based method for credit card fraud detection." In *Big Data–BigData 2019: 8th International Congress, Held as Part of the Services Conference Federation, SCF 2019, San Diego, CA, USA, June 25–30, 2019, Proceedings 8*, 18–32. Springer.

Zhang, Xiaoming, and Lean Yu. 2024. "Consumer credit risk assessment: A review from the state-ofthe-art classification algorithms, data traits, and learning methods." *Expert Systems with Applications* 237: 121484.

Zhao, Yue, Meng Li, Liangzhen Lai, Naveen Suda, Damon Civin, and Vikas Chandra. 2018. "Federated learning with non-iid data." *arXiv preprint arXiv:1806.00582* .

Zhu, Zhuangdi, Junyuan Hong, and Jiayu Zhou. 2021. "Data-free knowledge distillation for heterogeneous federated learning." In *International conference on machine learning*, 12878–12889. PMLR.