Journal of Hunan University (Natural Sciences)

Vol. 52 No. 11
November 2025

Available online at
https://jonuns.com

Open Access Article https://doi.org/10.55463/issn.1674-2974.52.11.1

Generative Models for Personalized Federated Learning via Class-Separable Latent Spaces and Meta-Adaptation

Sudhanshu Singh[1], Rokaya Akter[1], Tanzina Sultana[2], Ping Liu[3], Xinguo Lu[1]*

[1] Department of Computer Science and Electrical Engineering, Hunan University, Changsha, China,
[2] Department of Computer Science and Technology, Nanjing University of Posts and Telecommunications, Nanjing, China,
[3] Hunan Want Want Hospital, Changsha, China,

* Corresponding author: hnluxinguo@hnu.edu.cn

Article history
Received: October 21, 2025
Revised: November 23, 2025
Accepted: December 8, 2025
Published: December 30, 2025

Abstract: In decentralized learning scenarios such as federated learning (FL), client data are often highly heterogeneous, leading to non-IID distributions that degrade global model performance, slow convergence, and limit personalization. Conventional FL aggregation schemes inadequately capture client-specific structure and struggle to generalize across diverse data regimes. In this work, we propose a personalized federated learning framework that integrates Fisher Discriminant Analysis (FDA), meta-learning, and convolutional neural networks (CNNs) to jointly enhance class separability, adaptability, and representation quality. FDA is employed to project client-side features into an optimal class-separable latent space, enabling more discriminative modeling of local task characteristics under non-IID conditions.

On top of this representation, a meta-learning strategy is used to learn initialization parameters that support fast adaptation, allowing each client to efficiently fine-tune the model to its own data while preserving global knowledge. A CNN backbone is adopted to extract hierarchical feature representations, providing both fine-grained local patterns and robust global semantic embeddings suitable for heterogeneous federated environments. We conduct extensive experiments on CIFAR-10, CIFAR-100, and EMNIST-Writers under non-IID data partitions generated using a Dirichlet concentration parameter α = 0.5. The proposed method achieves 79.6% test accuracy on CIFAR-10, 45.7% on CIFAR-100, and 85.5% on EMNIST-Writers, consistently outperforming strong baseline FL methods in terms of adaptability, generalization, and personalization quality. The results demonstrate that combining class-separable latent spaces with meta-adaptation improves both global coherence and client-specific performance while preserving data privacy and offering good scalability. Future work will extend the framework to multi-modal and sequential data and investigate its deployment in cross-device and cross-silo FL settings.

Keywords: federated learning; personalized models; Fisher Discriminant Analysis; meta-learning; non-IID data; data privacy.

生成式人工智能引导的哨兵系统，用于自我优化的联邦网络安全与智能威胁检测

摘要：随着分散环境中客户数据的复杂性和异质性日益增加，传统的联合学习遇到了重大挑战，包括非 IID 客户数据、有限的泛化以及由于次优聚合和个性化策略不足而导致的收敛缓慢。 为了克服这些限制，我们提出了一种新的个性化联合学习框架，该框架将费舍尔判别分析（FDA）与元学习相结合。 美国食品药品监督管理局将特定于客户端的特征变成一个最大程度的类可分离的潜在空间，有效地捕捉异构的特定任务的细微差别。 我们采用元学习策略进行快速适应，实现在多样化和不断发展的客户数据分布中实现高效的微调和快速融合。 此外，我们的模型采用卷积神经网络（CNN）架构来提取分层表示，从而实现精细的局部特征提取和强大的全局语义嵌入。 在非 IID 设置下对基准数据集 CIFAR-10、CIFAR-100 和 EMNIST-Writers 进行全面评估。 该框架在 CIFAR-10（Dirichlet 0.5）上取得了 79.6%的成绩，在 CIFAR-100（Dirichlet 0.5）上取得了 45.7%的成绩，在 EMNIST-Writers 上取得了 85.5%的成绩，在适应性、泛化和个性化方面始终优于基线模型。 我们的结果强调了该框架在现实世界联合环境中增强全球一致性和客户特定准确性方面的隐私意识设计、可扩展性和有效性。 未来的工作将探索将该框架扩展到多模态和顺序数据，以及在跨设备和跨孤岛 FL 场景中的部署。

关键词：聯邦學習、個人化模型訓練、Fisher 判別分析、非獨立同分佈資料、資料隱私.

---

1. Introduction

Federated learning (FL) has emerged as a prominent paradigm for training machine learning models over decentralized data sources while preserving user privacy. Unlike conventional centralized learning, FL enables multiple clients to collaboratively train a shared model without transferring raw data to a central server. However, many classical FL algorithms implicitly assume that client data are independent and identically distributed (IID), an assumption that rarely holds in real-world applications such as financial services, healthcare, and educational systems, where transparency and security in handling user data are critical [1]. In practice, non-IID data distributions across clients constitute a fundamental challenge for FL [2]. In distributed systems [3], real-time data processing pipelines [4], cloud computing environments [5], and secure financial infrastructures [6], data are inherently partitioned among a large number of clients with distinct characteristics (e.g., behavioral patterns, usage intensity, or domain-specific constraints), which strongly influences model performance. Similar trends are observed in the educational domain, where machine learning is increasingly used to model student behavior and preferences, motivating privacy-preserving and adaptive learning environments [8]. These heterogeneous conditions reflect differences in data distributions, class proportions, and feature spaces, often exacerbated by device diversity, localized data collection, and user-specific activity patterns [9]. As a result, training a single globally optimal model using standard aggregation schemes becomes exceedingly difficult, particularly in the presence of label noise, class imbalance, and limited local sample sizes. Naïvely averaging model updates from heterogeneous and skewed clients typically leads to suboptimal convergence and poor generalization. These limitations underscore the need for advanced, adaptive, and personalized FL methods capable of handling non-IID data in decentralized settings.

Within decentralized learning, the ability to effectively adapt a global model to client-specific data distributions is now recognized as essential for achieving strong performance under non-IID conditions [10]. Personalized federated learning (PFL) has therefore gained attention as a family of approaches that aims to customize the global model to individual clients in order to improve local relevance and predictive accuracy [11]. Typical PFL methods employ techniques such as local fine-tuning, feature alignment, and regularization to accelerate convergence and enhance generalization across heterogeneous clients [12]. A variety of strategies have been proposed to address data heterogeneity: clustering-based methods group clients with similar distributions and train localized models [14], model-based approaches selectively share parameters or partition network layers to preserve client-specific components [16], and transfer-learning-based methods leverage prior knowledge to rapidly adapt to new client tasks [17]. Despite these advances, important challenges remain, including the preservation of class separability across clients, mitigation of feature-space misalignment, and robust personalization under severe data imbalance or limited local data. A natural way to alleviate these issues is to project features into a class-separable latent space, for example using Fisher Discriminant Analysis (FDA), and to couple this with meta-learning techniques that improve adaptation and support more effective client-specific personalization.

Building on the decomposition of model training into shared representation learning and personalized classifier adaptation in pFedFDA [18], we aim to further optimize this paradigm for highly heterogeneous decentralized environments. While pFedFDA separates global and local components in a conceptually appealing manner, its representational capacity and the flexibility of local classifier adaptation can still be limited, especially under strong class imbalance and severe feature mismatch. To address these shortcomings, our approach enhances the discriminative power of the shared feature extractor and reinforces the alignment between global and client-specific classifiers. By doing so, the proposed framework more effectively manages the bias–variance trade-off, promotes stable convergence, and improves both generalization and personalization, particularly in low-data and high-heterogeneity regimes.

Despite the considerable progress in PFL, several open challenges persist: misaligned feature spaces across clients, overfitting to client-specific idiosyncrasies, slow adaptation to new clients, and difficulties in balancing global bias against local variance. These limitations frequently hinder the ability of FL systems to generalize reliably in decentralized non-IID environments. To tackle these issues, we introduce a new framework that combines efficient discriminative feature learning with meta-adaptive personalization. Our method employs FDA to increase the discriminative power of shared feature statistics, thereby improving class separation and facilitating more effective local adaptation. In addition, we incorporate meta-learning to optimize the initial model parameters, enabling fast adaptation to new clients with minimal fine-tuning and limited local data. We evaluate the proposed framework on three widely used benchmark datasets CIFAR-10, CIFAR-100, and EMNIST which exhibit varying degrees of complexity and distributional heterogeneity. These datasets provide a rigorous testbed for assessing performance in realistic, privacy-preserving federated scenarios.

The main contributions of this work are summarized as follows:

1. We propose a discriminative aggregation strategy for FL that leverages Fisher Discriminant Analysis to project client-specific representations into a maximally class-separable latent space. This enhances the ability of the global model to preserve heterogeneous, task-specific structures across distributed nodes.
2. We employ a meta-learning scheme to learn a robust global initialization, enabling fast few-shot adaptation and efficient fine-tuning on limited local data. Operating primarily in the latent feature space, the framework maintains data privacy and scalability while supporting task-specific personalization and global coherence across diverse client distributions.
3. We adopt a convolutional neural network (CNN) backbone to learn hierarchical feature representations that capture both fine-grained local details and high-level semantic patterns. These expressive and transferable embeddings are crucial for downstream personalization under non-IID conditions.

4. We conduct extensive experiments on CIFAR-10, CIFAR-100, and EMNIST, demonstrating that the proposed method achieves strong generalization performance and robustness to statistical heterogeneity, client imbalance, and non-IID data distributions, consistently outperforming competitive PFL baselines.

2. Literature Review

Federated Learning (FL) supports decentralized model training over clients with varying data privacy. FedAvg[19] This scheme achieves global model optimization without sharing raw data. But the non-IID characteristics of data on clients cause great obstacle, resulting in the statistical heterogeneity which will harm model generalization and convergence. To overcome these difficulties, some refined algorithms were developed: personal strategies or more complicated aggregation methods [20][21][22].

For instance, [23] improved FedAvg by integrating it with the SE-ResNet18-E model to enhance classification performance for non-IID data while dealing with accuracy and privacy issues. Also, FedEL [10] relied on data diversity to train a variety of weak learners and obtained better performance in non-IID settings with higher accuracy. In another line of work, [24] introduced a novel algorithmic framework for Federated Learning (FL) formulated under primal-dual optimization to handle the instability of FedAvg in non-IID setting. Their method enhances optimization, communication efficiency, and robustness under heterogeneous local data distribution. Also, FLAD [25] enhances the FedAvg algorithm by integrating the update process of users' models with new attack profiles. This method improves the convergence time and the accuracy without sharing of test data, especially in unbalanced and heterogeneous datasets.

Some adaptive FL framework was also introduced in intrusion detection, where an adaptive FL framework for intrusion detection with integrating privacy-preserving training and class imbalance mitigated for security enhancement of distributed network Also FedDynamic [26] deals with non-IID data challenge via dynamically portioning aggregation weights on the basis of critical indices such as local model accuracy, data quality and model differences, which results in enhanced convergence and accuracy compared to existing methods including FedAvg, FedProx and Scaffold. Despite its success, federated learning struggles with the client heterogeneity and needs customized solutions for improved model accuracy and user data privacy in AI empowered systems [27].

Personalized Federated Learning (PFL) has improved traditional federated learning by training models that fit more individual client's data distributions. When dealing with non-IID data, federated learning has difficulty to capture the data distribution of which PFL deals owing to its local model adaption, leading a better generalization and convergence. Through the fusion of global collaboration and local personalization, PFL guarantees accuracy and efficiency over varied client data in distributed setting [28]. [29] suggested that the base layer should be used for local and global training in personalized federated learning without involving noise introduced by the head layer. They also propose a way to automatically select the fine-tuning iteration based on variance of accuracy, which can help the model generalize better and save communication cost. One way to handle data heterogeneity in personalized federated learning is with model-based personalization methods, which customize global models according to local client scenarios. The representative models that are inspired by this approach include pFedAvg [30], pFedFDA [18], FedPAC [31], FedALA [29] and FedNova [32] which all adopt distinct methods for locally adapting the model according to local data and globally aggregating the adapted models for improving non-IID performance. As an example, pFedFDA effectively generates personalized models by adapting global generative classifiers to local feature distributions. Analogously, FedALA encodes required information from the global model to enrich client models for personalized FL. Another work FedPAC directly align local and global features using global semantic knowledge to learn more discriminative representation-weighting, improving personalization and accuracy of client models in federated learning.

While federated learning continues to improve, challenges such as poor feature separability and slow adaptation in decentralized scenarios are still unresolved. Combining techniques such as Fisher Discriminant Analysis (FDA) and meta-learning is a potentially fruitful solution to these problems providing improved feature representation and increased convergence speed in heterogeneous data settings. This is desirable when FDA maps data into space where inter-class separation is increased and intra- class variance minimizations, enhancing the discriminative nature of the model. Aggregation methods that have been shown to work well in practice such as FedAvg [33]} and FedProx [23] perform reasonably, but do not feature heuristics that address preserving class separability for non-IID data. Recent literature has investigated the role of FDA in bridging these weaknesses. For instance, [34] illustrated the benefits of FDA to increase classification accuracy in complex systems, such as flow pattern identification for multiphase flows. At the same time, meta-learning has been proposed as an effective approach to achieve fast adaptation of models on new client data with few observations. By learning-to-learn, models are more adaptive and robust and thus can scale better, be more able to generalize and generalize with less forgetting under dynamic federated conditions [35].

In this paper, we propose a novel framework to cope with the two main remaining problems (poor discriminability of features and slow learning for personalized federated learning) in an unified way, which combines Fisher Discriminant Analysis (FDA) and meta-learning. Our approach exploits FDA for improving feature representation through data projection into a space that optimizes class separability, which renders it more applicable to heterogeneous data. In the federated setting FDA is conducted in a privacy preserving way, where models trained at client level are aggregated and using a point cloud of the data distribution intrinsic global covariance structure are computed across distributed nodes, to ensure stable and representative projections. To exploit the enriched feature space, meta-learning is utilized to fast adapt to new clients or tasks with few samples. This speeds up convergence and also improves generalization among different and non-iid client distributions. Our approach provides a more robust and efficient solution for decentralized learning with non-IID data by combining federated FDA with meta-learning, which simultaneously optimizes global knowledge sharing and local personalization.

3. Proposed Method

This section summarizes the proposed model and describes how key techniques are combined to work collaboratively. We introduce a Personalized Federated Learning (PFL) framework in which a server orchestrates training through clients while maintaining privacy by keeping data on individual devices. The use of CNN based hierarchical feature extractor/suppressor makes the model capable of dealing with complex data which automatically learn discriminant features and thereby improve upon heterogeneous client side data distributions. We then propose FDA as a key component of our model and further optimize the overall model by improving feature extraction, class scatter, and modelling local data distributions.

FDA enhances the model's capability to differentiate various classes, especially in non- IID settings by encouraging every client data to be more reflected of in global model. Meta-learning is employed in our model to support fast adaptation on new tasks and clients, thus facilitating model initialisation and generalisation across heterogeneous data distributions. Meta-learning speeds up the adaptation by making use of previous tasks to update the model more efficiently (\"update fast\") for new clients.

Personalization is addressed via local model adaptation and customer specific parameter optimization, which balances the global and the local knowledge. First, effective aggregation methods such as weighted averaging and noise mitigation are introduced to enhance parameter updates, thereby enhancing the efficient training and convergence rates. The overall architecture for the model is depicted in figure 1.

By adopting such a holistic view, our framework's objective is to enhance model personalization, privacy preservation and rapid adaptation capability, and optimize training process at the end of day deliver a strong and scalable solution for Personalized Federated Learning.

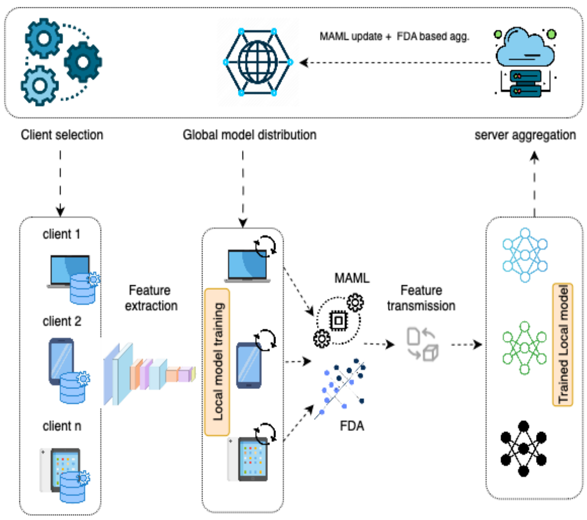

Figure 1. The overview of our proposed model, a detailed process for local training, Fisher Discriminant Analysis (FDA) and Meta-learning (MAML), global model aggregate approach

3.1. System Model

The model of the system improves PFL framework by combining Meta-learning and FDA to improve the performance of personalized learning, be more robust for non-IID data and guarantee privacy. This model is built on a CNN structure and is modified to exploit FDA for dimensionality reduction or classification purposes. The CNN feature is used to extract the features and FDA is performed on these extracted features for boosting discriminant power.

Incorporating meta- learning which helps to adapt the CNN model for new tasks, given a few examples and generalize well across different data distributions.

The key symbols employed in this system model are listed in Table 1. This strategy works in client-server architecture: the server collects the updates from various clients; our challenge is to have rendezvous between these clients with their own local content without downloading raw content. The model is bootstrapped with a global model, then tailored by meta-learning adaptation at each client, which enables rapid adaptation to various data distributions.

Table 1 Symbols and Their Descriptions

Symbol	Description
ϕ_0^i.	Model parameters for client i at initial stage
ϕ_r^i	Model parameters for client i at the r^{th} meta-learning round
μ_r^i	Local class mean for client i
Σ_r^i	Local covariance matrix for client i
x_j	Data samples at client i
N_i	Number of samples at client i
W_i	FDA projection matrix for client i
ϕ_r^g	Global CNN model parameters
$\Delta\phi_r^g$	Change in global CNN model parameters
η_{meta}	Meta-learning rate for global model update
S_W	Within-class scatter matrix for FDA
S_B	Between-class scatter matrix for FDA
∇L_{meta}^i	Meta-gradient for client i during meta-learning
α	Interpolation parameter balancing local and global knowledge
μ_g	Global class mean
$\sum_g$	Global covariance matrix
E_g	Number of global epochs
E_l	Number of local epochs
M_t	Number of clients
D_i	Client's private dataset
η	Learning rate for local training
y	Class labels for training
LocalLoss_i	Local loss for client i
LocalAcc_i	Local accuracy for client i
GlobalLoss_r	Global loss at the r^{th} global epoch
$GlobalAcc_r$	Global accuracy at the r^{th} global epoch

Then picture a process by which some of those prescriptions are recognized at hospitals worldwide, handwritten and filled in visually; records are processed visually at financial institutions; and visual patterns are analyzed on blackboards and PowerPoint presentations at academic centers. Each of these organizations acts as a client Ci and possesses its won private dataset from character images in EMNIST to natural image datasets including CIFAR-10 and CIFAR-100. These inherently diverse and non-IID datasets are derived from either different individual styles (in the case of EMNIST) or Dirichlet distributed splits using α = 0.1, and α = 0.5 which model distribution skews present in realistic data distributions. To solve the common task of training a joint but personalized recognition model, these clients join federated learning. At the local level, individual clients start by feeding raw images to a shared Convolutional Neural Network (CNN) model in parallel that extracts representative and hierarchical features from 28×28 grayscale (or RGB) input. To support adaptation, the system builds on meta-learning and creates local models that can rapidly refine themselves to be fine-tuned using a small fraction of the support query data split mimicking new tasks during training. Across a few local epochs, all the models are trained body model specific by learning through inner-loop optimization. Meanwhile, at this level of locality, the system uses Fisher Discriminant Analysis (FDA) in feature space to maximize inter-class distance and minimize intra-class variation. This ensures that though each client only concentrates on their own patterns, the generated local features are still highly discriminative. When the training is completed, clients do not submit raw data. Instead, they only hand through the updated model parameters and learned feature representations to the central server in a secure way. An intelligent aggregation process occurs there as based on FDA. The server calculates the between-class scatter matrix Sb and within-class scatter matrix Sw over all client models, learning a discriminative subspace that not only captures the information in individual client features but also maintains global structure of the model. Together with this, meta-optimization updates on each round polish the global model's initialization to boost learning in following rounds. The global model M which contains the generalized wisdom of all clients and it is personalized, sent back to each institution. Clients inherit this enriched model and are able to execute it swiftly against their own data, preserving accuracy and relevancy while impeding local information loss.

This iterative process of local training, FDA-improving representation, secure aggregation and global distribution repeats. The global model is enhanced to be more robust and context aware, whereas personalization offers each client its best performance on its own data domain at the same time. The method guarantees data privacy, and speeds up convergence as well as maintains scalability, achieving good performance on various datasets such as CIFAR-10, CIFAR-100 and EMNIST. In the end, what you get is not just a model of the world; it's a symphony of personal intelligence, played across institutions without ever breaking down the walls between private and public data.

In our model, we are aware that the clients participating in the federated learning have their data generalized to different distributions and computation capabilities. This means that every individual client may have imbalanced data and not contribute equally with computational resources, which may affect the convergence and performance of the global model. To deal with the problem, we employ Fisher Discriminant Analysis (FDA) to improve feature extraction ability by maximizing separability between classes. FDA calculates the local class averages and their covariance matrices against which it obtains a projection matrix that maximizes the discriminative power of the features. We also introduce meta-learning to enable the model learning new tasks more efficiently. Our method involves a CNN architecture that cooperates with FDA and can adapt to complex data, in addition to automatically learning discriminating features. This procedure is outlined in Algorithm 1, and each client starts the meta-learning adaptation process with a personal local model that will gradually adapts to the local data. By unifying these techniques, our model improves personalization and flexibility while preserving privacy across a wide range of client scenarios. Below is an overview of the details detailing how our modelization works,

For client training, each client i starts by initializing the local model ϕ_0^i from the global model ϕ_0^g . The CNN architecture, is used for efficient feature extraction and hierarchical representation learning. The client performs meta-learning adaptation for R_{meta} rounds, adjusting the model parameters ϕ_r^i to adapt to the local data distribution, $\phi_r^i = \text{Adapt}(\phi_r^g)$. The local training is driven by the loss function L_{local} in equation 1, and the model parameters are updated as equation 2.

$$L_{\text{local}} = \sum \text{Loss}(\phi_r^i(x), y) \quad \forall (x, y) \in D_i \tag{1}$$

$$\phi_r^i = \phi_r^i - \eta \cdot \nabla L_{\text{local}} \tag{2}$$

where η is the learning rate. The goal is to refine the model such that it improves its performance locally while maintaining global consistency. Local metrics like loss and accuracy are monitored in equation 3 and 4.

$$\text{LocalLoss}_i = L_{\text{local}} \tag{3}$$

$$\text{LocalAcc}_i = \text{Accuracy}(\phi_r^i(x), y) \tag{4}$$

Next, the client computes local statistics for FDA. The class mean μ_r^i and covariance matrix Σ_r^i are estimated from the local dataset denoted in equation 5 and 6. where x_j represents the data samples and N_i is the number of samples at client i. FDA is then applied to improve the discriminative power of the learned features.

$$\mu_r^i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_j \tag{5}$$

$$\Sigma_r^i = \frac{1}{N_i} \sum_{j=1}^{N_i} (x_j - \mu_r^i)(x_j - \mu_r^i)^\top \tag{6}$$

FDA is applied to enhance discriminative features via projection matrix W_i in equation 7.

$$W_i = \Sigma_r^{i-1} \cdot (\mu_r^i - \mu_g) \tag{7}$$

After applying FDA, the client computes the meta-gradient ∇L_{meta}^i , which reflects the optimal direction for updating the model parameters denoted in equation 8.

$$\nabla L_{\text{meta}}^i = \nabla_\phi L(\phi_r^i; \text{Query}) \tag{8}$$

For server aggregation, the model updates from the active clients are aggregated. The meta-update is computed by averaging the meta-gradients from all clients in equation 8.

$$\Delta\phi_r^g = \frac{1}{M_t} \sum_{i=1}^{M_t} \nabla L_{\text{meta}}^i \tag{9}$$

Where M_t represents the number of active clients participating in the current meta-learning round. The global model is updated as, $\phi_r^g = \phi_r^g - \eta_{\text{meta}} \cdot \Delta\phi_r^g$. where η_{meta} is the meta-learning rate. The server aggregates the local statistics from all clients to form a global model. The global class mean μ_r^g is calculated in equation 10 as the average of the local class means μ_r^i from each client.

$$\mu_r^g = M_t \frac{1}{M_t} \sum_{i=1}^{M_t} \mu_r^i \tag{10}$$

Similarly, the global covariance matrix Σ_r^g is the average of the local covariance matrices Σ_r^i denoted in equation 11, and the global FDA projection matrix W_g is the average of the local projection matrices W_i denoted in equation 12. Global metrics such as loss and accuracy are monitored in equation 13 and 14.

$$\Sigma_r^g = \frac{1}{M_t} \sum_{i=1}^{M_t} \tag{11}$$

$$W_g = \frac{1}{M_t} \sum_{i=1}^{M_t} W_i \tag{12}$$

$$\text{GlobalLoss}_r = \frac{1}{M_t}\sum_{i=1}^{M_t} LocalLoss_i \quad (13)$$

$$GlobalAcc_r = \frac{1}{M_t}\sum_{i=1}^{M_t} LocalAcc_i \quad (14)$$

This aggregation step is to enable the global model to become aware of collective knowledge from all clients while preserving their local information as well. By consolidating these parameters, the server constructs a global model that captures the variance and statistics of all participating clients.

Algorithm 1 Personalized Federated Learning with Fisher Discriminant Analysis and Meta-Learning

procedure FDATRAINING
  // Initialization
  E_g, E_l, M_t, D_i ϕ_g^0, η, η_{meta}, D, y
  // Client Training
  for i = 1 to M_t do // For each client
    $\phi_0^i \leftarrow \phi_0^g$ // Initialize local CNN model with global model
    for r = 1 to R_{meta} do // Meta-learning adaptation
      $\phi_r^i \leftarrow \text{Adapt}(\phi_r^g)$ // Adapt the CNN model to local data
    end for
    for e = 1 to E_l do // Local training
      $L_{\text{local}} = \sum \text{Loss}(\phi_r^i(x), y) \quad \forall (x,y) \in D_i$
      $\phi_r^i \leftarrow \phi_r^i - \eta \cdot \nabla L_{\text{local}}$
      // Monitor local training metrics
      $\text{LocalLoss}_i \leftarrow L_{\text{local}}$
      $\text{LocalAcc}_i \leftarrow \text{Accuracy}(\phi_r^i(x), y)$
    end for
    // Compute local statistics (Fisher Discriminant Analysis)
    $\mu_r^i \leftarrow \text{Mean}(x_j \text{ for } x_j \in D_i)$ // Local class mean
    $\Sigma_r^i \leftarrow \text{Covariance}(x_j \text{ for } x_j \in D_i)$ // Local covariance matrix
    // Compute FDA projection
    $W_i \leftarrow \Sigma_r^{i-1} \cdot (\mu_r^i - \mu_g)$ // FDA projection matrix
    $\nabla L_{\text{meta}}^i \leftarrow \nabla_\phi L(\phi_r^i; \text{Query})$ // Meta-gradient
    Send ϕ_r^i, μ_r^i, Σ_r^i, W_i, LocalLoss_i, LocalAcc_i to server
  end for
  // Server Aggregation
  for r = 1 to E_g do // For each global epoch
    $\Delta\phi_r^g \leftarrow \frac{1}{M_t}\sum_{i=1}^{M_t} \nabla L_{\text{meta}}^i$
    $\phi_r^g \leftarrow \phi_r^g - \eta_{\text{meta}} \cdot \Delta\phi_r^g$
    $\Delta\phi_r^g \leftarrow \frac{1}{M_t}\sum_{i=1}^{M_t} \phi_r^i$
    $\mu_r^g \leftarrow M_t \frac{1}{M_t}\sum_{i=1}^{M_t} \mu_r^i$
    $\Sigma_r^g \leftarrow \frac{1}{M_t}\sum_{i=1}^{M_t} \Sigma_r^i$
    $W_g \leftarrow \frac{1}{M_t}\sum_{i=1}^{M_t} W_i$
    // Monitor global training metrics
    $\text{GlobalLoss}_r = \frac{1}{M_t}\sum_{i=1}^{M_t} LocalLoss_i$
    $GlobalAcc_r = \frac{1}{M_t}\sum_{i=1}^{M_t} LocalAcc_i$
    Send ϕ_r^g, μ_r^g, Σ_r^i W_g to clients
  end for
return
end procedure

3.2. Generative Model of Feature Distributions: Motivation for Generative Classifiers

Generative classifiers aim at learning the distribution of data from each class and are able to make predictions using the full structure of data. While discriminative classifiers learn decision boundaries, generative classifiers model the complete data distribution, specifying the probability distribution of the features for each class. In our model setting, we adopt a generative approach by modeling distributions of features per class which maximizes the separability, in this way the model can distinguish between classes better and will have significantly more learning capacity when dealing with non-IID data distributions among clients.

3.2.1. Fisher Discriminant Analysis (FDA) in Generative Models

FDA is also critical in improving the class-conditional feature distributions separability (especially in generative classifiers and federated learning). In class-based models, it was generally assumed that given a class, the feature vectors in this class have Gaussian distribution with a mean vector μ_i and covariance matrix specific to the class Σ_i

FDA attempts to map the high dimensional data domain into a low-dimensional subspace in such a way that the class separation are maximized by enlarging between-class scatter and, at the same time, reducing within-class scatter. To make it more precise, FDA first calculates the within-class scatter matrix S_W, that represents how samples spread out in a class. It is computed in equation 15 and C represents the total number of classes, while D_i is the set of data samples belonging to class i, and μ_i is the mean vector of class i.

$$S_W = \sum i = 1^C \sum_{x_j \in D_i} (x_j - \mu_i)(x_j - \mu_i)^\top \quad (15)$$

Next, FDA computes the between-class scatter matrix S_B, which measures the separation of each class mean from the global mean μ_g. This is defined in equation 16, where N_i is the number of samples in class i, and $\mu_g = \frac{1}{\sum_{i=1}^{C} N_i}\sum_{i=1}^{C} N_i\mu_i$ represents the global mean of all class data.

$$S_B = \sum i = 1^C N_i(\mu_i - \mu_g)(\mu_i - \mu_g)^\top \quad (16)$$

The primary aim for FDA is to search for a projection matrix W such that the ratio of determinant between-class scatter and within-class scatter in the transformed space becomes largest. This optimization problem can be described by the Fisher criterion in Eq.(17).

$$W = \text{argmax} \frac{|W^\top S_B W|}{|W^\top S_W W|} \quad (17)$$

These optimal projection directions that maximally separate the class distributions are obtained as solution of the generalized eigenvalue problem. In the particular case when there are two classes the solution becomes much simpler, for in this situation there is at most one discriminant direction w given by Equation (18), where Σ and it is a weighted sum of class covariance matrices.

$$w = \Sigma^{-1}(\mu_1 - \mu_2) \quad (18)$$

This formulation maximizes the class separation along the vector connecting the class means while accounting for intra-class variability. As a result, FDA improves the discriminability of class distributions in the feature space, which is essential for effective classification performance, particularly in generative models trained under decentralized or heterogeneous data conditions.

3.2.2. Meta-learning in Generative Models

Meta-learning enhances generative classifiers by enabling rapid adaptation to new tasks or data distributions, particularly in settings like federated learning where client data is non-i.i.d. and heterogeneous. The core objective of meta-learning is to optimize the initialization of model parameters such that they can be efficiently fine-tuned with minimal data from new tasks. In the context of the proposed framework, meta-learning is integrated to personalize the generative classifier by adjusting the model parameters ϕ based on the unique local distributions of each client.

Each client i uses its local support set to adapt the shared model parameters ϕ, obtaining a locally adapted parameter ϕ_r^i. The adaptation is typically performed using gradient descent on the support set loss L_{support}^i denoted in equation 19, where η_{inner} is the inner-loop learning rate.

$$\phi_r^i = \phi - \eta_{\text{inner}} \nabla_\phi L_{\text{support}}^i(\phi) \quad (19)$$

After local adaptation, the performance of ϕ_r^i is evaluated on the query set to compute the meta-loss L_{meta}^i. The meta-gradient, derived from this meta-loss, is then used to update the global model parameters via equation 20.

$$\phi \leftarrow \phi - \eta_{\text{meta}} \nabla_\phi \sum_{i=1}^K L_{\text{meta}}^i(\phi_r^i) \quad (20)$$

where K is the number of participating clients in the current round, and η_{meta} is the meta-learning rate. This process allows the global model to learn an initialization ϕ that can quickly personalize to each client's local distribution with a few gradient steps.

By integrating this mechanism with Fisher Discriminant Analysis (FDA), the model not only benefits from a discriminative feature space but also gains the flexibility to adapt to diverse client data. While FDA enhances class separability, meta-learning ensures fast and personalized adaptation across non-i.i.d. client distributions, making our model highly effective in federated environments.

Further, by combining this framework with Fisher Discriminant Analysis (FDA), the model not only takes advantage of a discriminative feature space but also has flexibility to be adjusted to different client data. FDA improves class separability, and meta-learning achieves fast and personalized adaptation on non-i. i.d. client distributions, thus rendering our model extremely effective in federated settings.

3.3. Global Representation Learning

In our framework, global representation learning is centered around training a shared Convolutional Neural Network (CNN) feature extractor $\phi(\cdot;\theta)$, where θ denotes the CNN parameters. The CNN is responsible for mapping raw input data $x \in \mathbb{R}^{H\times W\times C}$ (e.g., images) into a lower-dimensional latent feature space $\phi(x) \in \mathbb{R}^d$, optimized to support classification through class-conditional generative modeling. The CNN architecture consists of a series of convolutional layers with non-linear activations (e.g., ReLU), followed by pooling and fully connected layers. Formally, a typical CNN feature mapping is defined recursively in equation 21, where $W^{(l)}$ and $b^{(l)}$ are the convolution weights and biases at layer l, σ is the activation function, and $*$ denotes the convolution operation. The final output $\phi(x) = \phi^{(L)}(x)$ is used for classification.

$$\phi^{(l)}(x) = \sigma(W^{(l)} * \phi^{(l-1)}(x) + b^{(l)}), \quad \phi^{(0)}(x) = x \quad (21)$$

To guide the CNN towards learning discriminative and generalizable features, we integrate two learning strategies: Fisher Discriminant Analysis (FDA) and meta-learning. The CNN is trained locally on each client by minimizing a generative cross-entropy loss using a fixed classifier derived from class-conditional Gaussian distributions denoted in equation 22 and 23.

$$\mathcal{L}(x, y; \phi, \mu, \Sigma, \pi) = -\sum_{c=1}^{C} y_c \log p(y_c | \phi(x), \mu_c, \Sigma, \pi) \quad (22)$$

with

$$p(y_c|\phi(x)) = \frac{\pi_c \cdot \mathcal{N}(\phi(x); \mu_c, \Sigma)}{\sum_{k=1}^{C} \pi_k \cdot \mathcal{N}(\phi(x); \mu_k, \Sigma)} \quad (23)$$

where $\mathcal{N}(\cdot; \mu_c, \Sigma)$ denotes the multivariate Gaussian density for class c. Instead of inverting Σ, we approximate $\Sigma^{-1}\mu_c$ via a least-squares solution to stabilize training. To align the CNN features with a discriminative subspace, we apply FDA to compute a local projection matrix W_i for each client computed by equation 24, where μ_r^i is the local class mean, μ_g is the global class mean, and Σ_i is the local covariance. These projections are aggregated to form a global FDA projection (equation 25).

$$W_i = \Sigma_i^{-1}(\mu_r^i - \mu_g) \quad (24)$$

$$W_g = \frac{1}{M_t} \sum_{i=1}^{M_t} W_i \quad (25)$$

which captures the overall class discriminability across clients. The CNN feature extractor is then trained so that $W_g \phi(x)$ lies in a space where class separation is maximized. Simultaneously, we incorporate meta-learning to improve the adaptability of the CNN to client-specific distributions. Each client uses a local adaptation of model parameters ϕ_r^i via a gradient-based meta-update showed in equation 26, where $\mathcal{L}_{\text{meta}}^i$ is the loss computed on a query set, and η_{meta} is the meta-learning rate.

$$\phi_r^i \leftarrow \phi_r^i - \eta_{\text{meta}} \nabla_\phi \mathcal{L}_{\text{meta}}^i \quad (26)$$

This update ensures that ϕ rapidly adapts to new local data with few examples, which is critical in federated settings with non-IID client distributions. The CNN learns feature representations $\phi(x)$ that are both discriminative, via FDA, and adaptive, via meta-learning.These dual goals allow the global model to generalize well across all clients since it is constantly running, rapidly customizing in case of need.

4. Results

Analyzing results In this section, we conduct a thorough analysis of our experimental results, which come from serial high-quality simulated experiments. It is the aim of this paper to provide a comprehensive analysis of the proposed framework, its strengths and limitations and future directions. The importance of our findings lies in the fact that they promptly brought the fundamental difficulties of federated learning based on data heterogeneity, feature discrimination and slow model adaptation under control. In these experiments we intend to show that our model is robust and scalable in multicast environments.

4.1. Experiment Setup

We used the PyTorch (v1. 0. 0) to realise the model design, meanwhile we applied CNN structure while setting weights manually from [8] for efficient feature abstraction. The experiment was implemented in a Conda ENV using Python 3.11 programming language and PyCharm Professional IDE for development. It incorporates Fisher Discriminant Analysis (FDA) and meta-learning to improve the feature adaptation and personalization. The training hyperparameters including batch size, learning rate, weight decay, client participation proportion and number of rounds are collected in the table. We use the meta-learning parameters (meta-learning rate 0.01, meta-batch size 2, and optimize for 2 inner steps) to adapt our model to quickly learn new tasks. Performance metrics such as training loss, test accuracy, round time and total training time are recorded for each round. On the server side, one-dimensional weighted averaging is used to aggregate models, thus facilitating computationally-efficient updates of global models. The framework exploits PyTorch and few basic dependencies such as torch, numpy, random, time which help to perform the training and evaluation tasks. For reproducibility and efficient training, we employ virtual environments to handle the dependencies and guarantee identical batch sizes and evaluation protocol. In order to ensure the model generalization and privacy, we include FDA for better feature representation in federated learning, and propose a meta learning for fast adaptation across clients under different data distributions. This method that guarantees the system's ability to handle non-IID data, and still offers strong privacy with an efficient model customization. The training happens with fixed random seeds to maintain the consistency across different experiments.

4.2. Dataset Overview

Two determinant datasets, the EMNIST, CIFAR-10 and the CIFAR-100 are employed to compute the efficiency of our federated learning framework. Similarly, EMNIST is an extension of MNIST which comprises handwritten characters over 814 classes and has been designed to be more complex than the

MNIST and wide range of characters. CIFAR-10 has 60,000 32x32 color images in 10 classes and has been designed to be more complex than the MNIST and wide range of characters. CIFAR-10 has 60,000 32x32 color images in 10 classes, and CIFAR-100 also offers the same number of samples but with 100 classes which makes it challenging for image classification models. In order to model realistic non-IID (non-Independent and Identically Distributed) data distributions we partition the datasets using the Dirichlet distribution, as is common in federated learning which places data unevenly across clients. Two different values of the Dirichlet distribution parameter α are also employed: α = 0.5 and α = 0.1. With α = 0.5, the data distribution becomes moderately imbalanced thereby creating a balanced non-IID setting. But in the case of α = 0.1, the data distribution is strongly imbalanced, simulating very non-iid situations where some clients have a lot more images from certain classes and others very few. These data partitioning schemes let us quantify how well the federated learning model generalizes for different kinds of non-IID data patterns and we get some understanding about the robustness and flexibility of our solution.

Table 2 Experiment Parameters

Parameters	Setting
Dataset	Extended MNIST/Cifar10/Cifar100
model	CNN
Number of Classes	62/10/100
Number of Clients	25, 50, 75, 100 (non-iid distribution)
Partition Path	Dirichlet distribution with α=0.5 and α=0.1
Training Proportion	1.0
Sampling Probability	0.2 (20% of clients participate per round)
Local Epochs	1
Global Rounds	200
Eval gap	200
Batch Size	32
Learning Rate	0.01
Weight Decay	5e-4
Meta Learning Rate	0.01
Meta Batch Size	2
Inner Steps	2
Support Set Size	10 samples per class
Query Set Size	5 samples per class

4.3. Results and Performance Analysis

In this section we make a detailed exploration of our proposal and namely the personalized federated learning framework through extensive experiments. We compared our method with 12 state-of-the-art baselines over several datasets - CIFAR-10, CIFAR-100, and EMNIST across different data distribution characteristics of the dataset, training data proportion and client characteristic.

The assessment centres around the adaptability, generalization, personalization and robustness to data heterogeneity of the framework as well as its scalability.

4.3.1. Comparison with Baselines under Data Scarcity

In order to evaluate its performance, we compared it directly with 12 baseline methods on the CIFAR10-S Dir(0.5) and the CIFAR100-S Dir(0.5) datasets in different proportions of training data.

The findings provided in Table 3 suggest the following observations.

Our approach significantly outperforms most previous models on all splits of the data for both datasets, with even worse competitive losing accuracy when the amount of training data reduces. This emphasizes its good performance in low-data environments. Although alternative approaches like pFedFDA and FedPAC offer a good performance in some configurations of the problem, our proposed model still remains competitive overall, especially under limited data regimes.

Moreover, while pFedFDA may be considered as a strong reference in the worst case because of its robustness to distribution shift at different n, our model outperforms pFedFDA even under limited data setting which further demonstrates its efficacy on diverse distributions.

Table 3. Average (standard deviation) test accuracy on CIFAR10/100-S for varying proportions of training data

Method	CIFAR10-S 100%	CIFAR10-S 75%	CIFAR10-S 50%	CIFAR10-S 25%	CIFAR100-S 100%	CIFAR100-S 75%	CIFAR100-S 50%	CIFAR100-S 25%
Local Only	.586(.12)	.476(.16)	.461(.15)	.435(.14)	.157(.05)	.136(.05)	.123(.04)	.093(.04)
FedAvg	.464(.13)	.410(.19)	.389(.17)	.321(.14)	.233(.06)	.212(.06)	.187(.05)	.114(.04)
FedAvg FT	.682(.10)	.579(.19)	.561(.17)	.526(.16)	.302(.06)	.273(.05)	.241(.06)	.160(.05)
APFL	.611(.12)	.520(.17)	.508(.16)	.504(.16)	.164(.05)	.148(.04)	.131(.05)	.105(.04)
Ditto	.668(.10)	.578(.18)	.558(.17)	.527(.16)	.295(.05)	.274(.06)	.239(.05)	.141(.05)
FedBABU	.602(.12)	.522(.17)	.495(.16)	.467(.15)	.187(.05)	.170(.05)	.148(.05)	.107(.04)
FedPAC	.679(.09)	.642(.19)	.594(.16)	.533(.18)	.360(.07)	.330(.07)	.283(.07)	.162(.05)
FedRep	.612(.10)	.541(.17)	.510(.16)	.486(.16)	.176(.05)	.158(.05)	.131(.04)	.100(.04)
FedRoD	.655(.11)	.554(.18)	.537(.18)	.499(.14)	.218(.05)	.186(.05)	.150(.04)	.115(.04)
FedAvg	.584(.13)	.483(.16)	.466(.15)	.433(.14)	.166(.05)	.153(.05)	.127(.05)	.091(.04)
pFedMe	.679(.10)	.583(.18)	.549(.17)	.523(.16)	.289(.06)	.268(.06)	.237(.06)	.153(.05)
pFedFDA	.724(.09)	.706(.10)	.661(.11)	.595(.12)	.361(.08)	.342(.08)	.326(.08)	.227(.07)
Our model	.796(.71)	.767(.77)	.780(.78)	.765(.59)	.457(.55)	.395(.58)	.435(.61)	.368(.66)

Table 3 Average (standard deviation) test accuracy on multiple datasets

Method	EMNIST (Writers)	CIFAR10 Dir(0.1)	CIFAR10 Dir(0.5)	CIFAR100 Dir(0.1)	CIFAR100 Dir(0.5)
Local	.242(.23)	.865(.13)	.585(.13)	.368(.09)	.150(.05)
FedAvg	.790(.14)	.545(.12)	.625(.07)	.245(.06)	.252(.05)
FedAvgFT	.844(.10)	.902(.10)	.742(.08)	.499(.09)	.314(.06)
APFL	.841(.10)	.882(.11)	.656(.11)	.388(.09)	.169(.05)
Ditto	.843(.10)	.898(.10)	.736(.08)	.504(.08)	.308(.06)
FedBABU	.728(.13)	.887(.11)	.678(.11)	.395(.09)	.193(.04)
FedPAC	.856(.09)	.908(.09)	.767(.07)	.560(.08)	.378(.06)
FedRep	.735(.12)	.889(.10)	.668(.10)	.398(.09)	.182(.05)
FedRoD	.747(.15)	.885(.11)	.713(.09)	.424(.08)	.224(.05)
LG-FedAvg	.666(.13)	.866(.13)	.599(.12)	.381(.09)	.162(.05)
pFedMe	.842(.10)	.900(.10)	.740(.09)	.493(.08)	.311(.06)
pFedFDA	.844(.10)	.902(.09)	.763(.07)	.523(.08)	.385(.07)
Our model	.855(.12)	.909(.10)	.796(.71)	.630(.71)	.457(.55)

4.3.2. Performance Across Multiple Datasets

To study the robustness of our approach with respect to data heterogeneity, we tuned α in Dir(α) on CIFAR-100. This served as a way to regulate the level of Non-IID in clients. We next discussed several approaches, with a summary shown in Table 4. Table 4 shows that our proposed model achieves remarkable accuracy on the varying levels of data heterogeneity across CIFAR100 (with an accuracy of 0.630 in Dir(0.1) and 0.457 in Dir(0.5)). This verifies the generalization of our model when dealing with different types of data distributions and surpassing other methods under these cases.

We also examined scalability by testing the model on 25, 50, 75 and 100 clients. The performance of our model in Figure 2 and Figure 3 under different client settings on CIFAR-10 and CIFAR-100, respectively. Both plots indicate a continuous increase in accuracy as the number of clients increases over 200 rounds, with the simpler CIFAR-10 dataset matching higher accuracy compared to more complex CIFAR-100 dataset. These findings demonstrate that our proposed model is scalable and effective, thus we also achieve better than when with more participating clients.

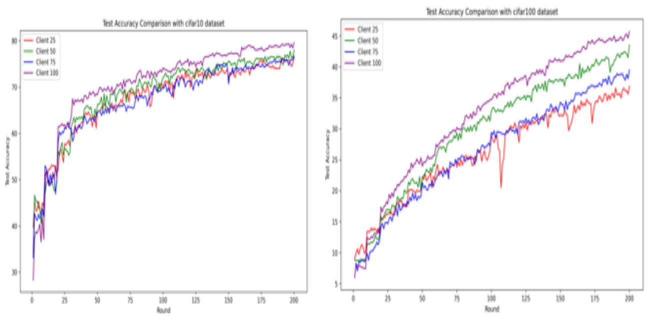

Figure 2. Test accuracy comparison between different clients

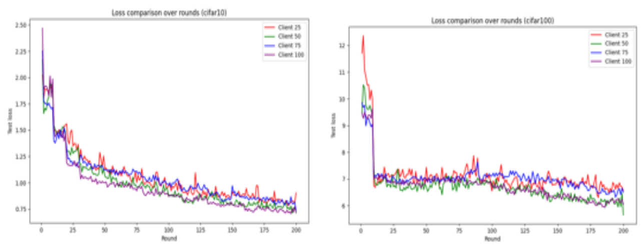

Figure 3 Test loss comparison between different clients

We illustrate the trade-off of test accuracy and training time with the number of clients for CIFAR-10, CIFAR-100 in Figure 4. This comparison illustrates the scale and effectiveness offered by federated learning, demonstrating that the number of clients strongly impacts both accuracy and training time. It also offers guidelines on how to find the right tradeoff between model performance and computational cost according to client characteristics.

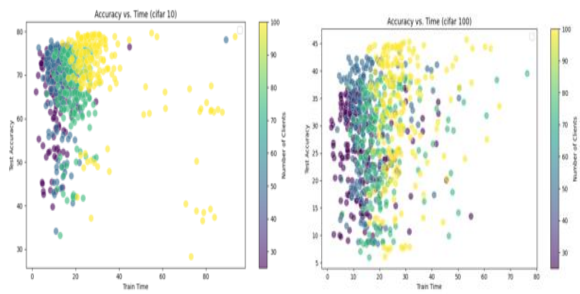

Figure 4 Comparison between clients count, time and accuracy

4.3.1. Significance of this study

The empirical results on numerous experiments demonstrate the importance and novelty of our proposed methodology within federated learning. First, the consistent superior performance across different datasets and data splits demonstrated that our approach successfully tackles the major challenges faced in FL – non-IID distributions, personalization, as well as scarce data. The FDA encourages feature discripability among clients and the meta-learning enables quick adaptation, thus our framework is adapted to real-world scenarios with heterogeneous and evolving user data.

Secondly, the stability of our model to extreme data heterogeneity, indicated by its better performance under low Dirichlet values, distinguishes our model from previous models that break down in the face of such heterogeneity. Third, the model's performance does not degrade much when data size is small and number of clients is large, which indicates a high scalability of this framework an essential feature for practical FL applications.

Last, we visualize the cooperation between clients, precision and training time to guide trade-off of computational cost and model performances. Collectively, this work introduces a generic flexible scalable FL solution and provides the basis for exploration of much more complex multimodal sequence-based FL schemes in future research. In summary, our results affirm that the proposed system is superior in terms of adaptability, personalization, robustness and scalability compared to prior federated learning solutions across diverse scenarios.

5. Conclusion and Future work

In this paper, we introduced a new personalized FL framework that can handle some of the core challenges in decentralized and heterogeneous learning settings. Classical federated learning methods usually fail when dealing with non-IID data distributions, and possess limited generalization ability, as well as inadequate personalization toward various clients. To address these limitations, the proposed framework combines Fisher Discriminant Analysis (FDA) and meta-learning into a homogeneous model capable of providing enhanced feature discrimination and fast adaptation. FDA allows mapping client-specific features to a maximally class-separable latent space which encourages better correspondence between local and global models. The meta-learning building block allows the model to fast-learn and efficiently adapt in rapidly varying client distributions, despite little local data. This framework also consists of a CNN-based hierarchical feature extractor, and can extract fine-grained and global semantic patterns to facilitate strong personalization and knowledge transfer across users. The framework is also developed with scalability in mind, and can be used in real-world scenarios with constantly increasing number of clients and complexity of data. Our approach is focused not only on technical merits, but also on privacy, due to the processing of latent feature space that naturally entails lower access on raw data. This study represents a significant advance toward more practical, adaptive and personalized federated learning. In the future, we plan to generalize our framework to handle multi-modal and sequential data, increase the adaptability of model via more sophisticated meta-learning approaches, and add stronger privacy protection such as differential privacy or secure aggregation. These planned additions will make our model more suitable for deployment in privacy critical domains such as healthcare, finance or mobile computing.

Funding

The work was supported by the National Natural Science Foundation of China under Grant 62372160.

Declaration of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

Conflicts of Interest

The author declares that there is no conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

References

[1] CHOWDHURY O., RISHAT M. A. S. A., AZAM M. H. B., and AMIN M. A. The rise of blockchain technology in Shariah based banking system. *Proceedings of the 2nd International Conference on Computing Advancements*, 2022: 349–358. https://doi.org/10.1145/3542954.3543005

[2] XU Y., LIAO Y., WANG L., XU H., JIANG Z., and ZHANG W. Overcoming noisy labels and non-iid data in edge federated learning. *IEEE Transactions on Mobile Computing*, 2024. https://doi.org/10.1109/TMC.2024.3398801

[3] HUANG G., WU Q., SUN P., MA Q., and CHEN X. Collaboration in federated learning with differential privacy: A stackelberg game analysis. *IEEE Transactions on Parallel and Distributed Systems*, 2024, 35(3): 455–469. https://doi.org/10.1109/TPDS.2024.3354713

[4] BABAR M., QURESHI B., and KOUBAA A. Review on federated learning for digital transformation in healthcare through big data analytics. *Future Generation Computer Systems*, 2024. https://doi.org/10.1016/j.future.2024.05.046

[5] QI Y., FENG Y., WANG X., LI H., and TIAN J. Leveraging federated learning and edge computing for recommendation systems within cloud computing networks. *arXiv*, 2024. https://doi.org/10.48550/arXiv.2403.03165

[6] CHOWDHURY O., RISHAT M. A. S. A., AL-AMIN M., and AZAM M. H. B. The decentralized Shariah-based banking system in Bangladesh using blockchain technology. *International Journal of Information Engineering and Electronic Business*, 2023, 15(3): 12–28. https://doi.org/10.5815/ijieeb.2023.03.02

[7] QUADIR B., MOSTAFA K., YANG J. C., SHEN J., and AKTER R. ARCS approach to PTA-based programming language practice sessions: Factors influencing programming problem-solving skills. *Education and Information Technologies*, 2023, 28(10): 13713–13735. https://doi.org/10.1007/s10639-023-11740-6

[8] ROKAYA A., ISLAM S. M. T., ZHANG H., SUN L., ZHU M., and ZHAO L. Acceptance of chatbot based on emotional intelligence through machine learning algorithm. *ICFEICT 2022*, 2022: 610–616. https://doi.org/10.1109/ICFEICT57213.2022.00111

[9] LU Z., PAN H., DAI Y., SI X., and ZHANG Y. Federated learning with non-iid data: A survey. *IEEE Internet of Things Journal*, 2024. https://doi.org/10.1109/JIOT.2024.3376548

[10] WU X., PEI J., HAN X. H., CHEN Y. W., YAO J., LIU Y., et al. Fedel: Federated ensemble learning for non-iid data. *Expert Systems with Applications*, 2024, 237: 121390. https://doi.org/10.1016/j.eswa.2023.121390

[11] ARIVAZHAGAN M. G., AGGARWAL V., SINGH A. K., and CHOUDHARY S. Federated learning with personalization layers. *arXiv*, 2019. https://doi.org/10.48550/arXiv.1912.00818

[12] GHARI P. M. and SHEN Y. Personalized federated learning with mixture of models for adaptive prediction and model fine-tuning. *arXiv*, 2024. https://doi.org/10.48550/arXiv.2410.21547

[13] ZHOU T., ZHANG J., and TSANG D. H. Fedfa: Federated learning with feature anchors to align features and classifiers for heterogeneous data. *IEEE Transactions on Mobile Computing*, 2023, 23(6): 6731–6742. https://doi.org/10.1109/TMC.2023.3325366

[14] YE Z., ZHANG X., CHEN X., XIONG H., and YU D. Adaptive clustering based personalized federated learning framework for next POI recommendation with location noise. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 36(5): 1843–1856. https://doi.org/10.1109/TKDE.2023.3312511

[15] CHIEN W. C., LIN C. H., ZHU T., DAI C., GARG S., and MUKHERJEE A. Density-clustering aggregation for personalized federated learning with AI-enabled aerial and edge computing in UAVs. *IEEE*

Internet of Things Journal, 2025. https://doi.org/10.1109/JIOT.2024.3522164
[16] TAN A. Z., YU H., CUI L., and YANG Q. Towards personalized federated learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 34(12): 9587–9603. https://doi.org/10.1109/TNNLS.2022.3160699
[17] ZHAO Y., PAN S., ZHAO Y., LIAO H., YE L., and ZHENG Y. Ultra-short-term wind power forecasting based on personalized robust federated learning with spatial collaboration. *Energy*, 2024, 288: 129847. https://doi.org/10.1016/j.energy.2023.129847
[18] MCLAUGHLIN C. and SU L. Personalized federated learning via feature distribution adaptation. *NeurIPS*, 2024, 37: 77038–77059. https://doi.org/10.52202/079017-2451
[19] GENG D., WANG S., and ZHANG Y. Multi-objective federated averaging algorithm. *Expert Systems*, 2025, 42(2): e13761. https://doi.org/10.1111/exsy.13761
[20] KARAMI M. and KARAMI A. Harmony in federated learning: A comprehensive review of techniques to tackle heterogeneity and non-IID data. *Cluster Computing*, 2025 (in press). https://doi.org/10.1007/s10586-025-05250-y
[21] THOMAS S. G. and MYAKALA P. K. Beyond the cloud: Federated learning and edge AI for the next decade. *Journal of Computer and Communications*, 2025, 13(2): 37–50. https://doi.org/10.4236/jcc.2025.132004
[22] ULLAH I., DENG X., PEI X., MUSHTAQ H., and UZAIR M. IoV-SFL: A blockchain-based federated learning framework for secure and efficient data sharing in the internet of vehicles. *Peer-to-Peer Networking and Applications*, 2025, 18(1): 1–20. https://doi.org/10.1007/s12083-024-01821-9
[23] LI R., WANG H., LU Q., YAN J., JI S., and MA Y. Research on medical image classification based on improved FedAvg algorithm. *Tsinghua Science and Technology*, 2025. https://doi.org/10.26599/TST.2024.9010184
[24] ZHANG X., HONG M., DHOPLE S., YIN W., and LIU Y. FedPD: A federated learning framework with adaptivity to non-IID data. *IEEE Transactions on Signal Processing*, 2021, 69: 6055–6070. https://doi.org/10.1109/TSP.2021.3115952
[25] DORIGUZZI-CORIN R. and SIRACUSA D. FLAD: Adaptive federated learning for DDoS attack detection. *Computers & Security*, 2024, 137: 103597. https://doi.org/10.1016/j.cose.2023.103597
[26] ZENG Y., MU Y., YUAN J., TENG S., ZHANG J., WAN J., et al. Adaptive federated learning with non-IID data. *The Computer Journal*, 2023, 66(11): 2758–2772. https://doi.org/10.1093/comjnl/bxac118
[27] AMIN M. S., KIM S., RISHAT M. A. S. A., TANG Z., and AHN H. A systematic literature review of privacy information disclosure in AI-integrated IoT technologies. *Sustainability*, 2024, 17(1): 8. https://doi.org/10.3390/su17010008
[28] ZHANG J., HUA Y., WANG H., SONG T., XUE Z., MA R., et al. GPFL: Simultaneously learning global and personalized feature information for personalized federated learning. *ICCV*, 2023: 5041–5051. https://doi.org/10.48550/arXiv.2308.10279
[29] JANG J. and CHOI B. J. Personalized federated learning via deviation tracking representation learning. *ICOIN 2024*, 2024: 762–766. https://doi.org/10.1109/ICOIN59985.2024.10572208
[30] SHEN W., WU S., and TAO Y. CLDP-pFedAvg: Safeguarding client data privacy in personalized federated averaging. *Mathematics*, 2024, 12(22): 3630. https://doi.org/10.3390/math12223630
[31] XU J., TONG X., and HUANG S. L. Personalized federated learning with feature alignment and classifier collaboration. *arXiv*, 2023. https://doi.org/10.48550/arXiv.2306.11867
[32] WANG J., LIU Q., LIANG H., JOSHI G., and POOR H. V. Tackling the objective inconsistency problem in heterogeneous federated optimization. *NeurIPS*, 2020, 33: 7611–7623. https://doi.org/10.48550/arXiv.2007.07481
[33] MCMAHAN B., MOORE E., RAMAGE D., HAMPSON S., and Y ARCAS B. A. Communication-efficient learning of deep networks from decentralized data. *AISTATS*, 2017: 1273–1282. https://doi.org/10.48550/arXiv.1602.05629
[34] KHAN U., PAO W., PILARIO K. E. S., SALLIH N., and KHAN M. R. Two-phase flow regime identification using multi-method feature extraction and explainable kernel Fisher discriminant analysis. *International Journal of Numerical Methods for Heat & Fluid Flow*, 2024, 34(8): 2836–2864. https://doi.org/10.1108/HFF-09-2023-0526
[35] BACCOUR E., ERBAD A., MOHAMED A., HAMDI M., and GUIZANI M. A blockchain-based reliable federated meta-learning for metaverse: A dual game framework. *IEEE Internet of Things Journal*, 2024. https://doi.org/10.1109/JIOT.2024.3383096

参考文

[1] CHOWDHURY O., RISHAT M. A. S. A., AZAM M. H. B., AMIN M. A. 基于区块链技术的沙里亚银行体系的兴起. Proceedings of the 2nd International Conference on Computing Advancements, 2022: 349–358. https://doi.org/10.1145/3542954.3543005
[2] XU Y., LIAO Y., WANG L., XU H., JIANG Z., ZHANG W. 在边缘联邦学习中克服噪声标签和非独立同分布数据问题. IEEE Transactions on Mobile Computing, 2024. https://doi.org/10.1109/TMC.2024.3398801

[3] HUANG G., WU Q., SUN P., MA Q., CHEN X. 差分隐私下的联邦学习协作：一种斯塔克尔伯格博弈分析. IEEE Transactions on Parallel and Distributed Systems, 2024, 35(3): 455–469. https://doi.org/10.1109/TPDS.2024.3354713
[4] BABAR M., QURESHI B., KOUBAA A. 运用大数据分析推动医疗数字化转型的联邦学习综述. Future Generation Computer Systems, 2024. https://doi.org/10.1016/j.future.2024.05.046
[5] QI Y., FENG Y., WANG X., LI H., TIAN J. 基于联邦学习与边缘计算的云计算网络推荐系统. arXiv, 2024. https://doi.org/10.48550/arXiv.2403.03165
[6] CHOWDHURY O., RISHAT M. A. S. A., AL-AMIN M., AZAM M. H. B. 使用区块链技术的孟加拉国去中心化沙里亚银行体系. International Journal of Information Engineering and Electronic Business, 2023, 15(3): 12–28. https://doi.org/10.5815/ijieeb.2023.03.02
[7] QUADIR B., MOSTAFA K., YANG J. C., SHEN J., AKTER R. 基于 PTA 的编程语言练习课的 ARCS 方法：影响编程问题解决能力的因素. Education and Information Technologies, 2023, 28(10): 13713–13735. https://doi.org/10.1007/s10639-023-11740-6
[8] ROKAYA A., ISLAM S. M. T., ZHANG H., SUN L., ZHU M., ZHAO L. 基于情绪智能和机器学习算法的聊天机器人接受度研究. ICFEICT 2022, 2022: 610–616. https://doi.org/10.1109/ICFEICT57213.2022.00111
[9] LU Z., PAN H., DAI Y., SI X., ZHANG Y. 面向非独立同分布数据的联邦学习综述. IEEE Internet of Things Journal, 2024. https://doi.org/10.1109/JIOT.2024.3376548
[10] WU X., PEI J., HAN X. H., CHEN Y. W., YAO J., LIU Y., et al. FedEL：用于非独立同分布数据的联邦集成学习. Expert Systems with Applications, 2024, 237: 121390. https://doi.org/10.1016/j.eswa.2023.121390
[11] ARIVAZHAGAN M. G., AGGARWAL V., SINGH A. K., CHOUDHARY S. 带有个性化层的联邦学习. arXiv, 2019. https://doi.org/10.48550/arXiv.1912.00818
[12] GHARI P. M., SHEN Y. 用模型混合实现自适应预测与模型微调的个性化联邦学习. arXiv, 2024. https://doi.org/10.48550/arXiv.2410.21547
[13] ZHOU T., ZHANG J., TSANG D. H. FedFA：利用特征锚点对齐异构数据中的特征和分类器的联邦学习. IEEE Transactions on Mobile Computing, 2023, 23(6): 6731–6742. https://doi.org/10.1109/TMC.2023.3325366
[14] YE Z., ZHANG X., CHEN X., XIONG H., YU D. 基于自适应聚类的个性化联邦学习：具备位置噪声的下一地点推荐. IEEE Transactions on Knowledge and Data Engineering, 2023, 36(5): 1843–1856. https://doi.org/10.1109/TKDE.2023.3312511
[15] CHIEN W. C., LIN C. H., ZHU T., DAI C., GARG S., MUKHERJEE A. 面向无人机的 AI 增强空地计算中基于密度聚类聚合的个性化联邦学习. IEEE Internet of Things Journal, 2025. https://doi.org/10.1109/JIOT.2024.3522164
[16] TAN A. Z., YU H., CUI L., YANG Q. 走向个性化联邦学习. IEEE Transactions on Neural Networks and Learning Systems, 2022, 34(12): 9587–9603. https://doi.org/10.1109/TNNLS.2022.3160699
[17] ZHAO Y., PAN S., ZHAO Y., LIAO H., YE L., ZHENG Y. 基于个性化鲁棒联邦学习与空间协作的超短期风电预测. Energy, 2024, 288: 129847. https://doi.org/10.1016/j.energy.2023.129847
[18] MCLAUGHLIN C., SU L. 基于特征分布自适应的个性化联邦学习. NeurIPS, 2024, 37: 77038–77059. https://doi.org/10.52202/079017-2451
[19] GENG D., WANG S., ZHANG Y. 多目标联邦平均算法. Expert Systems, 2025, 42(2): e13761. https://doi.org/10.1111/exsy.13761
[20] KARAMI M., KARAMI A. 联邦学习中应对异质性与非独立同分布数据的技术综述. Cluster Computing, 2025（待刊）. https://doi.org/10.1007/s10586-025-05250-y
[21] THOMAS S. G., MYAKALA P. K. 超越云端：面向未来十年的联邦学习与边缘人工智能. Journal of Computer and Communications, 2025, 13(2): 37–50. https://doi.org/10.4236/jcc.2025.132004
[22] ULLAH I., DENG X., PEI X., MUSHTAQ H., UZAIR M. IoV-SFL：一种基于区块链的车联网数据安全共享联邦学习框架. Peer-to-Peer Networking and Applications, 2025, 18(1): 1–20. https://doi.org/10.1007/s12083-024-01821-9
[23] LI R., WANG H., LU Q., YAN J., JI S., MA Y. 基于改进 FedAvg 算法的医疗图像分类研究. Tsinghua Science and Technology, 2025. https://doi.org/10.26599/TST.2024.9010184
[24] ZHANG X., HONG M., DHOPLE S., YIN W., LIU Y. FedPD：一种适应非独立同分布数据的联邦学习框架. IEEE Transactions on Signal Processing, 2021, 69: 6055–6070. https://doi.org/10.1109/TSP.2021.3115952
[25] DORIGUZZI-CORIN R., SIRACUSA D. FLAD：用于 DDoS 攻击检测的自适应联邦学习. Computers & Security, 2024, 137: 103597. https://doi.org/10.1016/j.cose.2023.103597
[26] ZENG Y., MU Y., YUAN J., TENG S., ZHANG J., WAN J., et al. 面向非独立同分布数据的自适应联邦学习. The Computer Journal, 2023, 66(11): 2758–2772. https://doi.org/10.1093/comjnl/bxac118
[27] AMIN M. S., KIM S., RISHAT M. A. S. A., TANG Z., AHN H. AI 融合物联网技术中的隐私信息披露系统文献综述. Sustainability, 2024, 17(1): 8. https://doi.org/10.3390/su17010008
[28] ZHANG J., HUA Y., WANG H., SONG T., XUE Z., MA R., et al. GPFL：同时学习全局与个性化特征信息的个性化联邦学习. ICCV, 2023: 5041–5051. https://doi.org/10.48550/arXiv.2308.10279
[29] JANG J., CHOI B. J. 通过偏差跟踪表示学习的个性化联邦学习. ICOIN 2024, 2024: 762–766. https://doi.org/10.1109/ICOIN59985.2024.10572208

[30] SHEN W., WU S., TAO Y. CLDP-pFedAvg：在个性化联邦平均中保护客户端数据隐私. Mathematics, 2024, 12(22): 3630. https://doi.org/10.3390/math12223630
[31] XU J., TONG X., HUANG S. L. 具备特征对齐与分类器协作的个性化联邦学习. arXiv, 2023. https://doi.org/10.48550/arXiv.2306.11867
[32] WANG J., LIU Q., LIANG H., JOSHI G., POOR H. V. 解决异构联邦优化中的目标不一致问题. NeurIPS, 2020, 33: 7611–7623. https://doi.org/10.48550/arXiv.2007.07481
[33] MCMAHAN B., MOORE E., RAMAGE D., HAMPSON S., Y ARCAS B. A. 从分散数据中高效通信地学习深度网络. AISTATS, 2017: 1273–1282. https://doi.org/10.48550/arXiv.1602.05629
[34] KHAN U., PAO W., PILARIO K. E. S., SALLIH N., KHAN M. R. 基于多方法特征提取和可解释核Fisher判别分析的两相流态识别. International Journal of Numerical Methods for Heat & Fluid Flow, 2024, 34(8): 2836–2864. https://doi.org/10.1108/HFF-09-2023-0526
[35] BACCOUR E., ERBAD A., MOHAMED A., HAMDI M., GUIZANI M. 面向元宇宙的基于区块链的可靠联邦元学习：双博弈框架. IEEE Internet of Things Journal, 2024. https://doi.org/10.1109/JIOT.2024.3383096

Word count: 10343 words, excluding references.

Peer-review record:
Fast-track status: Not fast-tracked
First-round reviews received: 3 reports
Revision cycles completed: 3 rounds
Final version submitted: December 8, 2025