AnaPFL: When Closed-Form Solutions Meet Generalization and Personalization in Personalized Federated Learning

Kejia Fan[1], Jianheng Tang[†2,7], Zhirui Yang[1], Yajiang Huang[1], Feijiang Han[3], Run He[4], Jiaxu Li[1], Songning Lai[5], Anfeng Liu[1], Houbing Song[6], Yunhuai Liu[†2,7], and Huiping Zhuang[4]

[1]Central South University, China [2]Peking University, China
[3]University of Maryland, USA [4]South China University of Technology, China
[5]The Hong Kong University of Science and Technology (Guangzhou), China
[6]University of Maryland, Baltimore County, USA
[7]Key Laboratory of High Confidence Software Technologies (PKU), China
tangentheng@gmail.com, yunhuai.liu@pku.edu.cn

Abstract. Personalized Federated Learning (PFL) has emerged as a prevalent paradigm to deliver personalized models to individual clients through collaborative training. Existing PFL methods often suffer from the issue of Non-IID data, due to their reliance on gradient-based updates. Recently, Analytic Learning (AL) has exhibited great potential to address this issue via analytical (i.e., closed-form) solutions in a gradient-free manner. However, there remains a significant gap in introducing AL into PFL, owing to the encountered generalization-personalization dilemma. In this paper, to bridge this gap and address the associated challenges, we propose an **Ana**lytic **P**ersonalized **F**ederated **L**earning approach, named **AnaPFL**, for addressing the Non-IID issue in PFL by introducing and advancing AL. In AnaPFL, we develop dual-stream analytic models with closed-form solutions, including (1) a shared primary stream for global generalization across all clients, and (2) a dedicated refinement stream for local personalization of each client. Experimentally, we give comprehensive results to show AnaPFL's superior performance with over 99% efficiency advantages against gradient-based methods.

Keywords: Personalized Federated Learning · Closed-Form Solution · Non-IID Data · Analytic Learning · Personalization

1 Introduction

Federated Learning (FL) emerges as a prominent paradigm for distributed machine learning, enabling the collaborative training of a global model across multiple clients while preserving individual data privacy [9, 22, 30]. Nevertheless, as shown in Figure 1, a single global model in traditional FL often falls short of

[†]Corresponding authors.

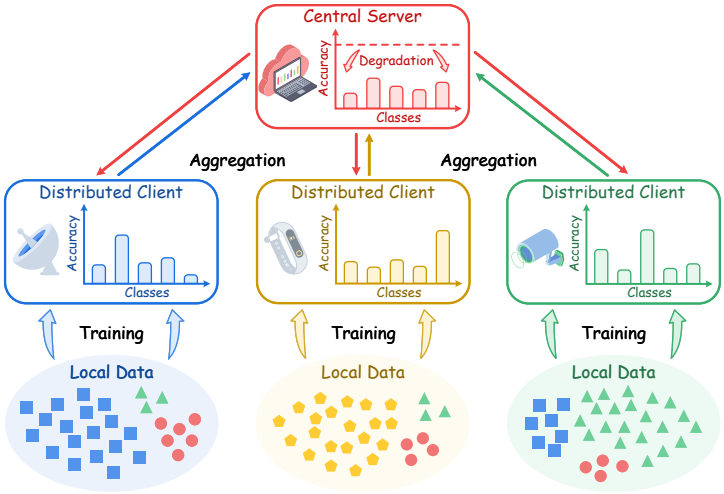

Fig. 1: Data heterogeneity across clients in PFL.

adapting to the diverse and personalized requirements of each client's local applications [32, 47]. In this context, Personalized Federated Learning (PFL) has emerged as an extended paradigm of FL, aiming to construct a personalized model for each client alongside the global model shared across clients [24, 32, 47].

The appeal of PFL lies in its potential to deliver the dual benefits of generalization and personalization [45]. First, it enables each local client to obtain the collective knowledge shared by all participants in the federation, enhancing the model's overall generalization capabilities [47]. Second, it facilitates individual personalization to allow models to adapt specifically to the unique distribution characteristics of the clients' local data, achieving targeted applications [2].

Data heterogeneity, often referred to as Non-Independent and Identically Distributed (Non-IID) data, represents a key challenge in PFL [24, 32]. In a Non-IID environment, the local gradients of client-side training are naturally biased towards their unique local data distributions. Such bias leads to distinct local optima and conflicting learning directions among clients. Hence, as illustrated in Figure 1, the diverse local datasets can cause the aggregated global model to drift away from a truly generalized representation, thus degrading its overall performance [2, 24, 32]. Many existing PFL methods are typically susceptible to this issue, which severely hinders collective generalization and, in turn, compromises the subsequent personalization efforts.

A core challenge in PFL is widely considered to be rooted in the long-standing reliance on gradient-based updates, which are inherently vulnerable under Non-IID data distributions [9, 11, 24, 46]. This observation has been broadly acknowledged in prior FL and PFL studies, and many influential works seek to alleviate the Non-IID issue by mitigating or compensating for the adverse effects of biased gradients [11, 32, 36, 39, 45, 47]. However, relatively few studies attempt to

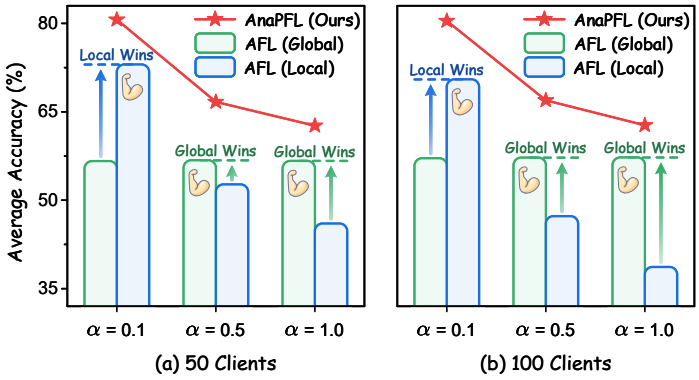

Fig. 2: Generalization-Personalization dilemma of AFL.

address Non-IID effects by explicitly circumventing gradient-based updates in PFL. To address the Non-IID problem more fundamentally, it is an interesting and promising direction to explore avoiding gradient-based updates (i.e., learning in a gradient-free manner). As a prominent gradient-free technique [4, 36], Analytic Learning (AL) has recently shown great promise in FL with closed-form solutions, leading to the recent approach named Analytic Federated Learning (AFL) [11]. Yet, AFL is severely limited by its foundation in traditional FL, suffering from the generalization-personalization dilemma in PFL.

As shown in Figure 2, we record the average accuracy of the AFL models across varying degrees of heterogeneity in PFL, using the Dirichlet distribution on CIFAR-100. When the degree of heterogeneity is low ($\alpha = 0.5, 1.0$), the global model performs better, as it aggregates more diverse data from the federation. Conversely, under high heterogeneity ($\alpha = 0.1$), the local model surprisingly surpasses the global model. This inversion implies that AFL is nearly ineffective under severe heterogeneity in PFL, as its global model is even inferior to the local models without FL aggregation. Thus, there remains a significant gap in introducing AL into PFL due to this generalization-personalization dilemma.

In this paper, to bridge this research gap and address the challenges, we propose an **Ana**lytic **P**ersonalized **F**ederated **L**earning approach, named AnaPFL, for addressing the Non-IID issue in PFL by introducing gradient-free AL. In our AnaPFL, we first utilize a foundation model as a frozen backbone for feature extraction. Then, we further develop our dual-stream analytic model with analytical (i.e., closed-form) solutions. Specifically, the dual-stream model in our AnaPFL includes: (1) a shared primary stream for global generalization across all clients, and (2) a dedicated refinement stream for local personalization of each individual client. To enhance AnaPFL's fitting ability in generalization and personalization, we employ random projections and nonlinear activations for dual streams to form two distinct feature spaces. As shown in Figure 2, our AnaPFL consistently outperforms AFL across all settings, demonstrating a superior and robust balance between generalization and personalization.

In summary, our key contributions in this paper are as follows:

1. Conceptually, our AnaPFL investigates and brings AL into PFL to address the Non-IID issue while balancing generalization and personalization.
2. Technically, we develop dual analytic streams based on least squares to achieve both collective generalization and individual personalization in PFL.
3. Experimentally, AnaPFL shows superior performance on benchmark datasets, outperforming the SOTA baselines with 1.57%–16.71% accuracy improvements, and over *99% efficiency advantages* against gradient-based methods.

2 Related Work

2.1 Personalized Federated Learning

FL emerges as a prominent distributed machine learning paradigm, allowing clients to collaboratively train a shared global model [9, 11, 24, 46]. FL is inherently compatible with the decentralized data acquisition mechanism of mobile crowd sensing in the Internet of Things [3, 5, 37]. In such scenarios, mobile edge devices continuously generate sensory observations and naturally serve as distributed FL clients. Meanwhile, the data collected by these devices often contain sensitive user- or context-related information, making direct centralized collection impractical or undesirable [4, 13, 34]. Therefore, FL can be interpreted as a distributed knowledge integration paradigm, where collective intelligence is extracted and aggregated from multiple sensing participants without requiring raw data sharing. Nonetheless, this "one-size-fits-all" style of traditional FL often falls short when clients have diverse data distributions and unique application requirements [32,39,45,47]. In this context, PFL has recently emerged promise to extend the traditional FL, aiming to construct models that are not only globally informed but also individually tailored to each client's requirement [2].

PFL's core appeal lies in its dual promise for both collective generalization and individual personalization [24, 47]. Despite the promising prospects offered by PFL, a critical accompanying challenge is the presence of Non-IID data across clients. It is widely recognized that this challenge in PFL stems from the vulnerability of gradient-based updates to Non-IID data [32,36]. Many significant efforts have been devoted to addressing this issue, but relatively few studies attempt to directly circumvent gradient-based updates in PFL [24, 31]. Differing from prior gradient-based PFL research, our AnaPFL aims to fundamentally avoid gradient-based updates through analytical solutions, enabling it to achieve the ideal invariance to heterogeneity. Notably, our AnaPFL mainly aims at efficient PFL over pretrained frozen features, based on AFL [11], rather than replacing all gradient-based methods with trainable representations [1, 28]. Therefore, although many effective representation-learning-based PFL methods have been developed [1, 20, 26, 28], they address a setting that is different from the one considered in this work. In our setting, the feature extractor remains fixed, and the key challenge is how to simultaneously enable global generalization and local personalization without relying on iterative gradient-based optimization.

2.2 Analytic Learning

AL is a popular gradient-free alternative to traditional backpropagation to circumvent the gradient-related issues such as vanishing and exploding gradients [17, 36, 49, 51]. AL is also referred to as pseudoinverse learning due to its reliance on matrix inversion [4, 8, 29, 50]. Its core idea is to directly derive analytical (i.e., closed-form) solutions utilizing least squares [4, 17, 36, 52]. With the utilization of the block-wise recursive Moore-Penrose inverse [50], AL has shown superiority in many modern applications, particularly in continual learning [15, 23, 51]. While AFL has introduced AL into FL [11], it is not well-suited to the PFL scenarios, where it suffers from the generalization-personalization dilemma, as shown in Figure 2. In addition to AFL, several recent studies have investigated the efficiency and effectiveness of closed-form solutions in FL [6, 35, 36]. Specifically, [6] leverages analytic classifiers to accelerate heterogeneous FL. Meanwhile, [35] extends analytic classifiers from static FL settings to dynamic continual FL environments with incremental data, and [36] enhances the representation learning capacity of closed-form classifiers by deepening the analytic networks. However, to the best of our knowledge, no existing work has explicitly investigated how to reconcile generalization and personalization in PFL while preserving analytical solutions. Our AnaPFL bridges this gap via dual-stream least squares, achieving both generalization and personalization in PFL. Notably, our AnaPFL is *orthogonal* to the aforementioned AFL-based advances [35, 36], and can be readily combined with them to support continual settings and stronger representations.

3 Our Proposed AnaPFL

We consider the standard PFL scenario with one server and K clients. Each client k possesses its local dataset $D_k = \{\mathbf{X}_k, \mathbf{Y}_k\}$, where $\mathbf{X}_k \in \mathbb{R}^{N_k \times l \times w \times h}$ and $\mathbf{Y}_k \in \mathbb{R}^{N_k \times c}$ represent the N_k local data samples and their corresponding labels, respectively. Here, $l \times w \times h$ denotes the three dimensions of input images, and c is the number of classes. On the one hand, clients aim to improve their models' generalization by leveraging collective knowledge within the federation. On the other hand, the clients seek to enhance their models' personalization capabilities by adapting to the unique characteristics of their own local data.

3.1 Overview

In our AnaPFL, we aim to fundamentally avoid the reliance on gradient-based updates in PFL, with the help of gradient-free AL. Taking into account the dual requirements of generalization and personalization in PFL, we design our dual-stream analytic model to achieve a balance. The overall design of our AnaPFL is illustrated in Figure 3. Motivated by the widespread use of *pre-trained foundation* models in FL [40, 41], we employ a *self-supervised foundation model*, e.g., ViT-MAE [10], as the frozen backbone for feature extraction. The server can construct it using non-private datasets or simply by downloading a public model, with

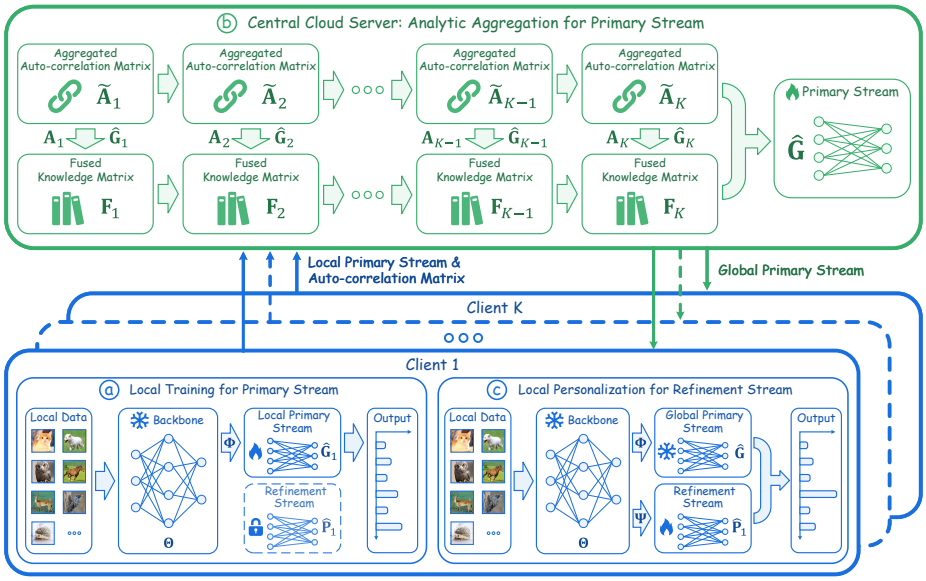

Fig. 3: Detailed design of our proposed AnaPFL.

few privacy concerns. Following this backbone, we develop the primary stream $\hat{\mathbf{G}}$ to aggregate collective knowledge from all K clients in the federation for enhancing global generalization. Finally, for each client k, we further construct the individual refinement stream $\hat{\mathbf{P}}_k$ to capture and adapt to the client's unique local preferences, thereby enabling local personalization.

Thus, our goal is to obtain a primary stream $\hat{\mathbf{G}}$ and a total of K refinement streams $\{\hat{\mathbf{P}}_k\}_{k=1}^{K}$, as formulated below.

Minimize:

$$\sum\nolimits_{k=1}^{K} \|(\mathbf{Y}_k - \mathbf{\Phi}_k\hat{\mathbf{G}}) - \mathbf{\Psi}_k\hat{\mathbf{P}}_k\|_{\mathrm{F}}^2 + \beta \sum\nolimits_{k=1}^{K} \|\hat{\mathbf{P}}_k\|_{\mathrm{F}}^2, \tag{1}$$

Subject to:

$$\hat{\mathbf{G}} = \arg\min_{\mathbf{G}} \|\mathbf{Y}_{1:K} - \mathbf{\Phi}_{1:K}\mathbf{G}\|_{\mathrm{F}}^2 + \gamma\|\mathbf{G}\|_{\mathrm{F}}^2. \tag{2}$$

Here, $\mathbf{\Phi}_k$ and $\mathbf{\Psi}_k$ represent the client k's activated feature matrices, both derived from the same frozen backbone but subsequently passed via distinct random projections and nonlinear activation functions. The objective (1) serves for the refinement streams to minimize the sum of each client's local empirical risk for local personalization. The constraint (2) ensures that the primary stream minimizes the global empirical risk for achieving global generalization.

Notably, since the Cross-Entropy (CE) loss does not admit analytical solutions, we adopt the Mean Squared Error (MSE) as the loss. In fact, recent studies have broadly shown that the MSE loss is more robust and can achieve performance comparable to or even better than the common CE loss [7, 11, 17, 48]. Then, using least squares, we devise the corresponding computation process for

our AnaPFL to derive the analytical solutions to (2) and (1), thereby attaining superiority in both generalization and personalization.

3.2 Primary Stream for Global Generalization

To begin with, all the clients collaboratively train a shared primary stream for global generalization. To reach this goal, we meticulously design the computation and aggregation process to derive the closed-form solution to (2) analytically, rendering the primary-stream solution invariant to client-wise data partitioning.

Specifically, each client k first uses the established backbone of the foundation model for feature extraction, thereby obtaining its primary-stream feature matrix $\mathbf{\Phi}_k$ as follows:

$$\mathbf{\Phi}_k = \sigma_{\mathrm{P}}(\mathrm{Backbone}(\mathbf{X}_k, \Theta)\mathbf{R}_{\mathrm{P}}), \tag{3}$$

Here $\mathbf{\Phi}_k \in \mathbb{R}^{N_k \times d_{\mathrm{P}}}$ denotes the activated feature matrix of client k and d_{P} is the corresponding feature dimension. Meanwhile, $\mathrm{Backbone}(\cdot, \Theta)$ represents the backbone with its parameters Θ, while $\mathbf{R}_{\mathrm{P}}$ and $\sigma_{\mathrm{P}}(\cdot)$ represent the random projection matrix (initialized using Gaussian values) and the nonlinear activation, which are all shared among clients.

Subsequently, the client utilizes its local activated feature $\mathbf{\Phi}_k$ and label $\mathbf{Y}_k$ to construct its local primary stream $\hat{\mathbf{G}}_k$. Specifically, the problem of training the local primary stream can be formulated as follows:

$$\hat{\mathbf{G}}_k = \arg\min_{\mathbf{G}_k} \|\mathbf{Y}_k - \mathbf{\Phi}_k\mathbf{G}_k\|_{\mathrm{F}}^2 + \gamma\|\mathbf{G}_k\|_{\mathrm{F}}^2, \tag{4}$$

where γ is the regularization parameter and $\|\cdot\|_{\mathrm{F}}^2$ is the Frobenius norm. The analytical solution to (4) can be directly derived using the least squares method:

$$\hat{\mathbf{G}}_k = (\mathbf{\Phi}_k^\top\mathbf{\Phi}_k + \gamma\mathbf{I})^{-1}\mathbf{\Phi}_k^\top\mathbf{Y}_k. \tag{5}$$

Concurrently, the client k computes its *Auto-correlation Matrix* according to (6), and uploads this matrix $\mathbf{A}_k$ and its local primary stream $\hat{\mathbf{G}}_k$ to the server.

$$\mathbf{A}_k = \mathbf{\Phi}_k^\top\mathbf{\Phi}_k + \gamma\mathbf{I}. \tag{6}$$

Upon receiving $\{\mathbf{A}_k\}_{k=1}^K$ and $\{\hat{\mathbf{G}}_k\}_{k=1}^K$ from the clients, the server can aggregate all the clients' local knowledge and produce the global primary stream $\hat{\mathbf{G}}$. Specifically, the server first updates the *Aggregated Auto-correlation Matrix*:

$$\tilde{\mathbf{A}}_k = \tilde{\mathbf{A}}_{k-1} + \mathbf{A}_k = \sum\nolimits_{i=1}^{k} \mathbf{A}_i, \quad k > 1. \tag{7}$$

Then, the server produces the *Fused Knowledge Matrix*, which is initialized with $\mathbf{F}_1 = \hat{\mathbf{G}}_1$ and recursively updated to include each client's local knowledge:

$$\mathbf{F}_k = \mathbf{\Lambda}_k\mathbf{F}_{k-1} + \mathbf{\Delta}_k\hat{\mathbf{G}}_k, \quad k > 1, \tag{8}$$

where

$$\begin{cases} \mathbf{\Lambda}_k = \mathbf{I} - (\tilde{\mathbf{A}}_{k-1})^{-1}\mathbf{A}_k(\mathbf{I} - (\tilde{\mathbf{A}}_k)^{-1}\mathbf{A}_k), \\ \mathbf{\Delta}_k = \mathbf{I} - (\mathbf{A}_k)^{-1}\tilde{\mathbf{A}}_{k-1}(\mathbf{I} - (\tilde{\mathbf{A}}_k)^{-1}\tilde{\mathbf{A}}_{k-1}). \end{cases} \tag{9}$$

Once the total knowledge from all K clients has been integrated into the final $\mathbf{F}_K$, the server computes the global primary stream $\hat{\mathbf{G}}$ via (10):

$$\hat{\mathbf{G}} = [\tilde{\mathbf{A}}_K - (K-1)\gamma\mathbf{I}]^{-1}\tilde{\mathbf{A}}_K\mathbf{F}_K. \tag{10}$$

Lastly, the server distributes the primary stream $\hat{\mathbf{G}}$ to all clients. It is worth noting that the resulting global model $\hat{\mathbf{G}}$ exactly equals the optimal solution for (2), as established by Theorem 1 in Section 3.4. More encouragingly, the final result $\hat{\mathbf{G}}$ is also independent of the order of clients. This order-invariance also allows the server to update the *Fused Knowledge Matrix* asynchronously upon client arrivals, without requiring a fixed aggregation order.

3.3 Refinement Stream for Local Personalization

After receiving the primary stream $\hat{\mathbf{G}}$, each client k leverages its local data to construct the corresponding refinement stream for local personalization. Similarly, we also derive the closed-form solution of (1) to serve as each client's individual refinement stream, aiming at further correcting the prediction bias of the primary stream on the local dataset.

Specifically, each client k first extracts its refinement-stream feature matrix $\mathbf{\Psi}_k$ using a distinct random projection and activation function to enhance the linear separability:

$$\mathbf{\Psi}_k = \sigma_{\mathrm{R}}(\mathrm{Backbone}(\mathbf{X}_k, \Theta)\mathbf{R}_{\mathrm{R}}). \tag{11}$$

Here, $\mathbf{\Psi}_k \in \mathbb{R}^{N_k \times d_{\mathrm{R}}}$ denotes the feature matrix for the client k's refinement stream, while σ_{R} and $\mathbf{R}_{\mathrm{R}}$ are the corresponding activation function and random projection matrix. To form distinct feature spaces, $\mathbf{R}_{\mathrm{R}}$ is different from $\mathbf{R}_{\mathrm{P}}$. Moreover, σ_{R} and $\mathbf{R}_{\mathrm{R}}$ can be client-specific for privacy.

Subsequently, the client k uses the obtained global primary stream $\hat{\mathbf{G}}$, the activated feature matrices $\mathbf{\Phi}_k$ and $\mathbf{\Psi}_k$, and the label matrix $\mathbf{Y}_k$ to construct its individual refinement stream $\hat{\mathbf{P}}_k$. The overall optimization objective in (1) can then be formulated locally for each client as:

$$\hat{\mathbf{P}}_k = \arg\min_{\mathbf{P}_k} \|(\mathbf{Y}_k - \mathbf{\Phi}_k\hat{\mathbf{G}}) - \mathbf{\Psi}_k\mathbf{P}_k\|_{\mathrm{F}}^2 + \beta\|\mathbf{P}_k\|_{\mathrm{F}}^2, \tag{12}$$

Here, β denotes the regularization parameter for the refinement stream. Similarly, by leveraging the least squares method, the closed-form solution of (12) can be expressed via (13), as established by Theorem 2 in Section 3.4.

$$\hat{\mathbf{P}}_k = (\mathbf{\Psi}_k^\top\mathbf{\Psi}_k + \beta\mathbf{I})^{-1}(\mathbf{\Psi}_k^\top\mathbf{Y}_k - \mathbf{\Psi}_k^\top\mathbf{\Phi}_k\hat{\mathbf{G}}). \tag{13}$$

Thus, the primary stream $\hat{\mathbf{G}}$ and refinement stream $\hat{\mathbf{P}}_k$ are combined to construct the final model for each client k. During inference, for any input $\hat{\mathbf{X}}_k$, the corresponding inference result $\hat{\mathbf{Y}}_k$ is computed as:

$$\hat{\mathbf{Y}}_k = \hat{\mathbf{\Phi}}_k \hat{\mathbf{G}} + \lambda \hat{\mathbf{\Psi}}_k \hat{\mathbf{P}}_k, \tag{14}$$

where $\hat{\mathbf{\Phi}}_k$ and $\hat{\mathbf{\Psi}}_k$ are the activated feature matrices obtained from $\hat{\mathbf{X}}_k$ via (3) and (11), respectively. The balance parameter $\lambda \in (0, 1]$ controls the trade-off between generalization and personalization, where a larger λ places greater emphasis on personalization.

3.4 Theoretical Analyses

Here, we provide detailed theoretical analyses of our AnaPFL. First, in Lemma 1, we establish the validity of the recursive update formulation for the *Fused Knowledge Matrix* and derive its closed-form analytical expression. Then, we prove that the primary stream and refinement streams obtained by our AnaPFL correspond exactly to the analytical solutions of (2) and (1), respectively. Finally, in Theorem 3, we establish the desirable property of *invariance to heterogeneity* enjoyed by AnaPFL and the corresponding corollary.

Lemma 1: Consider the following recursive update formulation for the *Fused Knowledge Matrix*:

$$\mathbf{F}_k = \mathbf{\Lambda}_k \mathbf{F}_{k-1} + \mathbf{\Delta}_k \hat{\mathbf{G}}_k, \tag{15}$$

where

$$\begin{cases} \mathbf{\Lambda}_k = \mathbf{I} - (\tilde{\mathbf{A}}_{k-1})^{-1} \mathbf{A}_k (\mathbf{I} - (\tilde{\mathbf{A}}_k)^{-1} \mathbf{A}_k), \\ \mathbf{\Delta}_k = \mathbf{I} - (\mathbf{A}_k)^{-1} \tilde{\mathbf{A}}_{k-1} (\mathbf{I} - (\tilde{\mathbf{A}}_k)^{-1} \tilde{\mathbf{A}}_{k-1}). \end{cases} \tag{16}$$

The resulting *Fused Knowledge Matrix* $\mathbf{F}_k$ admits the closed-form expression:

$$\mathbf{F}_k = (\mathbf{\Phi}_{1:k}^\top \mathbf{\Phi}_{1:k} + k\gamma \mathbf{I})^{-1} \mathbf{\Phi}_{1:k}^\top \mathbf{Y}_{1:k}. \tag{17}$$

Proof. Due to space limitations, we only briefly outline the main proof idea here for reference and provide the detailed proof in Appendix A. Specifically, the proof proceeds by mathematical induction. For the base case, we first verify that the initialization $\mathbf{F}_1$ satisfies the closed-form expression in (17). For the inductive step, assuming that an arbitrary $\mathbf{F}_{k-1}$ satisfies (17), we substitute its closed-form expression into the recursive update in (15) and apply the *Woodbury Matrix Identity* to show that the resulting $\mathbf{F}_k$ also satisfies (17). Therefore, by the principle of mathematical induction, every *Fused Knowledge Matrix* $\mathbf{F}_k$ obtained by our AnaPFL admits (17).

Theorem 1: The primary stream $\hat{\mathbf{G}}$ derived from our AnaPFL is exactly equivalent to the optimal solution that is obtained through empirical risk minimization centrally over the complete dataset $D_{1:K}$, as defined by the objective (2).

Proof. Due to space limitations, we only briefly outline the main proof idea here for reference and provide the detailed proof in Appendix A. We first derive the optimal solution to the centralized objective in (2) using the least-squares

method. Then, by substituting the expression of $\mathbf{F}_K$ established in Lemma 1 into (10), we show that the resulting $\hat{\mathbf{G}}$ is identical to the optimal solution.

Theorem 2: The refinement streams $\{\hat{\mathbf{P}}_k\}_{k=1}^K$ derived from our AnaPFL are exactly equivalent to the optimal solutions obtained through empirical risk minimization over the local datasets $\{D_k\}_{k=1}^K$, as defined by the objective (1).

Proof. Due to space limitations, we only briefly outline the main proof idea here for reference and provide the detailed proof in Appendix A. Based on Theorem 1, we treat $\hat{\mathbf{G}}$ as fixed and decompose the objective in (1) into independent client-wise subproblems. For each subproblem, we derive its closed-form optimal solution using the least-squares method and show that it is identical to the refinement stream $\hat{\mathbf{P}}_k$ computed by our AnaPFL.

Theorem 3 (Invariance to Heterogeneity): The primary-stream global model $\hat{\mathbf{G}}$ within our AnaPFL is independent of the data distributions across clients, as long as the complete dataset $D_{1:K}$ remains unchanged. Formally, given two arbitrary PFL systems with distinct data distributions $\{D_i\}_{i=1}^K$ and $\{D_i'\}_{i=1}^K$. As long as $\bigcup_{i=1}^K D_i = \bigcup_{i=1}^K D_i'$, their global models $\hat{\mathbf{G}}$ and $\hat{\mathbf{G}}'$ derived from our AnaPFL will be identical.

Proof. Due to space limitations, we only briefly outline the main proof idea here for reference and provide the detailed proof in Appendix B. According to the analytical expression of the primary stream proved in Theorem 1, $\hat{\mathbf{G}}$ depends only on $\mathbf{\Phi}_{1:K}^\top \mathbf{\Phi}_{1:K}$ and $\mathbf{\Phi}_{1:K}^\top \mathbf{Y}_{1:K}$. When the complete dataset remains unchanged, different client data distributions correspond only to permutations of the samples. Since both terms are invariant to such permutations, the resulting primary stream $\hat{\mathbf{G}}$ remains unchanged.

Based on Theorem 3, we can further deduce our AnaPFL's *Dual Invariance to Heterogeneity* for any client k to the data distributions of all other clients, as shown in Corollary 1 from Appendix B. Notably, in scenarios with partial client participation (e.g., with dropouts), the complete dataset $D_{1:K}$ in Theorem 3 and Corollary 1 should be defined as the union of data from all the participants.

4 Experiments

4.1 Experimental Setup

Datasets & Settings. We comprehensively evaluate AnaPFL on three visual benchmark datasets: CIFAR-100 [14], ImageNet-R [12], and Tiny-ImageNet [16]. To simulate diverse Non-IID scenarios in PFL, we partition the data among 50 and 100 clients via the Dirichlet distribution [21] with the concentration parameters $\alpha \in \{0.1, 0.5, 1.0\}$. Notably, a larger number of clients (e.g., 100 clients) and a smaller α (e.g., $\alpha = 0.1$) indicate more severe data heterogeneity, as a more challenging Non-IID scenario.

Baselines & Metrics. We compare our AnaPFL against up to *twelve* state-of-the-art baselines, including FL and PFL ones. The FL baselines include FedAvg [25], FedProx [19], their personalized variants (abbreviated as "-FT"), and the analytic-learning-based approach AFL [11]. The PFL ones include Ditto [18],

FedALA [43], FedDBE [42], Per-FedAvg [2], FedPCL [33], FedSelect [31], and FedAS [39]. All gradient-based methods are set with 3 local training epochs and 200 global communication rounds. To ensure a fair comparison, we follow the frozen-feature setting adopted by AFL [11], where all methods use the same frozen backbone for feature extraction. Building upon this, to address potential data privacy concerns associated with the original supervised backbone used in AFL (i.e., a pre-trained ResNet-18), we adopted a self-supervised ViT-MAE-Base [10] as the pre-trained backbone and utilized its CLS tokens for feature extraction. Following the PFL convention, we use the average accuracy of each client's final model on its local test set as the core metric. Moreover, we use the overall running time and communication overhead to evaluate the computational and communication efficiency of different methods.

Implementation Details. For our AnaPFL, the balance parameter λ is selected by randomly holding out 10% of the training data as a validation set, without accessing the test set. The selected values of λ are 0.5, 0.3, and 0.2 for CIFAR-100, Tiny-ImageNet, and ImageNet-R, respectively. Once λ is determined, the validation samples are restored, and the final model is trained using the complete training set. The remaining parameters are empirically set to $\gamma = 10^{-2}$, $\beta = 1$, $d_{\mathrm{P}} = d_{\mathrm{R}} = 1024$, and $\sigma_{\mathrm{P}}(\cdot) = \sigma_{\mathrm{R}}(\cdot) = \mathrm{Tanh}$, without exhaustive tuning for their optimal values. For the baselines, AFL and FedSelect are implemented using their official codebases, while the remaining baselines are implemented based on PFLLib. For each baseline, we evaluate several recommended configurations provided by its official repository or reference implementation and select the best-performing configuration using the same validation protocol described above. All experiments are implemented in PyTorch and conducted on NVIDIA RTX 4090 GPUs.

4.2 Overall Comparisons

As detailed in Table 1, we compared the average accuracy performance of our AnaPFL with the other 12 baselines across various settings in terms of datasets, client scale, and data heterogeneity. In general, our AnaPFL consistently achieves the best performance across all the experimental settings, outperforming the SOTA baselines with significant improvements of **at least 1.57%–16.71%**.

Notably, since AFL focuses on improving the generalization ability of the global model without accounting for personalization, it shows the most stable performance across all heterogeneity scenarios. In contrast, the classic FL methods (i.e., FedAvg and FedProx) exhibit an increase in performance as α increases, as the damage to their aggregation process from Non-IID data is reduced. Conversely, most PFL methods achieve peak performance at $\alpha = 0.1$, with a decline in performance as α increases, suggesting that personalization benefits are more pronounced when data heterogeneity is high (i.e., $\alpha = 0.1$). When data heterogeneity is low (i.e., $\alpha = 1.0$), the local data distribution across all clients tends to be IID, leading to minimal gain from the personalization process.

As a result, AFL is the strongest baseline with relatively low data heterogeneity (i.e., $\alpha = 0.5$ and $\alpha = 1.0$), benefiting from its strong ability to aggre-

Table 1: Performance comparisons of the average accuracy (%) among our AnaPFL and baselines. The best results are highlighted in **bold**, and the second-best ones are underlined. The "**Advance ↑**" refers to the absolute performance advantage of our AnaPFL compared with the second-best result. The comparisons involve three benchmark datasets, two different client scales, and three data heterogeneity settings.

Baseline	CIFAR-100						Tiny-ImageNet						ImageNet-R					
	50 Clients			100 Clients			50 Clients			100 Clients			50 Clients			100 Clients		
	$\alpha=0.1$	$\alpha=0.5$	$\alpha=1.0$	$\alpha=0.1$	$\alpha=0.5$	$\alpha=1.0$	$\alpha=0.1$	$\alpha=0.5$	$\alpha=1.0$	$\alpha=0.1$	$\alpha=0.5$	$\alpha=1.0$	$\alpha=0.1$	$\alpha=0.5$	$\alpha=1.0$	$\alpha=0.1$	$\alpha=0.5$	$\alpha=1.0$
FedAvg	43.53%	47.75%	48.91%	39.28%	43.31%	43.28%	40.86%	44.08%	43.99%	36.39%	39.90%	40.20%	6.30%	7.14%	7.61%	4.90%	4.60%	5.00%
FedProx	42.74%	46.73%	48.19%	38.37%	42.38%	42.54%	39.37%	42.77%	42.62%	34.80%	38.38%	38.74%	6.28%	7.36%	7.21%	4.85%	4.50%	4.99%
FedAvg-FT	58.81%	41.38%	37.73%	49.05%	31.37%	26.91%	49.61%	34.21%	30.15%	41.56%	25.28%	21.84%	20.42%	8.14%	6.15%	18.68%	7.52%	4.84%
FedProx-FT	57.63%	41.32%	37.73%	48.64%	31.38%	26.91%	49.59%	34.25%	30.14%	41.59%	25.28%	21.84%	20.40%	8.11%	6.13%	18.70%	7.53%	4.82%
Per-FedAvg	49.21%	40.39%	39.83%	49.65%	37.00%	34.26%	46.28%	36.10%	32.04%	46.10%	33.00%	27.63%	14.49%	7.57%	6.33%	13.83%	5.37%	4.34%
Ditto	69.80%	50.18%	43.72%	66.80%	41.32%	33.46%	62.10%	45.28%	44.99%	59.08%	41.26%	42.05%	25.35%	8.57%	6.05%	22.85%	7.88%	5.10%
FedALA	72.89%	52.60%	46.42%	70.63%	47.11%	38.20%	63.71%	43.25%	37.16%	61.16%	37.65%	30.88%	32.01%	12.29%	8.86%	28.20%	10.15%	5.91%
FedDBE	48.52%	49.96%	50.11%	43.20%	44.58%	44.41%	44.14%	45.24%	44.80%	39.28%	40.72%	40.86%	9.20%	8.46%	8.31%	7.40%	5.61%	5.78%
FedAS	56.70%	41.37%	37.78%	47.79%	30.73%	27.28%	48.64%	33.33%	29.78%	40.19%	24.12%	21.33%	20.32%	7.96%	6.00%	18.33%	7.37%	4.65%
FedPCL	23.99%	11.33%	12.52%	25.83%	11.82%	11.13%	14.63%	6.24%	7.58%	15.64%	6.88%	6.80%	0.83%	0.88%	1.13%	1.72%	1.11%	1.23%
FedSelect	72.46%	49.44%	41.58%	68.44%	42.71%	34.41%	62.29%	40.17%	33.53%	58.75%	33.50%	26.26%	34.83%	14.94%	9.59%	28.11%	11.01%	7.09%
AFL	56.63%	56.78%	56.66%	57.11%	57.23%	57.25%	46.63%	46.69%	46.25%	46.85%	46.87%	46.99%	28.87%	28.18%	28.32%	28.19%	28.58%	28.93%
AnaPFL	**80.64%**	**66.96%**	**62.63%**	**79.22%**	**65.70%**	**61.42%**	**69.01%**	**55.16%**	**51.39%**	**68.57%**	**54.47%**	**50.81%**	**46.23%**	**34.02%**	**31.03%**	**44.91%**	**31.20%**	**30.50%**
Advance ↑	**7.75%**	**10.18%**	**5.97%**	**8.59%**	**8.47%**	**4.17%**	**5.30%**	**8.47%**	**5.14%**	**7.41%**	**7.60%**	**3.82%**	**11.40%**	**5.84%**	**2.71%**	**16.71%**	**2.62%**	**1.57%**

gate client knowledge for global generalization. In highly Non-IID scenarios (i.e., $\alpha = 0.1$), FedALA and FedSelect emerge as the top performers among baselines, as the increasing heterogeneity among clients' local data makes personalization more effective. Distinct from all these baselines, our AnaPFL maintains the best performance as it properly balances generalization and personalization while mitigating the detrimental impact of Non-IID data in a gradient-free manner.

Concurrently, as the number of clients increases (i.e., from 50 to 100 clients), the performance of all methods generally declines. This evidence is because a larger number of clients not only detrimentally affects model aggregation but also results in each client's local dataset becoming sparser, making effective personalization more difficult. Despite our extensive tuning efforts, the average accuracy of most gradient-based baselines on the challenging dataset ImageNet-R still languish below 10%, highlighting their inherent limitations.

4.3 Efficiency Evaluations

Subsequently, we further conduct comprehensive efficiency comparisons between our AnaPFL and other baselines. Here, we set α to 0.1, the number of clients to 50, and $d_{\mathrm{P}} = d_{\mathrm{R}}$ to 1024 on all evaluations. Specifically, Figure 4 evaluates the Accuracy-Efficiency balance of all methods, while Figure 5 focuses on the efficiency advantages of our AnaPFL over the gradient-based baselines. Notably, the computation overhead for each method encompasses the running time of both feature extraction and classifier training for fairness.

We first compare the performance of AnaPFL and the baselines under different numbers of aggregation rounds, as shown in Figure 4(a). Since AnaPFL and AFL each require only a single aggregation round per client, they are represented by the red and blue dashed lines, respectively. In contrast, the gradient-based baselines iteratively update their models over as many as 200 aggregation rounds, with their average accuracies generally increasing and gradually converging as

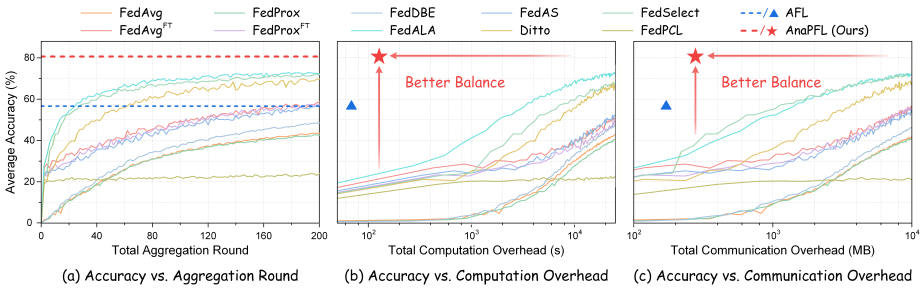

Fig. 4: Evaluations for the Accuracy-Efficiency balance on the CIFAR-100 dataset.

the number of rounds grows. As can be observed, our AnaPFL achieves superior average accuracy through gradient-free analytic learning after only one aggregation round, consistently outperforming all gradient-based methods even with up to 200 rounds of iterative aggregation.

Furthermore, Figures 4(b)–(c) jointly evaluate the trade-off between accuracy and computational or communication overhead. For clearer visualization of the substantial differences in overhead across methods, we use logarithmic scales for the horizontal axes in both plots. As can be observed, AnaPFL, denoted by the red pentagram ★, is consistently positioned in the **upper-left region** of both plots, indicating higher accuracy with lower running time and communication overhead. These results demonstrate that our AnaPFL not only achieves superior predictive performance, but also provides a favorable Accuracy–Efficiency balance in terms of both computation and communication.

Although AFL is slightly more efficient than our AnaPFL because it learns only a global model without performing client-specific personalization, its average accuracy is substantially lower than that of AnaPFL, with performance gaps of up to 17.36%–24.01% across the three datasets. Moreover, AFL remains less accurate than several gradient-based methods, indicating that efficiency alone is insufficient without an effective personalization mechanism. In contrast, AnaPFL combines the single-round efficiency with client-specific refinement, making it the only evaluated method that consistently outperforms all gradient-based baselines in both accuracy and efficiency across the considered PFL settings.

Then, we further quantify the efficiency advantage of our AnaPFL over the gradient-based methods in Figure 5. To highlight the superiority of our AnaPFL, we selected two baselines as representative examples for comparison: FedAvg for the FL category and FedPCL for the PFL category. As indicated in Figure 5, our AnaPFL achieves an overhead reduction of over **99%** compared to the gradient-based baselines in terms of both computation (e.g., 100–300 s vs. 45,000–66,000 s) and communication (e.g., 200–400 MB vs. 20,000–40,000 MB). Moreover, our AnaPFL has further room for future improvement via compression or sketching techniques (e.g., low-rank matrix decomposition) [38, 44].

Notably, these efficiency gains are measured under the controlled frozen-backbone protocol adopted in our experiments, where all methods use the same

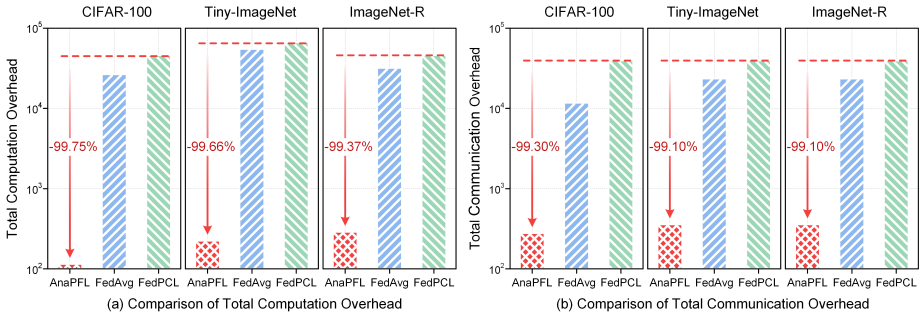

Fig. 5: Efficiency comparisons regarding computation and communication overhead.

pretrained backbone for feature extraction. This unified setting enables a more controlled and fair evaluation of the Accuracy–Efficiency trade-offs among different methods. By contrast, if the backbone were unfrozen, the overhead of the gradient-based methods would increase substantially, since the backbone would need to be optimized over multiple local epochs and communicated across multiple global rounds. For instance, each communication round would require uploading and downloading the updated backbone parameters, increasing the communication overhead by orders of magnitude. This observation further underscores the substantial efficiency advantage of AnaPFL.

5 Conclusion and Discussion

In this paper, we propose our AnaPFL to address the Non-IID issue while balancing generalization and personalization in a gradient-free manner. Specifically, our AnaPFL brings AL into PFL, and employs dual-stream analytic models to achieve both collective generalization and individual personalization. Notably, our AnaPFL mainly aims at efficient PFL over pretrained frozen features, based on AFL, rather than replacing all gradient-based methods with trainable representations. Extensive experiments yield impressive results, validating the superior performance of our AnaPFL in terms of accuracy and efficiency.

In our AnaPFL, we adopt a pretrained ViT-MAE as the backbone for feature extraction, following the established tradition of employing a backbone in analytic learning (e.g., ResNet-18 in AFL [11]). In this way, we also address the previous privacy concerns of AFL regarding data leakage, as the self-supervised pre-training for ViT-MAE only necessitates image reconstruction and avoids the use of image labels [10]. Moreover, given the widespread adoption of ViT-based backbones in the FL community, the resulting additional storage overhead (< 0.5 GB) is typically acceptable. Meanwhile, our AnaPFL is also readily compatible with other stronger self-supervised models, such as DINO series [27], for further enhancement with only minimal modifications.

Acknowledgment

This work is supported partly by the National Natural Science Foundation of China (NSFC) under grants 62576013 and U24A20248, the National Key Research and Development of China under grant 2024YFC2607404, and the Ningxia Domain-Specific Large Model Health Industry R&D Program 2024JBGS001.

References

1. Collins, L., Hassani, H., Mokhtari, A., Shakkottai, S.: Exploiting shared representations for personalized federated learning. In: Meila, M., Zhang, T. (eds.) Proceedings of the 38th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 139, pp. 2089–2099. PMLR (18–24 Jul 2021) 4
2. Fallah, A., Mokhtari, A., Ozdaglar, A.: Personalized federated learning: A meta-learning approach. arXiv preprint arXiv:2002.07948 (2020) 2, 4, 11
3. Fan, K., Guo, J., Li, R., Li, Y., Liu, A., Tang, J., Wang, T., Dong, M., Song, H.: RMDF-CV: A reliable multi-source data fusion scheme with cross validation for quality service construction in mobile crowd sensing. IEEE Transactions on Services Computing **18**(1), 399–413 (2025) 4
4. Fan, K., Huang, Y., He, J., Han, F., Tang, J., Zhuang, H., et al.: CALM: A ubiquitous crowdsourced analytic learning mechanism for continual service construction with data privacy preservation. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **9**(2) (2025) 3, 4, 5
5. Fan, K., Liu, J., Tang, J., Liu, A., Wang, T., Liu, Y., Dong, M., Song, H.: PCQ-SP: A privacy-preserving, cost-aware, and quality-enabled service planning scheme for multi-task spatial crowdsourcing. IEEE Transactions on Mobile Computing **25**(5), 6629–6643 (2026) 4
6. Fani, E., Camoriano, R., Caputo, B., Ciccone, M.: Accelerating heterogeneous federated learning with closed-form classifiers. In: International Conference on Machine Learning. pp. 13029–13048. PMLR (2024) 5
7. Ghosh, A., Kumar, H., Sastry, P.S.: Robust loss functions under label noise for deep neural networks. In: Proceedings of the AAAI conference on artificial intelligence. vol. 31 (2017) 6
8. Guo, P., Lyu, M.R., Mastorakis, N.: Pseudoinverse learning algorithm for feedforward neural networks. Advances in Neural Networks and Applications **1**(321-326) (2001) 5
9. Han, P., Wang, S., Jiao, Y., Huang, J.: Federated learning while providing model as a service: Joint training and inference optimization. In: IEEE INFOCOM 2024 - IEEE Conference on Computer Communications. pp. 631–640 (2024) 1, 2, 4
10. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15979–15988 (2022) 5, 11, 14
11. He, R., Tong, K., Fang, D., Sun, H., Zeng, Z., Li, H., Chen, T., Zhuang, H.: AFL: A single-round analytic approach for federated learning with pre-trained models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2025) 2, 3, 4, 5, 6, 10, 11, 14
12. Hendrycks, D., Basart, S., Mu, N., Kadavath, S., Wang, F., Dorundo, E., Desai, R., Zhu, T., Parajuli, S., Guo, M., et al.: The many faces of robustness: A critical analysis of out-of-distribution generalization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8340–8349 (2021) 10

13. Huang, Y., Guo, J., Yang, S., Liu, J., Liu, A., Tang, J., Wang, T., Dong, M., Song, H.: QLP-DCS: A quality-aware, low-cost, and privacy-preserving data collection service for mobile crowd sensing. IEEE Transactions on Services Computing **18**(4), 2326–2341 (2025) 4
14. Krizhevsky, A., Hinton, G.: Learning multiple layers of features from tiny images. Technical Report (2009) 10
15. Lai, S., Liao, M., Hu, Z., Yang, J., Chen, W., Xiao, H., Tang, J., Liao, H., Yue, Y.: Learning new concepts, remembering the old: Continual learning for multimodal concept bottleneck models. In: Proceedings of the 33rd ACM International Conference on Multimedia. pp. 12314–12322. MM '25 (2025) 5
16. Le, Y., Yang, X.: Tiny imagenet visual recognition challenge. CS 231N **7**(7), 3 (2015) 10
17. Li, J., Li, R., Qi, J., Lai, S., Lv, L., Fan, K., Tang, J., Yue, Y., Zhou, D., Liu, Y., Zhuang, H.: CFSSeg: Closed-form solution for class-incremental semantic segmentation of 2d images and 3d point clouds. In: Proceedings of the 33rd ACM International Conference on Multimedia. p. 247–256. MM '25 (2025) 5, 6
18. Li, T., Hu, S., Beirami, A., Smith, V.: Ditto: Fair and robust federated learning through personalization. In: International Conference on Machine Learning. pp. 6357–6368. PMLR (2021) 10
19. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. Proceedings of Machine Learning and Systems **2**, 429–450 (2020) 10
20. Liang, P.P., Liu, T., Ziyin, L., Allen, N.B., Auerbach, R.P., Brent, D., Salakhutdinov, R., Morency, L.P.: Think locally, act globally: Federated learning with local and global representations. arXiv preprint arXiv:2001.01523 (2020) 4
21. Lin, T., Kong, L., Stich, S.U., Jaggi, M.: Ensemble distillation for robust model fusion in federated learning. In: Advances in Neural Information Processing Systems. vol. 33, pp. 2351–2363. Curran Associates, Inc. (2020) 10
22. Liu, Y., Chang, S., Liu, Y., Li, B., Wang, C.: FairFed: Improving fairness and efficiency of contribution evaluation in federated learning via cooperative shapley value. In: IEEE INFOCOM 2024 - IEEE Conference on Computer Communications. pp. 621–630 (2024) 1
23. Liu, Z., Du, C., Lee, W.S., Lin, M.: Locality sensitive sparse encoding for learning world models online. In: The Twelfth International Conference on Learning Representations (2024) 5
24. Luo, J., Mendieta, M., Chen, C., Wu, S.: PGFed: Personalize each client's global objective for federated learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3946–3956 (2023) 2, 4
25. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial Intelligence and Statistics. pp. 1273–1282. PMLR (2017) 10
26. OH, J.H., Kim, S., Yun, S.: FedBABU: toward enhanced representation for federated image classification. In: 10th International Conference on Learning Representations, ICLR 2022. International Conference on Learning Representations (ICLR) (2022) 4
27. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al.: DINOv2: Learning robust visual features without supervision. Transactions on Machine Learning Research Journal pp. 1–31 (2024) 14

28. Ozkara, K., Girgis, A.M., Data, D., Diggavi, S.: A statistical framework for personalized federated learning and estimation: Theory, algorithms, and privacy. In: The Eleventh International Conference on Learning Representations (ICLR) (2023) 4
29. Park, J., Sandberg, I.W.: Universal approximation using radial-basis-function networks. Neural Computation **3**(2), 246–257 (1991) 5
30. Ren, C., Yu, H., Peng, H., Tang, X., Zhao, B., Yi, L., Tan, A.Z., Gao, Y., Li, A., Li, X., Li, Z., Yang, Q.: Advances and open challenges in federated foundation models. IEEE Communications Surveys & Tutorials pp. 1–41 (2025) 1
31. Tamirisa, R., Xie, C., Bao, W., Zhou, A., Arel, R., Shamsian, A.: FedSelect: Personalized federated learning with customized selection of parameters for fine-tuning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 23985–23994 (2024) 4, 11
32. Tan, A.Z., Yu, H., Cui, L., Yang, Q.: Towards personalized federated learning. IEEE Transactions on Neural Networks and Learning Systems **34**(12), 9587–9603 (2022) 2, 4
33. Tan, Y., Long, G., Ma, J., Liu, L., Zhou, T., Jiang, J.: Federated learning from pre-trained models: A contrastive learning approach. Advances in Neural Information Processing Systems **35**, 19332–19344 (2022) 11
34. Tang, J., Fan, K., Yang, S., Liu, A., Xiong, N.N., Song, H.H., Leung, V.C.M.: CPDZ: A credibility-aware and privacy-preserving data collection scheme with zero-trust in next-generation crowdsensing networks. IEEE Journal on Selected Areas in Communications **43**(6), 2183–2199 (2025) 4
35. Tang, J., He, J., Fan, K., He, R., Wang, J., Liu, A., Song, H.H., Wang, L., Zhu, Z., Zhuang, H., Liu, Y.: AFCL: Achieving spatio-temporal invariance to data heterogeneity in federated continual learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Findings. pp. 7768–7778 (2026) 5
36. Tang, J., Huang, Y., Fan, K., Han, F., Li, J., Xu, J., He, R., Liu, A., Song, H.H., Zhuang, H., Liu, Y.: DeepAFL: Deep analytic federated learning. In: The Fourteenth International Conference on Learning Representations (ICLR) (2026) 2, 3, 4, 5
37. Tang, J., Huang, Z., Yang, S., Fan, K., Liu, A., Wang, T., Liu, Y., Dong, M., Song, H.H.: Q-DBPP: A quality-aware dual-bilateral privacy preserving scheme in mobile crowd sensing. IEEE Transactions on Mobile Computing **25**(4), 4904–4919 (2026) 4
38. Xue, Y., Lau, V.: Riemannian low-rank model compression for federated learning with over-the-air aggregation. IEEE Transactions on Signal Processing **71**, 2172–2187 (2023) 13
39. Yang, X., Huang, W., Ye, M.: FedAS: Bridging inconsistency in personalized federated learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11986–11995 (2024) 2, 4, 11
40. Yang, Y., Long, G., Shen, T., Jiang, J., Blumenstein, M.: Dual-personalizing adapter for federated foundation models. In: Globerson, A., Mackey, L., Belgrave, D., Fan, A., Paquet, U., Tomczak, J., Zhang, C. (eds.) Advances in Neural Information Processing Systems. vol. 37, pp. 39409–39433. Curran Associates, Inc. (2024) 5
41. Yu, H., Yang, X., Gao, X., Kang, Y., Wang, H., Zhang, J., Li, T.: Personalized federated continual learning via multi-granularity prompt. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. p. 4023–4034. KDD '24, Association for Computing Machinery, New York, NY, USA (2024) 5

42. Zhang, J., Hua, Y., Cao, J., Wang, H., Song, T., XUE, Z., Ma, R., Guan, H.: Eliminating domain bias for federated learning in representation space. Advances in Neural Information Processing Systems **36**, 14204–14227 (2023) 11
43. Zhang, J., Hua, Y., Wang, H., Song, T., Xue, Z., Ma, R., Guan, H.: FedALA: Adaptive local aggregation for personalized federated learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 11237–11244 (2023) 11
44. Zhang, P., Xu, L., Mei, L., Xu, C.: Sketch-based adaptive communication optimization in federated learning. IEEE Transactions on Computers (2024) 13
45. Zhang, R., Chen, Y., Wu, C., Wang, F., Li, B.: Multi-level personalized federated learning on heterogeneous and long-tailed data. IEEE Transactions on Mobile Computing **23**(12), 12396–12409 (2024) 2, 4
46. Zhang, Y., Zhu, H., Tan, A.Z., Yu, D., Huang, L., Yu, H.: pFedMxF: Personalized federated class-incremental learning with mixture of frequency aggregation. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 30640–30650 (2025) 2, 4
47. Zhou, X., Yang, Q., Zheng, X., Liang, W., Wang, K.I.K., Ma, J., Pan, Y., Jin, Q.: Personalized federated learning with model-contrastive learning for multi-modal user modeling in human-centric metaverse. IEEE Journal on Selected Areas in Communications **42**(4), 817–831 (2024) 2, 4
48. Zhuang, H., Chen, Y., Fang, D., He, R., Tong, K., Wei, H., Zeng, Z., Chen, C.: GACL: Exemplar-free generalized analytic continual learning. Advances in Neural Information Processing Systems **37**, 83024–83047 (2024) 6
49. Zhuang, H., He, R., Tong, K., Zeng, Z., Chen, C., Lin, Z.: DS-AL: A dual-stream analytic learning for exemplar-free class-incremental learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 17237–17244 (2024) 5
50. Zhuang, H., Lin, Z., Toh, K.A.: Blockwise recursive moore–penrose inverse for network learning. IEEE Transactions on Systems, Man, and Cybernetics: Systems **52**(5), 3237–3250 (2021) 5
51. Zhuang, H., Weng, Z., He, R., Lin, Z., Zeng, Z.: GKEAL: Gaussian kernel embedded analytic learning for few-shot class incremental task. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7746–7755 (2023) 5
52. Zhuang, H., Weng, Z., Wei, H., Xie, R., Toh, K.A., Lin, Z.: ACIL: Analytic class-incremental learning with absolute memorization and privacy protection. Advances in Neural Information Processing Systems **35**, 11602–11614 (2022) 5