SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Fed-RCPA: Representative Class Prototype Aggregation-based One-Shot Federated Learning

Maria Altaib Badawi

Department of Computer Science and Information, College of Science Zulfi, Majmaah University, Al-Majmaah 11952, Saudi Arabia.
M.badawi@mu.edu.sa

ABSTRACT
Federated Learning enables numerous rounds for server-clients communication, leading to critical challenges such as network instability, and vulnerability to privacy attacks, and high communication overhead. One-shot FL (OFL) represents a Federated Learning paradigm that overcomes above drawbacks by a single round of server-clients communication. However, existing OFL faces expensive computation cost and fails to address non Independent and Identically Distributed (non-IID) data efficiently and robustly. To overcome these issues, this work introduces Fed-RCPA, a novel OFL that capture Representative Class Prototypes for aggregation. Using global feature statistics enables heterogeneity-resistant and training-free aggregation process. In addition, we ameliorate the prototypes quality using Dual Tree Complex Wavelet Transform (DTCWT) for efficient aggregation. Furthermore, we incorporate explainability using Grad-CAM tool for improving trust and the transparency. Fed-RCPA application to two different scenarios, using benchmarks and real medical case, demonstrates the method effectiveness across varying degrees of data heterogeneity settings. Experimental evaluation on CIFAR-10, CIFAR-100, and SVHN demonstrates that Fed-RCPA achieves accuracies of 65.55%, 41.08%, and 61.94%, respectively, while maintaining stable performance across the investigated non-IID heterogeneity settings. In addition, its application to chest X-ray images for pneumonia detection achieves 69.55% test accuracy with 100% sensitivity for pneumonia cases. These results demonstrate the robustness of Fed-RCPA to data heterogeneity and its applicability to both benchmark and real-world medical imaging scenarios.

Keywords- One-Shot Federated Learning; prototypes; non-IID data; pre-trained models; Explainable ML; Grad-CAM tool.

1. INTRODUCTION

Federated learning (FL) is a decentralized Machine Learning (ML) framework that allows multiple independent devices to jointly optimize a global model without exchanging their proprietary local datasets. This paradigm not only preserves data privacy by keeping raw information on the edge, but also facilitates collaborative learning across heterogeneous, cross-organizational domains [1][2][3]. One-shot Federated Learning (OFL) resolves the challenges associated with traditional FL, including slow convergence and substantial communication expenses, by training global model in just one round of communication. This significantly boosts efficiency, particularly in an environment having limited resources and privacy-sensitive [4][5]. However, the current OFL methods often face difficulties due to the heterogeneity of data across clients, leading to performance degradation, especially in real applications like medical imaging [6][7]. To address these issues, innovative approaches that combine feature-aware hierarchical token sequences with knowledge distillation have been introduced to improve the OFL effectiveness, particularly when dealing with non-Independent and Identically Distributed (non-IID) image data settings [8][9]. For example, FALCON utilized hierarchical token sequences along with knowledge distillation to enhance the OFL performance on non-IID image datasets [10]. Similarly, in [11], the novel framework named FedHydra resolves both data heterogeneity and model heterogeneity via a two-stage learning mechanism. It incorporates model stratification and heterogeneity-aware stratified aggregation [11]. This method introduces a dual-prototype strategy, combining specific and generalized prototypes, to capture class representations for clustering, alongside a prototype-contrastive training mechanism to maximize intra-cluster prototype consistency [12].

In addition, the diversity of local prototypes in FL that results from data heterogeneity leads to significant aggregation bias in conventional weighted averaging strategies [13]. This heterogeneity stems from the fact that a single prototype is often inadequate to represent a class when intra-class feature distributions vary significantly across clients [14]. Furthermore, handling the pervasive issue of non-IID data distributions among clients represents a critical challenge in FL. This issue causes performance degradation, especially where clients possess heterogeneous local data distributions [15]. As depicted by Guo et al. [16], both intensive cross-client migration and the development of unbiased global models can significantly increase computational overhead and training time. The major advantage of OFL approaches

lies in enabling optimal knowledge transfer and aggregation within a single communication round, therefore reducing privacy risks and improving overall system efficiency [17]. For instance, [5] presented FedCGS, a novel FL algorithm designed to capture global feature statistics through pre-trained models. By employing these statistics, the FedCGS algorithm achieves training-free operation while remaining robust to data heterogeneity in OFL settings.

The proposed Representative Class Prototype Aggregation-based OFL, Fed-RCPA, aims to ameliorate convergence within a single communication round under non-IID data settings. By employing selective layer prototypes, the Fed-RCPA approach aims to refine the abstraction of knowledge across different feature hierarchies. Thus, the model's ability to learn robust and discriminative representations is improved even under severe data imbalance and domain shift conditions. This paper further evaluates Fed-RCPA's efficacy using CIFAR-10, CIFAR-100, SVHN, and real medical datasets, demonstrating superior performance in image classification tasks under various heterogeneous conditions. This extensive evaluation confirms the robustness and adaptability of Fed-RCPA, showcasing its potential to overcome persistent challenges in data and model heterogeneity inherent in OFL.

The remainder of this paper is organized as follows. Section 2 provides background on One-Shot Federated Learning (OFL). Section 3 details our methodology and introduces the proposed Fed-RCPA framework that combines the weighted average method for global feature identification and the Dual Tree Complex Wavelet Transform (DTCWT) for ameliorating the representation quality. Section 4 presents the experimental setup and findings. Section 5 describes the evaluation study for comparing Fed-RCPA with existing works. Finally, Section 5 concludes the paper.

2. Literature Review

One-Shot Federated Learning (OFL) is an advanced distributed ML paradigm that enables multiple decentralized clients to cooperatively learn a global model using strictly a single communication round [18]. While OFL promises efficient training through single-round collaboration, existing methods often fall short in accuracy compared to multi-round counterparts due to the strict requirements of single-round knowledge transfer [19]. Despite the utilisation of pre-trained models and statistical aggregation to surpass representational gaps, OFL approaches face challenges in heterogeneous federated settings,paticularly with diverse data distributions [20]. The persistent challenges of data and model heterogeneity in OFL remain a significant obstacle. This necessitates more sophisticated aggregation strategies that can successfully reconcile the diverse learning outputs from different clients within a single communication round [21][22]. Addressing this, certain approaches have demonstrated that increasing model diversity plays a key role in reducing the effects of non-IID data, leading to a more representative global model [23]. However, these methods often introduce additional complexity or computational overhead [24]. Accordingly, to maintain effeciency, OFL requires new solutions that effectively address model and data heterogeneity [25][26]. Due to the lack of iterative communication rounds for model reconciliation, proposing novel OFL approaches that minimize this bias are crucial [27].

Several existing OFL methods aim to overcome these limitations by focusing on efficient knowledge distillation and model aggregation. Employing dataset distillation or generating synthetic data on the server side enable improving the performance of global model [28][29]. Several methods employ knowledge distillation techniques for transferring learned representations from multiple client models to a global server model in a single round [18][30][31]. Despite these advancements, preserving privacy and mitigating the risks of data reconstruction attacks remain paramount concerns [32]. In real-world FL settings, client data is rarely IID. As a result, local models tend to over-optimize for their specific datasets, which ultimately degrades the generalizability of the final global model [33]. This issue is specialy acute in OFL, where the lack of iterative aggregation prevents the server from correcting local biases, ultimately amplifying the negative effects of data heterogeneity on the global model [34][35].

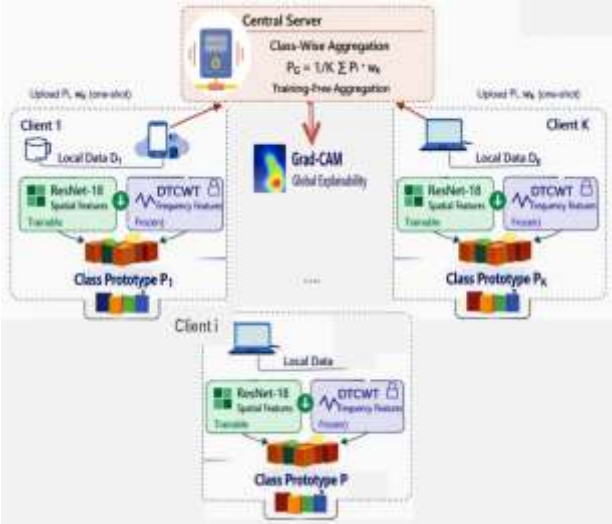
The widespread adoption of pre-trained models like ResNet [36] and ViT [37] has fundamentally improved the performance of downstream tasks. As a result, adapting these models for FL has become a key focus of recent research, providing a promising approach to address the challenges of training from scratch in decentralized settings [38][39]. Several OFL-based pre-trained models have been introduced in the literature. For instance, Beitollahi et al. (2024) developed FedPFT, which exploits the transferable knowledge of pre-trained models to improve both the generalization performance and the computational efficiency of OFL [19]. This FedPFT achieves competitive performance with minimal computational overhead while remaining robust to data heterogeneity. More recently, in [5], authors introduced FedCGS, a new FL algorithm designed to capture global feature statistics using pre-trained models. By utilizing these statistics, their approach enables training-free operation while maintaining robustness to data heterogeneity in OFL scenarios. In 2026, the Co-Boosting++ framework improves the global model through a combined optimization of synthetic data quality and ensemble aggregation. By regarding these two

objectives as mutually dependent, it offers a principled, unified solution to the interconnected data and model heterogeneity challenges typical of OFL.
Locally trained generators pose a significant privacy risk, as their direct exposure to client data can lead to the memorization and inadvertent leakage of sensitive information [40]. OFL methodologies elegantly resolve this by training generators without ever accessing raw data, thereby upholding the strict privacy mandates of modern decentralized systems.
Unlike existing OFL approaches, Fed-RCPA explicitly constructs class-specific prototypes from a hybrid representation combining high-level semantic features extracted by ResNet-18 with multi-scale and directional information captured by DTCWT as explained in the following section.

3. Methodology

This section outlines the specific methodologies employed to develop the proposed Fed-RCPA framework, focusing on selective feature representation and class-wise weighted aggregation (described in Algorithm 1). As depicted in Fig.1, it details the architectural design including different stages designed to address the challenges of non-IID data distribution, data heterogeneity, and communication efficiency in a OFL paradigm. In addition, we use Grad-CAM tool for improving trust and the transparency.
Specifically, we delineate the novel method to prototype generation at feature representation, designed to capture essential feature representations while minimizing data leakage and computational overhead. As shown in Fig.1, this novel approach is named DTCWT-enhanced feature representation. The Dual-Tree Complex Wavelet Transform (DTCWT) is used to improve feature representation by capturing multi-scale and directional information while offering better shift stability compared to conventional discrete wavelet transforms [41].

Fig. 1. Overview of Fed-RCPA framework.



In the proposed framework, DTCWT serves as a complementary feature extraction method alongside the ResNet-18 backbone. It breaks down the input images into multiple frequency sub-bands at various scales and orientations, allowing for the extraction of both spatially localized structural patterns and fine-grained high-frequency details.
Specifically, low-frequency components capture coarse structural and contextual details, while high-frequency components retain local variations, edges, textures, and other fine-grained features that can be important for differentiating between classes. At the fusion phase, the resulting wavelet coefficients offer complementary information to the semantic representations obtained from the convolutional backbone. For this reason, we employ colored cubes to mention the integration of various representations with Dimension D_T.This dimension represents the sum of D_{DTCWT} and $D_{ResNet\text{-}18}$ introspectively of DTCWT output and ResNet-18 Layer 4 feature.
The fused representation therefore contains both semantic information and complementary frequency/spatial information. Class-specific prototypes computed from this enriched representation provide a more comprehensive summary of the local class distribution without requiring transmission of raw images or individual feature vectors.
This DTCWT-enhanced representation is especially appropriate for the OFL scenario, in which each client condenses its local data into class-specific prototypes instead of sending full model updates or individual

feature representations. This framework also integrates a sophisticated aggregation mechanism that prioritizes contributions from clients exhibiting higher data diversity and model accuracy.

To generate global prototypes, we employ the class-wise weighted aggregation of client prototypes based on the number of local samples per class. For each class, a client's contribution is weighted according to the number of samples it has for that class. The server aggregates these local prototypes by using the class-specific sample counts as weights. Consequently, clients with a larger number of samples for a given class have a proportionally greater influence on the corresponding global prototype. This class-wise weighted aggregation generates a representative global prototype while addressing the uneven class distributions across clients in the non-IID setting.

To improve the interpretability of the proposed Fed-RCPA framework, Grad-CAM (Gradient-weighted Class Activation Mapping) is utilized as a post-aggregation explainability method [42]. After the server-side class-wise weighted aggregation of local prototypes, the resulting global representations are incorporated into the classification model to generate predictions for representative test samples. Grad-CAM leverages the gradients of the target class with respect to the final convolutional feature maps to produce a class-discriminative localization map, highlighting the regions of the image that contribute most significantly to the prediction [42]. The resulting activation maps are up-sampled and overlaid on the original images to provide an intuitive visual explanation of the model's decision. It is important to emphasize that Grad-CAM is used exclusively as a post-prediction explainability mechanism. After server-side classification, it is applied to representative test images to visualize the image regions that most strongly contribute to the predicted class. It thereby provides an interpretable assessment of the decisions produced by the aggregated prototype-based classifier.

This analysis allows for the investigation of whether the aggregated representations retain meaningful and class-relevant visual patterns despite the heterogeneous and non-IID distributions across clients.

Algorithm 1: Procedure of Fed-RCPA

1. for client $i = 1$ to M do (in parallel)
2. $\{P_i^j, N_i^j\}_{j=1}^{C} \leftarrow \text{ClientPrototypes}(D_i)$
3. end for
4. for each class $j \in C$ do
5. $N^j \leftarrow \sum_{i=1}^{M} N_i^j$
6. $P^j \leftarrow \frac{\sum_{i=1}^{M} N_i^j P_i^j}{N^j}$
7. end for
8. $\mathcal{P} \leftarrow \{P^j\}_{j=1}^{C}$
9. return $\mathcal{P}$

ClientPrototypes(D_i):

1. for each $(x, y) \in D_i$ do
2. $r(x) \leftarrow \text{ResNet18-Layer4}(x)$
3. $w(x) \leftarrow \text{DTCWT}(x)$
4. $f(x) \leftarrow \text{Fusion}(r(x), w(x))$
5. end for
6. for each class $j \in C$ do
7. $D_i^j \leftarrow \{x \in D_i \mid y = j\}$
8. $N_i^j \leftarrow |D_i^j|$
9. $P_i^j \leftarrow \frac{1}{N_i^j} \sum_{x \in D_i^j} f(x)$
10. end for
11. return $\{(P_i^j, N_i^j)\}_{j=1}^{C}$

As mathematical interpretation, we can mention that the most important part of our algorithm is represented in the following equation:

$$P_i^j = \frac{1}{N_i^j} \sum_{x \in D_i^j} f(x) \quad (1)$$

i denotes the client index; j denotes the class index; C denotes the total number of classes; and P denotes prototype and the fused representation is:

$$f(x) = Fusion(f_{ResNet18-L4}(x), f_{DTCWT}(x)) \quad (2)$$

Then, the server performs the class-wise weighted aggregation:

$$P^j = \frac{\sum_{i=1}^{M} N_i^j P_i^j}{\sum_{i=1}^{M} N_i^j} \quad (3)$$

Where N_i^j represents the number of samples of class j on client i. In our Fed-CPA, each client i extracts class-specific features from its own dataset D_i for each class $j \in \{1, \ldots, C\}$ and f denotes the feature extractor of the pre-trained model (ResNet18).

The server models each class as a Gaussian centered on its aggregated prototype, with a single shared covariance computed from the clients' summed statistics. Using Bayes' rule with the class prior (class count over total samples), the log-posterior simplifies to a linear score. Each test sample is then assigned to the class with the highest score representing the maximum a posteriori decision, which requires no training or gradients. Furthermore, our OFL respect the privacy constraint that characterizes the FL as indicated in line number 2 in Algorithm 1. The server will receive information represented by I to be used for the global feature statistics to perform global OFL. Accordingly, we can speak about secure aggregation because each client upload P and class-wise label count N for each class.

4. Experimental Results

This section describes the experimental methodology, including the datasets used, the specification of model hyper-parameters, and the presentation of empirical results. To test our proposal, we conduct experiments both on benchmark datasets and real medical dataset to demonstrate the performance of our OFL.

A. Experimental settings

For our experiments, we used three well-known image datasets for classification task, which are CIFAR100 [43], CIFAR10 [44], and SVHN [45]. These datasets are described in the following table:

TABLE I. BENCHMARK DATASETS DESCRIPTION

Dataset	**Description**		
	Total images	***Image size***	***Classes***
CIFAR-10	60,000 color images	32×32 pixels	10 classes
CIFAR-100	60,000 color images	32×32 pixels	100 classes
SVHN	600,000 digit images	32×32 pixels	10 digits

In this work the proposed Fed-RCPA algorithm is implemented in Google-Colab (GPU + TPU with 32 GB RAM) environment. We use ResNet18 that is pre-trained on ImageNet for all datasets. In the local training phase, for each client, we use mini-batch SGD as the local optimizer and set the batch size to 128. The momentum is set to 0.5, the learning rate is set to 0.01, the weight decay is set to 5e-4. The number of clients is set to 30 across all datasets.

As a first step of our implementation, we verified the generated output of ResNet18 Layer 3 and ResNet18 Layer 4 using real image and sample of CIFAR-10 images (see Fig. 2). Our aim is to determine which feature representation is more informative. In our case, it is important to select the most informative representation to be used by the server. Fig. 2, illustrates the difference between ResNet18-Layer 4 and Layer 3 feature representations using the same input image. The figure shows that Layer 3 creates a fairly detailed and spread-out activation pattern. Its feature map picks out several parts of the image, including bits of the animal and the areas around it. This shows that Layer 3 is kind of in the middle, keeping quite a bit of local and spatial info like textures, edges, and structural patterns. On the other hand, Layer 4 makes a more focused activation pattern, with stronger activity around the main semantic parts of the image, especially the animal's head and upper body. This means that the deeper Layer 4 picks up more high-level, class-relevant semantic info, while gradually losing some of the finer spatial details.

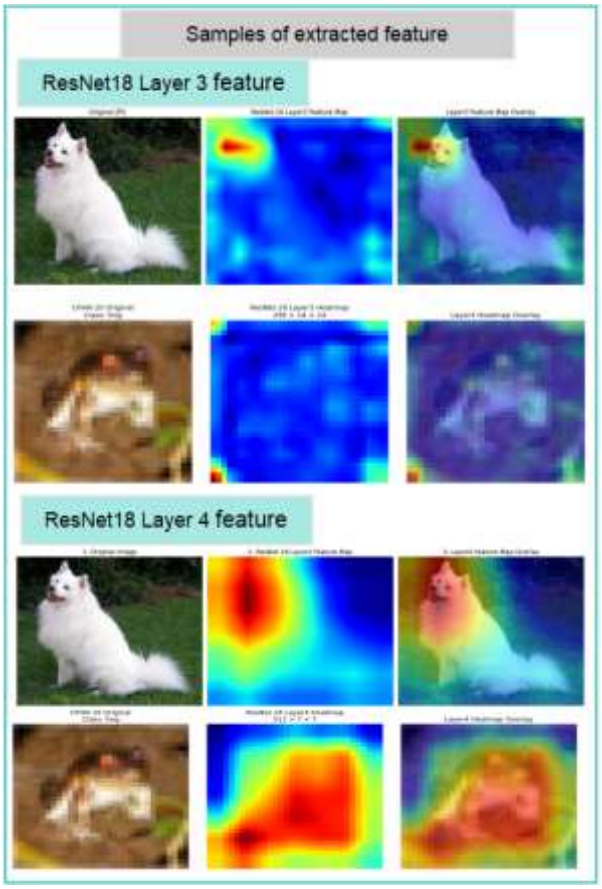


Fig. 2. ResNet18 Layer 3 vs Layer 4 feature representations.

For this reason, in the proposed framework, Layer 4 of the pre-trained ResNet-18 backbone was selected because it provides the deepest convolutional representation and therefore contains higher-level semantic and class-discriminative information than earlier layers

In this paper, we test the effectiveness of our OFL framework for diverse data heterogeneity stetting. Heterogeneity refers to the variations in data distributions among clients, i.e., non-IID data where clients might exhibit different class proportions or feature distributions. In our experiments, α (the Dirichlet parameter) determines the extent of this data heterogeneity, where a lower α signifies more heterogeneity.We therefore consider α = 0.03, 0.05, 0.1, 0.3, and 0.5 to cover a broad spectrum ranging from severe to relatively mild non-IID conditions and to systematically evaluate the robustness of Fed-RCPA to different degrees of client heterogeneity.

B. Experimental results using Benchmark datasets

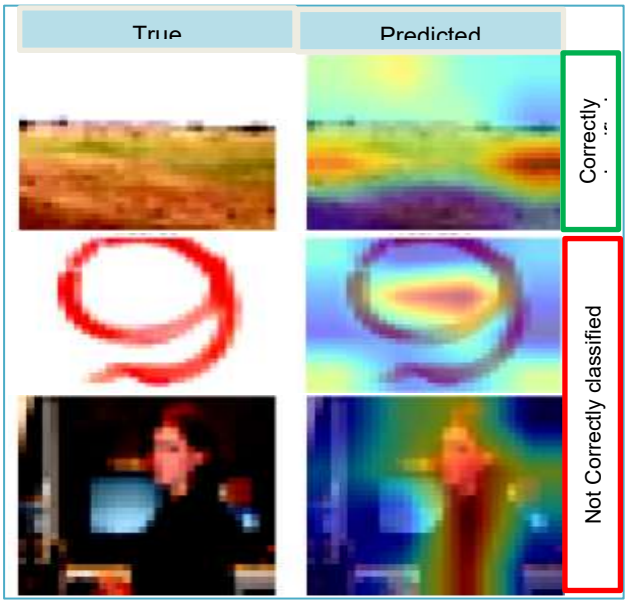
In our implementation, we use ResNet18 as classification networks for CIFAR-10, CIFAR-100 and SVHN. We employ accuracy as a metric for performance evaluation and comparison. Accuracy (ACC) measures the set of instances that are correctly identified by the pre-trained model ResNet18. Table II illustrates the experimental results for each dataset using ResNet-18 Layer 4.

TABLE II. EXPERIMENTAL RESULTS.

Dataset	**Accuracy**	
	α=0.03	***α=0.5***
CIFAR-10	65.55%	65.55%
CIFAR-100	41.08%	41.08%
SVHN	61.94%	61.94%

As reported in Table II, the experimental results across all benchmark datasets using ResNet-18 (Layer 4) with α values of 0.03 and 0.5, where a lower α indicates greater heterogeneity; accordingly, both the smallest and largest values were selected to illustrate the performance range. Table II shows the evaluation results

of the proposed OFL framework on three benchmark datasets, CIFAR-10, CIFAR-100, and SVHN, under two heterogeneity settings.



Grad-CAM visualizations for server-side aggregated prototype classifier using CIFAR-10 (α=0.1).

The results indicate that our Fed-RCPA maintains consistent accuracy across all datasets and heterogeneity levels. For this reason, the obtained outcomes demonstrate the robustness and stability of the learning framework, which will be further analyzed in evaluation section. Specifically, CIFAR-10 achieves 65.55 % accuracy, CIFAR-100 reaches 41.08 %, and SVHN attains 61.94 %, with no noticeable performance degradation as α increases. The visualizations in Fig. 3 depict the semantically meaningful zones highlighted by Grad-CAM for the test images. This result explain the importance of the aggregation method for the final classification in the server side. These examples encompass both correctly and incorrectly classified instances, thereby offering interpretability regarding how prototype-based decisions are made.

C. Experimental results using madical dataset.

In addition to standard benchmarks, we evaluate our proposed OFL framework using real medical imaging dataset introduced by Kermany et al. (2018) [46]. This dataset contains X-ray images from pediatric patients aged between 1-5 years old for detecting medical diagnoses and treatable diseases in Medical Center in Guangzhou. The publicly available chest X-ray (CXR) data used in this study were obtained from Kaggle, which has already been used in numerous scientific research studies. The dataset includes 5,863 images divided into three sub-datasets, training, testing, and validation, which are summarized in Table III.

TABLE III. DESCRIPTION OF MEDICAL DATASET.

	Dataset size
Train and validation	5232 (training=4185, validation=1047)
Test	624

Each part has sub-folders with labeled Pneumonia and Normal as shown in Fig. 4.

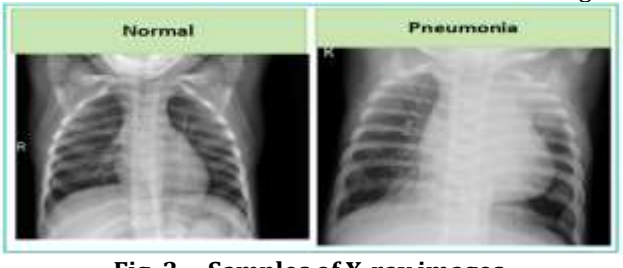


Fig. 3. Samples of X-ray images.

The objective of this OFL application is to detect pneumonia from the provided medical images. To assess the performance of our approach, we utilize a confusion matrix shown in Fig. 5. In this evaluation, True Positives (TP) refer to pneumonia cases correctly identified, totaling 390 instances, while False Negatives (FN) represent pneumonia cases that were missed, which in this case is 0. Conversely, True Negatives (TN) indicate non-pneumonia images correctly classified, amounting to 44, and False Positives (FP) refer to non-pneumonia images that were incorrectly classified as pneumonia, totaling 190. These results underscore the model's perfect sensitivity in detecting pneumonia but also demonstrate a tendency toward over-prediction, as evidenced by the relatively high number of false positives. The obtained accuracy for test dataset is 69.55%.

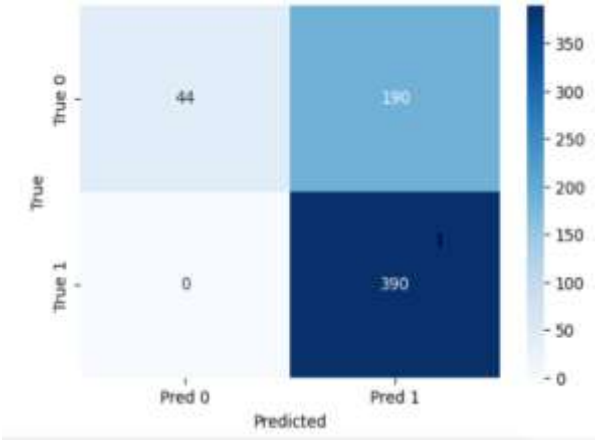


Fig. 4. Confusion matrix.

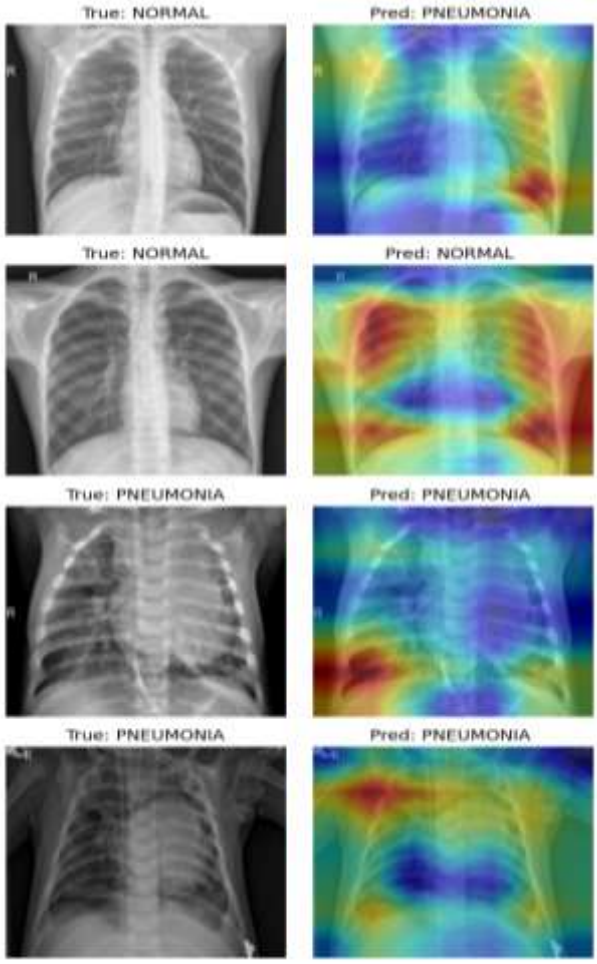


Fig. 5. Grad-CAM visualizations for server-side aggregated prototype classifier using medical dataset.

Beyond the illustrated quantitative results, we used Grad-CAM on the server side to provide interpretability. These visualizations highlight the distinctive zones that guided the global prototype-based classification, offering transparency into the decision-making process for both correct and incorrect predictions.

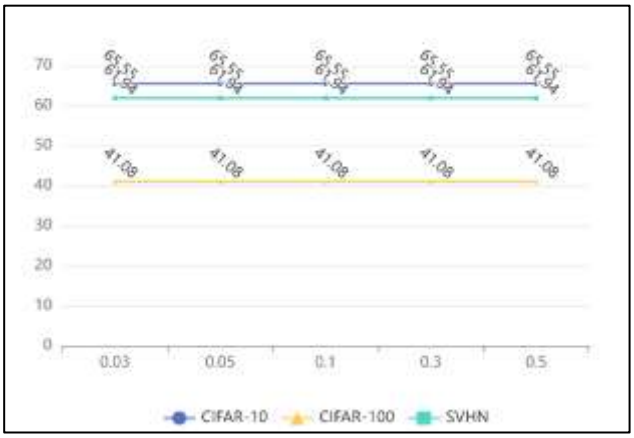


Fig. 6. Fed-RCPA performance for various levels of data heterogeneity.

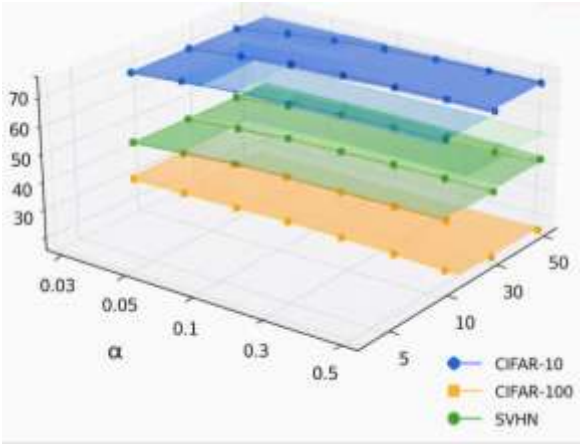


Fig. 7. Fed-RCPA performance for various levels of data heterogeneity and number of clients.

5. FED-RCPA PERFORMANCE EVALUATION

To assess the proposed OFL framework, we systematically evaluate its robustness and stability under varying levels of data heterogeneity. To control it, we apply different values of α , 0.03, 0.05, 0.1, 0.3, and 0.5. As shown in Fig. 7, accuracy remains constant across all α values, indicating strong robustness and stability of our learning framework. This consistency suggests that the OFL method effectively mitigates the impact of data heterogeneity, sustaining stable performance even under extreme distributional shifts. In our OFL, local prototypes corresponding to the same class are combined using their class-wise sample counts. Consequently, variations in how samples of a given class are distributed among clients can be partially compensated during aggregation because the server reconstructs a global class representation from all available client contributions for that class. This class-wise aggregation mechanism therefore reduces the sensitivity of the resulting global representation to changes in client-level class imbalance.Furthermore, we test our OFL for various number of clients. As depicted in Fig. 8, the Fed-RCPA performance remains stable using 5, 10, 30, and 50 clients.

To evaluate our proposed method, we compare it with the existing works in [5], which are the baseline FedAvg [47], Ensemble, and other state-of-the-art OFL methods: DENSE [48], Co-Boosting [18], FedPFT [19], and recent work FedCGS [5].

TABLE IV. COMPARISON OF TEST ACCURACY (%) WITH DIFFERENT METHODS.

Method	α	CIFAR-10	CIFAR-100	SVHN
FedAvg [45]	***0.05***	13.97 ± 1.26	17.57 ± 0.58	16.75 ± 0.63

	0.1	27.28 ± 1.34	21.46 ± 0.38	24.88 ± 0.39
	0.5	51.91 ± 0.46	35.26 ± 0.22	44.39 ± 0.46
Ensemble	***0.05***	38.81 ± 0.89	22.43 ± 0.63	42.26 ± 0.54
	0.1	57.29 ± 0.54	28.07 ± 0.47	53.34 ± 0.23
	0.5	66.00 ± 0.42	37.89 ± 0.49	82.93 ± 0.29
DENSE [46]	***0.05***	31.26 ± 0.73	14.31 ± 0.61	37.49 ± 0.42
	0.1	56.21 ± 0.74	17.21 ± 0.36	51.53 ± 0.37
	0.5	62.42 ± 0.43	26.49 ± 0.32	77.44 ± 0.26
Co-Boosting [16]	***0.05***	44.37 ± 0.42	20.30 ± 0.76	41.90 ± 0.38
	0.1	60.41 ± 0.67	24.63 ± 0.64	57.13 ± 0.28
	0.5	**67.43 ± 0.36**	34.43 ± 0.45	84.65 ± 0.24
FedPFT [17]	***0.05***	56.08 ± 0.49	36.79 ± 0.21	42.55 ± 0.24
	0.1	56.43 ± 0.23	37.16 ± 0.43	43.03 ± 0.14
	0.5	56.80 ± 0.18	37.95 ± 0.27	43.84 ± 0.42
FedCGS [4]	***0.05***	63.95 ± 0.00	39.95 ± 0.00	57.77 ± 0.00
	0.1	63.95 ± 0.00	39.95 ± 0.00	57.77 ± 0.00
	0.5	63.95 ± 0.00	39.95 ± 0.00	57.77 ± 0.00
Our Fed-RCPA	***0.05***	**65.55 ± 0.00**	**41.08± 0.00**	**61.94± 0.00**
	0.1	**65.55 ± 0.00**	**41.08± 0.00**	**61.94± 0.00**
	0.5	65.55 ± 0.00	**41.08± 0.00**	61.94± 0.00

Table IV exhibits the comparative study of several FL methods performance across three benchmark datasets (CIFAR-10, CIFAR-100, and SVHN). As shown, we evaluate our OFL under different heterogeneity levels (α = 0.05, 0.1, 0.5) and we highlighted higher performance in Bold value. The baseline FedAvg demonstrates limited accuracy, especially on CIFAR-100 (17.57–35.26%) and SVHN (16.75–44.39%), indicating its sensitivity to non-IID distributions. More advanced methods such as Ensemble, DENSE, and Co-Boosting achieve significant improvements, with Co-Boosting reaching 84.65% on SVHN at α = 0.5. FedPFT achieves moderate gains but remains less competitive.

FedCGS exhibits stable performance across all α values, consistently achieving 63.95% on CIFAR-10, 39.95% on CIFAR-100, and 57.77% on SVHN, underscoring its robustness to heterogeneity. Moreover, our proposed Fed-RCPA method outperforms all baseline models, maintaining 65.55% on CIFAR-10, 41.08% on CIFAR-100, and 61.94% on SVHN, regardless of α values. This consistency highlights Fed-RCPA's capability to balance accuracy and stability, surpassing existing methods under various conditions.

Recently, Yang et al. [49] introduced novel OFL named Co-Boosting++ where synthetic data generation and ensemble construction mutually enhance each other in an iterative fashion. Co-Boosting++ is extended to support the simultaneous generation of multiple heterogeneous target models, enabling efficient adaptation to diverse device constraints. Our proposed Fed-RCPA attains 65.55% accuracy using ResNet-18 across all α values, demonstrating notable invariance to heterogeneity.

TABLE V. COMPARISON WITH METHODS IN [47] USING CIFAR-10.

Method	α	ResNet
Local	***α = 0.05***	46.49 ± 1.97
	α = 0.1	51.70 ± 1.14
	α = 0.3	67.61 ± 1.54
FedDF	***α = 0.05***	56.51 ± 1.87
	α = 0.1	59.46 ± 1.91
	α = 0.3	70.53 ± 1.19
F-ADI	***α = 0.05***	55.64 ± 1.91
	α = 0.1	58.58 ± 1.48
	α = 0.3	71.61 ± 1.24
F-DAFL	***α = 0.05***	55.14 ± 1.51
	α = 0.1	58.58 ± 1.48
	α = 0.3	71.92 ± 1.95
DENSE	***α = 0.05***	55.79 ± 1.75
	α = 0.1	59.82 ± 1.61
	α = 0.3	72.98 ± 1.61
Co-Boosting++	***α = 0.05***	60.01 ± 1.29
	α = 0.1	63.44 ± 1.35
	α = 0.3	**76.06 ± 1.09**
Our Fed-RCPA	***α = 0.05***	**65.55 ± 0.00**
	α = 0.1	**65.55 ± 0.00**
	α = 0.3	65.55 ± 0.00

This consistency enables Fed-RCPA to outperform all competing methods at α = 0.05 and α = 0.1, while at α = 0.3 it remains the second-best performing method, only slightly trailing Co-Boosting++. The underlined results emphasize this performance, highlighting Fed-RCPA's capacity to balance accuracy and robustness across diverse heterogeneity scenarios. For our evaluation, we specifically used ResNet-18 on CIFAR-10, which serves as the benchmark for evaluating the robustness of our proposed Fed-RCPA method. The baseline Local and FedDF approaches demonstrate moderate improvements as α increases, whereas F-ADI and F-DAFL achieve competitive results that fluctuate. DENSE and Co-Boosting++ consistently outperform the earlier baselines, with Co-Boosting++ achieving 76.06% on ResNet at α = 0.3.

TABLE VI. COMPARISON WITH AVERAGED TEST ACCURACY OF FIVE HOMOGENEOUS TARGET SERVER MODELS OF DIFFERENT METHODS IN [47]

Dataset	α	FedDF	F-ADI	F-DAFL	Co-Boosting++	Our Fed-RCPA
SVHN	***0.05***	60.79 ± 0.52	56.58 ± 1.05	59.38 ± 1.60	67.46 ± 1.08	61.94 ± 0.00
	0.1	68.98 ± 0.63	66.33 ± 1.69	67.77 ± 0.34	74.84 ± 1.16	61.94 ± 0.00
	0.3	79.78 ± 0.55	76.75 ± 1.47	78.01 ± 1.02	83.66 ± 0.89	61.94 ± 0.00
CIFAR-10	***0.05***	37.53 ± 0.67	36.94 ± 1.70	37.82 ± 1.30	49.36 ± 0.97	**65.55 ± 0.00**
	0.1	49.63 ± 0.84	47.19 ± 1.94	46.32 ± 1.69	60.98 ± 1.45	**65.55 ± 0.00**

	0.3	67.18 ± 0.60	60.60 ± 1.32	65.89 ± 1.69	72.28 ± 0.98	65.55 ± 0.00
CIFAR-100	***0.05***	16.07 ± 0.54	13.75 ± 1.01	15.79 ± 0.21	20.86 ± 1.34	**41.08± 0.00**
	0.1	22.07 ± 0.43	19.60 ± 1.17	22.21 ± 1.41	25.16 ± 1.05	**41.08± 0.00**
	0.3	30.71 ± 0.53	26.14 ± 1.37	28.79 ± 1.25	32.52 ± 0.89	**41.08± 0.00**

Table VI illustrates that on CIFAR-10 our proposed Fed-RCPA shows clear superiority at lower heterogeneity levels, consistently achieving 65.55% accuracy at α = 0.05 and α = 0.1. This in-variance to α underscores the method's robustness, as it maintains stable performance where other approaches vary. Although Co-Boosting++ slightly outperforms Fed-RCPA at α = 0.3, our method remains highly competitive, retaining the second-highest ranking. On CIFAR-100, Fed-RCPA achieves a significant advantage over all baseline methods, attaining 41.08% accuracy compared to the 25–32% range observed for competing approaches. This result highlights Fed-RCPA's robust resilience in highly heterogeneous settings.

On SVHN, Fed-RCPA consistently maintains an accuracy of 61.94% across all α values. At α = 0.05, it surpasses FedDF, F-ADI, and F-DAFL, whereas at higher α values, Co-Boosting++ emerges as the efficient OFL. Nevertheless, Fed-RCPA's stability in the face of heterogeneity demonstrates its capability to deliver reliable performance across varied conditions.

6. Conclusion

To tackle the challenges posed by heterogeneity and limited communication in OFL, we introduce Fed-RCPA, an innovative framework that captures Representative Class Prototypes for aggregation. Utilizing global feature statistics, Fed-RCPA facilitates a training-free and heterogeneity-resistant aggregation approach. To further improve prototype quality, we incorporate the Dual Tree Complex Wavelet Transform (DTCWT), offering ameliorated outcomes for aggregation. Additionally, we enhance explainability through Grad-CAM, boosting transparency and trust in decision-making. Fed-RCPA's effectiveness is validated on benchmark datasets (CIFAR-10, CIFAR-100, SVHN) as well as a real-world medical application (pneumonia detection), showing superior performance over existing methods. These findings underscore Fed-RCPA's capacity to balance accuracy, robustness, and interpretability, marking it as a promising direction for practical OFL. The real application underscores the model's perfect sensitivity in detecting pneumonia but also demonstrate a tendency toward over-prediction, as evidenced by the relatively high number of false positives. For this reason, we plan to employ optimization methods for hyper-parameter tuning. Furthermore, the combination and alignment of various encoders constitutes a novel approach for pre-trained model-based OFL. Additional future work will investigate adaptive feature-layer selection, optimization of the fusion between DTCWT and deep representations, improved prototype aggregation strategies, and broader evaluation under different forms of data and model heterogeneity. We plan to deploy the proposed OFL in real-world federated settings like [50], where data is naturally spread out and non-IID across the participating institutions.

DECLARATION OF COMPETING INTERESTS

The author declares that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

ACKNOWLEDGMENT

The authors declare that they received no financial support for the research, authorship, and/or publication of this article. No individuals, institutions, or infrastructures beyond the listed authors contributed to this work.

DATA AVAILABILITY

The datasets used in this study are publicly accessible and can be obtained from established repositories. No proprietary or restricted data were employed.

References

[1] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," IEEE Signal Processing Magazine, vol. 37, no. 3, pp. 50–60, May 2020.

[2] A. Alsarhan, M. Barhoush, B. Khassawneh, M. Al-Essa, M. Aljaidi, and Q. Al-Na'amneh, "Deep learning utilization for DDoS attack detection with federated learning: A case study on the CICDDoS2019 dataset," Eng., Technol. Appl. Sci. Res., vol. 16, no. 1, pp. 31203–31208, 2026.

[3] N. Arfaoui, A. Ksibi, N. A. Almujally, and R. Ejbali, "Empowering e-learning approach by the use of federated edge computing," Cluster Computing, vol. 27, no. 10, pp. 13737–13748, 2024

[4] F. Amato, G. Cozzolino, A. De Luca, and M. Panella, "One-shot federated learning: Efficient global model training with minimal communication," Future Generation Computer Systems, Apr. 2025.

[5] Z. Guan, Y. Zhou, and X. Gu, "Capture global feature statistics for one-shot federated learning," in Proc. AAAI Conf. Artificial Intelligence, vol. 39, no. 16, Apr. 2025, pp. 16942–16950.

[6] J. Chen, L. Wang, H. Zhang, and Y. Liu, "Addressing heterogeneity in one-shot federated learning for medical imaging," IEEE Journal of Biomedical and Health Informatics, 2025.

[7] M. H. Elhajjej, S. Said, N. Arfaoui, and R. Ejbali, "Asynchronous and Flexible Federated Learning for COVID-19 Detection from Chest X-Ray Images," Journal of Advanced Computational Intelligence and Intelligent Informatics, vol. 29, no. 4, pp. 857–867, 2025

[8] Y. Liu, H. Zhang, and K. Wang, "Feature-aware hierarchical token sequences for one-shot federated learning," IEEE Trans. Neural Netw. Learn. Syst., vol. xx, no. xx, Jan. 2026a.

[9] Y. Liu, H. Zhang, and K. Wang, "Knowledge distillation strategies for non-IID federated learning," Pattern Recognition Letters, Mar. 2026b.

[10] Y. Liu, R. Dai, Y. Zhang, A. Li, T. Liu, and B. Han, "FALCON: Federated learning with hierarchical token sequences and knowledge distillation," IEEE Trans. Pattern Anal. Mach. Intell., 2026.

[11] Y. Bai, L. Chen, and Z. Wu, "FedHydra: Resolving data and model heterogeneity in federated learning via two-stage stratified aggregation," Proc. Int. Conf. Machine Learning (ICML), Vienna, Austria, Jul. 2025, .

[12] H. Gao, J. Sun, and M. Zhou, "Dual-prototype contrastive learning for federated clustering," IEEE Trans. Knowledge Data Eng., Oct. 2025.

[13] Y. Zheng, F. Li, and P. Xu, "Aggregation bias in federated learning under heterogeneous prototypes," IEEE Trans. Big Data, Feb. 2026.

[14] H. Fang, J. Li, and Y. Zhao, "Class feature distribution variability across federated clients," IEEE Trans. Neural Netw. Learn. Syst., vol. xx, no. xx, 2025.

[15] X. Meng, L. Sun, and P. Zhang, "Addressing non-IID data distributions in federated learning," Pattern Recognition, 2025.

[16] Y. Guo, T. Wang, and R. Chen, "Cross-client migration and unbiased global model development in federated learning," IEEE Trans. Parallel Distrib. Syst., 2024.

[17] J. Chen, L. Wu, and M. Huang, "Privacy-aware one-shot federated learning for efficient knowledge transfer," Future Generation Computer Systems, 2024.

[18] R. Dai, Y. Zhang, A. Li, T. Liu, X. Yang, and B. Han, "Enhancing one-shot federated learning through data and ensemble co-boosting," in Proc. Int. Conf. Learning Representations (ICLR), Vienna, Austria, May 2024.

[19] M. Beitollahi, A. Bie, S. Hemati, L. M. Brunswick, X. Li, X. Chen, and G. Zhang, "Parametric feature transfer: One-shot federated learning with foundation models," arXiv preprint arXiv:2402.01862, Feb. 2024.

[20] M.Ishaq, M.Bilal, Sultan Alshehery, A.Thaljaoui, S. M Eljack and M. Abushaega, "Deep neural network-based predictive modeling of entropy generation in electroosmotic transport of six-constant ," Nonlinear Dynamics 2025.

[21] Y. Liu, H. Zhang, and K. Wang, "Sophisticated aggregation strategies for reconciling diverse client outputs in one-shot federated learning," IEEE Trans. Neural Netw. Learn. Syst., 2024.

[22] X. Yang, R. Dai, Y. Zhang, A. Li, T. Liu, and B. Han, "Increasing model diversity to mitigate non-IID effects in federated learning," Proc. Int. Conf. Artificial Intelligence and Statistics (AISTATS), Apr. 2024.

[23] J. Wang, L. Chen, and Z. Wu, "Model diversity for representative global aggregation in federated learning," Pattern Recognition, 2024.

[24] T. Le, P. Nguyen, and H. Tran, "Balancing efficiency and complexity in one-shot federated learning," IEEE Trans. Big Data, 2024.

[25] Y. Luo, F. Xu, and M. Zhou, "Computational overhead in heterogeneous federated learning aggregation," Future Generation Computer Systems, 2024.

[26] R. Mendieta, S. Patel, and J. Kim, "Minimizing bias in one-shot federated learning without iterative reconciliation," Proc. Int. Conf. Machine Learning (ICML), Vienna, Austria, Jul. 2024.

[27] Z. Wang, F. Yan, T. Wang, C. Wang, Y. Shu, P. Cheng, and J. Chen, "Fed-DFA: Federated Distillation for Heterogeneous Model Fusion Through the Adversarial Lens," in Proc. AAAI Conf. Artificial Intelligence, vol. 39, no. 20, pp. 21429–21437, 2025. doi:10.1609/aaai.v39i20.35444.
[28] P. Wang, C. Shen, M. Oda, C.-S. Fuh, K. Mori, W. Wang, and H. R. Roth, "Conditional Distillation for Federated Learning with Partially Labeled Data," IEEE J. Biomed. Health Inform., early access, 2025. (arXiv:2412.18833).
[29] M. Yao, Z. Wang, W. Chen, J. Liu, T. Xu, T. Ji, and X. Li, "FedMKAN: Federated Meta Kolmogorov-Arnold Network on Non-IID Data," in Proc. Int. Conf. Artificial Intelligence, Springer LNCS, 2025.
[30] H. Holland, J. Xu, and K. Wang, "When Federated Learning Meets Knowledge Distillation," IEEE Trans. Neural Netw. Learn. Syst., vol. 35, no. 7, 2024.
[31] J. Zhang, C. Shan, and J. Han, "FedGMKD: An Efficient Prototype Federated Learning Framework Through Knowledge Distillation and Discrepancy-Aware Aggregation," in Proc. NeurIPS, 2024. doi:10.52202/079017-3757.
[32] Y. Li, L. Fu, T. Wang, J. Lou, B. Chen, L. Yang, J. Shen, Z. Zheng, and C. Chen, "Clients Collaborate: Flexible Differentially Private Federated Learning with Guaranteed Improvement of Utility-Privacy Trade-off," in Proc. 42nd Int. Conf. Machine Learning (ICML), PMLR vol. 267, pp. 34333–34354, 2025.
[33] A. K. Qin, W. Mao, B. Yu, and C. Zhang, "FedKT: Federated Learning with Knowledge Transfer for Non-IID Data," Pattern Recognition, vol. 158, Mar. 2025.
[34] H. Xu, J. Li, W. Wu, and H. Ren, "Federated Learning with Sample-Level Client Drift Mitigation," in Proc. AAAI Conf. Artificial Intelligence, vol. 39, no. 20, pp. 21752–21760, 2025. doi:10.1609/aaai.v39i20.35480.
[35] R. Zhao, Y. Chan, R. Zhou, Z. Jiang, and E. Ngai, "Internal Cross-Layer Gradients for Extending Homogeneity to Heterogeneity in Federated Learning," in Proc. Int. Conf. Learning Representations (ICLR), 2024.
[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 770–778.
[37] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, J. Houlsby, et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in Proc. Int. Conf. Learning Representations (ICLR), Vienna, Austria, May 2021.
[38] J. Chen, X. Liu, and Y. Wang, "Federated learning with pre-trained models: Challenges and opportunities," IEEE Trans. Neural Netw. Learn. Syst., vol. 33, no. 12, pp. 7070–7083, Dec. 2022.
[39] T. Nguyen, H. Pham, and M. Le, "Adapting vision transformers for federated learning," in Proc. Int. Conf. Neural Information Processing Systems (NeurIPS), New Orleans, LA, USA, Dec. 2022.
[40] Y. Liu, J. Peng, J. James, and Y. Wu, “PPGAN: Privacy-preserving gener ative adversarial network,” in Proc. IEEE 25th Int. Conf. Parallel Distrib. Syst., 2019, pp. 985–989.
[41] N. G. Kingsbury, "The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters," in Proc. IEEE Digital Signal Processing Workshop, Bryce Canyon, UT, USA, Aug. 1998, pp. 120–125.
[42] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 618–626, 2017. DOI: 10.1109/ICCV.2017.74.
[43] A. Krizhevsky, V. Nair, and G. Hinton, "CIFAR-100," Technical Report, University of Toronto, Canadian Institute for Advanced Research (CIFAR), 2009. Available: http://www.cs.toronto.edu/~kriz/cifar.html
[44] A. Krizhevsky, V. Nair, and G. Hinton, "Learning multiple layers of features from tiny images," Technical Report, University of Toronto, Canadian Institute for Advanced Research (CIFAR), 2009.
[45] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, "Reading digits in natural images with unsupervised feature learning," in Proc. Neural Information Processing Systems Workshop on Deep Learning and Unsupervised Feature Learning, Granada, Spain, Dec. 2011.
[46] D. S. Kermany, M. Goldbaum, W. Cai, C. C. S. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, J. Dong, M. K. Prasadha, J. Pei, M. Y. L. Ting, J. Zhu, C. Li, S. Hewett, J. Dong, I. Ziyar, and A. Shi, "Identifying medical diagnoses and treatable diseases by image-based deep learning," Cell, vol. 172, no. 5, pp. 1122–1131.e9, Feb. 2018. doi: 10.1016/j.cell.2018.02.010
[47] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Aguera y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. 20th Int. Conf. Artificial Intelligence and Statistics (AISTATS), Fort Lauderdale, FL, USA, Apr. 2017, pp. 1273–1282.
[48] J. Zhang, C. Chen, B. Li, L. Lyu, S. Wu, S. Ding, C. Shen, and C. Wu, "DENSE: Data-Free One-Shot Federated Learning," Advances in Neural Information Processing Systems (NeurIPS), vol. 35, Dec. 2022, Art. no. 2112.12371.

[49] X. Yang, R. Dai, Y. Zhang, A. Li, T. Liu, and B. Han, "Co-Boosting++: Coupled optimization of data and ensemble for one-shot federated learning," IEEE Trans. Pattern Anal. Mach. Intell., 2026.
[50] S. Ghazal, R. Manan, A. Khalid, M. Shahid, U. F. Khattak, and M. A. Khan, "Modified federated learning for Parkinson's disease prediction," Eng., Technol. Appl. Sci. Res., vol. 16, no. 2, pp. 34235–34240, 2026.