

Review

Not peer-reviewed version

A Survey of Recent Advances for Tackling Data Heterogeneity in Federated Learning

Alexander Johnson *

Posted Date: 18 February 2025

doi: 10.20944/preprints202502.1159.v1

Keywords: data heterogeneity; federated learning; NonIID



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Review

A Survey of Recent Advances for Tackling Data Heterogeneity in Federated Learning

Alexander Johnson

New York, United States of America; alexanderjohnsonaj123@gmail.com

Abstract: Federated Learning (FL), as a distributed machine learning framework, allows multiple parties to collaboratively train models without sharing their data, thereby protecting privacy and data security. However, the issue of data heterogeneity—where data distributions, feature spaces, and label spaces vary significantly across different clients—poses a critical challenge to the effectiveness of federated learning. To address this problem, researchers have proposed various solutions, including techniques to mitigate local model drift, adaptive model aggregation, local data augmentation, and personalized federated learning. These strategies collectively enhance the capability of federated learning in handling data heterogeneity, promoting its widespread application across numerous fields. This review aims to summarize and discuss the latest advancements in these technologies.

Keywords: data heterogeneity; federated learning; NonIID

1. Introduction

With the increasing awareness of data privacy protection and advancements in technology, Federated Learning (FL)[1,2] has emerged as a novel distributed machine learning framework, gradually becoming a hot research area in both academia and industry. Federated learning enables multiple parties to collaboratively train models without sharing their local data, thereby effectively protecting user privacy and data security. This approach is not only suitable for cross-organizational data collaboration in fields such as smart transportation [3], medical image processing [4–6] and recommendation [7,8], significantly expanding the scope of machine learning applications.

However, despite its numerous advantages, federated learning faces a significant challenge in practical applications: the issue of data heterogeneity. Data heterogeneity refers to the significant differences in data distribution including feature space [9,10] and label space [11,12] across different clients. For example, in medical image analysis, hospitals may collect vastly different imaging data due to geographical location and patient demographics; in financial risk assessment, customer profiles among different banks can vary greatly. These differences lead to increased discrepancies between local models, affecting the performance and robustness of the global model. Therefore, how to effectively address the problem of data heterogeneity has become a critical subject in federated learning research.

To tackle this challenge, researchers have proposed various solutions, including but not limited to mitigating local model drift, personalized federated learning, local data augmentation, improved aggregation mechanisms [13]. Local model drift mitigation aims to reduce the impact of data heterogeneity by minimizing the differences between local models on different clients [14]. Personalized federated learning [15] seeks to generate models tailored to each client's specific data environment through techniques like local fine-tuning, meta-learning, or multi-task learning, improving the alignment between local models and local data while ensuring consistency and stability of the global model. Local data augmentation [16] enriches the training datasets of clients using data augmentation techniques (such as data expansion, synthetic data generation), reducing the negative impacts of data sparsity and heterogeneity. Improved aggregation mechanisms [17] optimize traditional simple average aggregation methods by adopting importance-weighted aggregation, diversified aggregation

rules, and adaptive aggregation algorithms to mitigate the influence of data heterogeneity on the global model. While these methods have different focuses, their common goal is to enhance the ability of federated learning systems to handle data heterogeneity, thereby improving the overall performance of the global model.

Recently, several surveys have also summarized the latest research progress [18]. For instance, [19] systematically categorizes data heterogeneity, statistical heterogeneity, system heterogeneity, and model heterogeneity, proposing hierarchical solutions for each type of heterogeneity. Meanwhile, [20] summarizes innovative applications of knowledge distillation techniques in federated learning concerning privacy protection, communication optimization, and personalization. It classifies knowledge distillation into feature distillation and output distillation according to technical approaches, comparing these methods with traditional FL methods, thereby highlighting the potential of knowledge distillation in enhancing efficiency and effectiveness. Despite the extensive coverage provided by these surveys on data heterogeneity issues and their solutions, there are still some methodological shortcomings in investigating data heterogeneity. Although each survey attempts to analyze data heterogeneity from different perspectives, they adopt varying classification standards and frameworks. This inconsistency may make it challenging for researchers to obtain a comprehensive and unified understanding. Most surveys focus on reviewing existing technologies and methods, with inadequate predictions about future development trends and how emerging technologies might influence strategies for handling data heterogeneity. This limitation reduces their guidance value for long-term research directions.

In conclusion, with the rapid development of federated learning technology and its broad application prospects [21], addressing the issue of data heterogeneity has become a key factor in advancing this field. It is hoped that through this review, readers will gain a comprehensive understanding of the various methods currently used to overcome data heterogeneity and recognize the necessity of continuous technological innovation for achieving more efficient, reliable, and widely applicable federated learning systems. Future research will continue to focus on developing advanced algorithms and technologies to meet the increasingly complex demands of real-world application scenarios, propelling federated learning to new heights. This review will focus on discussing several primary methods addressing data heterogeneity in federated learning and their latest research advancements. Initially, we will elaborate its specific impacts on federated learning, helping readers understand why it is an urgent issue to resolve. Subsequently, the article will delve into the fundamental principles, implementation steps, and application scenarios of each solution, showcasing their practical outcomes through specific case studies. Lastly, the paper will summarize the main contributions and limitations of existing research findings, offering suggestions for further investigation, aiming to provide valuable references for researchers in related fields.

2. Local Model Drift Mitigation

To train a global model, numerous efforts have been dedicated to addressing the Non-IID (independent and identically distributed) challenge in Federated Learning (FL) [14,22–24]. In response to this challenge, the research community has primarily developed two types of solutions. On one hand, some works aim to alleviate the issue of uneven data distribution by adding a regularization term to the local loss function. This approach seeks to introduce additional constraints so that each client's model updates are not only based on its own local data but also take into account the consistency requirements across the entire network. The goal is to reduce model bias caused by differences in data distribution. On the other hand, other studies focus on calibrating local model updates with global information to achieve consistency among clients. Specifically, these methods might use global model parameters or statistical information as reference points to guide participating parties in adjusting their update directions and steps. The objective is to ensure that knowledge contributed by each client

can be effectively integrated into the global model, even when there are significant differences in data distribution, thereby enhancing the overall performance and generalization ability of the model.

2.1. Regularization-Based Drift Mitigation

Several methods propose adding regularization to the local loss function to achieve consistency among locally updated models [25–28]. For instance, FedProx [26] introduces a proximal term into the local objective function, which encourages the local updates to move towards both the local optimum and the previously received global model. This approach helps mitigate the impact of data heterogeneity by ensuring that local updates remain close to the global model. Another innovative approach is proposed in [27], where a dynamic regularizer is introduced based on the current local model and the received global model. This method aims to achieve the same stationary point across all clients, thereby enhancing the overall consistency of the federated learning process. In [29], the authors integrate the local training process with a primal-dual algorithm to enhance the consistency among different variable models. This dual approach ensures that while each client's model is optimized for its local data, it also remains aligned with the broader federated model. FedSpeed [30] addresses the issue that the prox-term can introduce bias into the local updates. To counteract this, they apply a prox-correction term to the current local updates, effectively reducing the bias and improving convergence efficiency. The work presented in [31] highlights the impact of hyperparameters in the local update process on achieving consistency. By carefully tuning these parameters, they regularize the local update process, ensuring better alignment between local and global models. Building on this, [32] propose a surrogate loss function for quadratic models, demonstrating that the local learning rate decay can balance the trade-off between the convergence rate and inconsistency. Additionally, [33] employs a local fixed-point strategy to implicitly control the convergence of the local model, ensuring stability and consistency within the federated learning framework. Finally, [34] propose an adaptive method for tuning the global step size by computing a regularized term of all local updates, addressing inconsistencies in local updates and promoting more coherent global model evolution. These approaches collectively contribute to advancing the state-of-the-art in federated learning by addressing the critical challenge of maintaining model consistency in the face of data heterogeneity, thus paving the way for more robust and scalable federated learning systems.

2.2. Calibration-Based Drift Mitigation

Other works consider that the local gradient is biased and then correct the local gradient to align the global gradient [30,35–38]. For instance, [35] first identifies the issue of local update drift in Federated Learning (FL) and proposes Scaffold, a method that leverages the bias between the local gradient and the global gradient to mitigate these drifts. Similarly, [39] acknowledges that the drift between the local optimal model and the global optimal model inherently exists and suggests learning a drift term to compensate for the local gradient. Another approach, proposed by [40], involves incorporating global gradient information into the local training process to reduce local bias. These methods collectively aim to address the challenges posed by data heterogeneity and non-IID (non-independent and identically distributed) data distributions in FL. In addition to gradient correction, some studies introduce momentum into FL frameworks. Momentum, which encapsulates information from past gradients across clients, implicitly helps calibrate the drift. For example, [41–44] explore the use of momentum to stabilize training and improve convergence in FL settings. By leveraging historical gradient information, these methods aim to reduce the divergence between local and global models, particularly in scenarios where clients have highly heterogeneous data.

Another line of research focuses on addressing the bias in the last layer of neural networks, which is critical for classification tasks. FedRS [45] introduces an asymmetric loss function to calibrate the bias of the last-layer parameters across different classes. Similarly, other works seek to align the outputs of the last layer or the penultimate layer with the global model to achieve consistency. For instance,

[46–48] propose techniques to fine-tune the final layers of the model, ensuring that local updates do not deviate significantly from the global model. These approaches are particularly effective in classification tasks, where the last layer plays a pivotal role in determining the model's performance. Recently, a more fine-grained approach to addressing client drift has emerged. Unlike previous works that treat the training of deep neural networks (DNNs) as a whole and consider the drift of the entire local update, FedNLR [49] reframes the problem of model drift as a neuron drift problem. This perspective views client drift as a phenomenon occurring at the neuron level, offering a more granular understanding of the issue. To address this, FedNLR adopts neuron-wise learning rates during the local training process, allowing for more precise adjustments to individual neurons and thereby reducing overall model drift. This approach represents a significant advancement in addressing the challenges of FL, particularly in scenarios involving highly heterogeneous data and complex model architectures. In summary, while traditional FL methods often struggle with issues such as local gradient bias, model drift, and data heterogeneity, recent advancements have introduced innovative techniques to mitigate these challenges. From gradient correction and momentum-based methods to fine-grained neuron-level adjustments, these approaches collectively enhance the robustness and efficiency of FL, paving the way for more effective collaborative learning in distributed environments.

3. Effective Model Aggregation

Federated Averaging (FedAvg) aggregates models by computing a weighted average of them based on the amount of data each client possesses. However, while FedAvg has proven effective in many scenarios, it may struggle with highly non-IID data distributions, leading to suboptimal global models. Considering this, some aggregation techniques are designed to enhance the model aggregation step on the server side, aiming to produce a more robust and accurate global model [17].

3.1. Adaptive Aggregation Weights

One of the typical research directions in federated learning is the determination of adaptive aggregation weights, which play a crucial role in balancing the contributions of different clients to the global model. Several studies have explored this area, such as [50–52], which propose various methods to dynamically adjust aggregation weights based on client-specific characteristics. For instance, AUTO-FEDAVG [53] tailors aggregation weights based on distinct institutional medical datasets, enabling personalized medicine by accounting for the unique data distributions and requirements of each institution. Similarly, L2C [50] identifies similar peers in decentralized federated learning settings and adapts aggregation weights using local data, ensuring that clients with similar data distributions contribute more significantly to the global model. While these approaches have demonstrated effectiveness in creating personalized models for individual clients, they primarily focus on optimizing local performance rather than achieving a robust global model. Considering this, some other works shift the focus toward acquiring a high-quality global model that generalizes well across all clients. This is particularly important in scenarios where the primary goal is to develop a unified model that can be deployed universally, rather than tailoring models to individual clients. Recently, FedLAW [54] has made strides in this direction by learning aggregation weights to achieve a global model. However, a common limitation of these methods, including FedLAW, is their reliance on a proxy dataset available on the server. This dependency can be problematic, as it assumes the server has access to representative data, which may not always be feasible or practical, especially in privacy-sensitive applications where data cannot be shared or centralized.

3.2. Model Fusion

Due to the permutation invariance of neural network parameters, several works have highlighted that the ordering of parameters across different local models on clients may vary significantly, especially when data is Non-IID [55–58]. This variability can cause significant issues during the aggregation

process. Specifically, a straightforward coordinate-wise average of local models might result in mismatches between corresponding parameters from different clients' local models, leading to degraded performance of the aggregated global model. To address this challenge, researchers have proposed various strategies aimed at reordering or aligning the parameters of local models before aggregation. These approaches seek to ensure that similar parameters are matched across different clients, thereby improving the effectiveness of the aggregation process. One approach involves using the Hungarian algorithm [59] to optimally match the parameters of local models. This method formulates the parameter alignment problem as an assignment problem where the goal is to minimize the overall mismatch cost between corresponding parameters across different clients. Another strategy employs Bayesian methods [60] to infer the most probable parameter alignment. By modeling the uncertainty in parameter positions, this approach provides a probabilistic framework for aligning parameters, which can be particularly useful in scenarios with high variability in data distributions. A more recent development involves leveraging graph matching algorithms [61] to align the parameters. In this context, each local model's parameters are represented as nodes in a graph, and edges are formed based on similarity metrics. The graph matching algorithm then seeks to find the optimal alignment that maximizes the total edge weight, effectively aligning similar parameters across different clients. These techniques aim to mitigate the negative impact of parameter permutation invariance by ensuring that the aggregated model benefits from well-aligned contributions from all clients. However, these methods also introduce additional computational overhead and complexity, necessitating careful consideration of trade-offs between alignment accuracy and efficiency. Moreover, future research could explore hybrid approaches that combine multiple alignment strategies or integrate them with other federated learning enhancements such as differential privacy or adaptive aggregation. By doing so, it may be possible to achieve both robustness against data heterogeneity and efficient model aggregation, paving the way for more scalable and effective federated learning systems.

3.3. Federated Distillation

Knowledge Distillation (KD) is a widely used technique to transfer knowledge from one or more pre-trained networks, referred to as teachers, to another network, known as the student [62]. The core principle of knowledge distillation lies in aligning the soft predictions (e.g., logits or probability distributions) of the student model with those of the teacher model. This alignment enables the student to mimic the behavior of the teacher, often leading to improved generalization and performance, even when the student model is smaller or less complex [63–66]. Building on this idea, federated distillation extends the concept to distributed learning environments by employing ensemble distillation. In this approach, the logits or soft predictions of multiple local models (acting as teachers) are aggregated and averaged, and the global model (student) is trained to align with this ensemble output [67–70]. This method allows the global model to benefit from the collective knowledge of all participating local models, even in the absence of direct data sharing. A significant advancement in federated distillation was introduced by [71,72], who proposed a server-side knowledge distillation technique. This method leverages an unlabeled proxy dataset on the server to transfer knowledge from multiple local models to the global model. By distilling the ensemble of local models into a single global model, this approach addresses some of the challenges posed by data heterogeneity and non-IID (non-independent and identically distributed) data distributions across clients. However, a limitation of these methods is that they assign equal weights to all local models during the distillation process, disregarding the varying quality and relevance of each local model's knowledge. To address this, DaFKD [22] introduced an adaptive weighting mechanism that assigns specific weights to each local model based on its contribution to the ensemble distillation. This adaptive weighting strategy better accounts for data heterogeneity and ensures that more reliable or relevant local models have a greater influence on the global model.

Despite these advancements, a common limitation of many federated distillation methods is their reliance on an auxiliary or proxy dataset available on the server. In real-world federated learning scenarios, such a dataset may not be accessible due to privacy concerns, regulatory restrictions, or practical constraints. To overcome this limitation, recent studies have explored data-free federated distillation techniques. For example, [22,73,74] proposed replacing the proxy dataset with synthetically generated data. These methods use generative models or other data synthesis techniques to create representative samples that mimic the distribution of the local data, enabling ensemble distillation without requiring actual data on the server. This data-free approach not only preserves privacy but also enhances the practicality of federated distillation in scenarios where data sharing is prohibited or impractical.

In summary, federated distillation represents a powerful paradigm for knowledge transfer in distributed learning environments. By leveraging ensemble distillation, adaptive weighting, and data-free techniques, researchers have made significant strides in addressing challenges such as data heterogeneity, privacy preservation, and the absence of proxy datasets. These advancements pave the way for more robust and scalable federated learning systems, enabling the deployment of high-performance global models in a wide range of real-world applications.

4. Client Selection

Client selection has emerged as a critical method to optimize the coordination and efficiency of the Federated Learning (FL) process by strategically choosing a subset of participating clients for each training round. Since the introduction of the original random sampling approach, several advanced client selection techniques have been developed to address the limitations of naive random selection [75–78]. Recent research has demonstrated that well-designed client selection strategies can significantly enhance model performance by improving the convergence rate of FL training [79,80], reducing the number of required training rounds [81], and promoting fairness in scenarios with unbalanced data distributions [82–84]. These improvements are achieved through various mechanisms, such as sampling clients based on the size of their local datasets [85,86], clustering clients with similar data characteristics [87], or prioritizing clients based on the magnitude of their local loss values [88–90]. These strategies aim to ensure that the selected clients contribute meaningfully to the global model update, thereby accelerating training and improving model accuracy.

However, many of these client selection strategies overlook the inherent heterogeneity in client capabilities, such as computational power, communication bandwidth, and data quality. This oversight can lead to inefficiencies, as clients with limited resources may slow down the training process or provide updates that are less informative. To address this limitation, state-of-the-art methods like DivFL [91] have introduced more sophisticated selection criteria. DivFL, for instance, proposes selecting a small but diverse subset of clients whose aggregated updates can approximate the contributions of the entire client population. By prioritizing clients with representative gradient information, DivFL not only boosts training efficiency but also ensures that the selected subset captures the diversity of the overall data distribution. This approach mitigates the risk of bias introduced by non-representative client selection and enhances the robustness of the global model.

In addition to improving efficiency and fairness, advanced client selection techniques also play a crucial role in addressing challenges such as stragglers (clients that are slow to respond) and dropouts (clients that fail to participate in a round). For example, methods that account for client resource heterogeneity can dynamically adjust the selection process to prioritize clients with sufficient computational and communication capabilities, thereby reducing delays and improving the overall reliability of the FL system. Furthermore, incorporating client selection strategies that consider data quality and relevance can help mitigate the impact of noisy or low-quality data, leading to more accurate and reliable global models.

In summary, client selection has evolved from simple random sampling to sophisticated strategies that consider factors such as data distribution, client capabilities, and gradient diversity. These advancements have significantly improved the efficiency, fairness, and robustness of FL systems. By addressing the challenges posed by client heterogeneity and resource limitations, state-of-the-art methods like DivFL demonstrate the potential of intelligent client selection to unlock the full potential of federated learning in real-world applications. As FL continues to scale to larger and more diverse environments, further innovations in client selection will be essential to ensure optimal performance and scalability.

5. Data Augmentation

Data augmentation is a powerful technique that enhances model generalization by leveraging additional samples, which can be generated through various methods such as synthetic data generation [16] or by utilizing publicly available datasets [92]. In the context of federated learning, where client data is often heterogeneous and unevenly distributed, traditional training approaches may struggle to achieve optimal model performance. Data augmentation addresses these challenges by introducing a variety of transformations, including geometric modifications like rotations, flips, and scaling, as well as more advanced techniques such as generative adversarial networks (GANs) to create synthetic data. These approaches help alleviate issues related to data scarcity and distribution shifts, ultimately improving the robustness and generalization capabilities of federated learning models [93].

One notable data augmentation method is Mixup, which generates new training samples through linear interpolation between pairs of existing samples. This technique has been successfully integrated into federated learning frameworks to enhance model performance, particularly in scenarios with highly non-independent and identically distributed (non-IID) data. For example, FedMix [94] incorporates the Mixup technique to improve model accuracy while preserving user privacy by avoiding the direct sharing of raw data. Additionally, generative models like GANs are employed to synthesize data that closely resembles the real data distribution, effectively addressing the issue of insufficient local data on client devices. Another innovative approach, RandAugment [16], automates the data augmentation process by randomly applying a sequence of image transformations, thereby generating diverse and robust training samples that enhance model performance on large-scale datasets.

Beyond traditional data augmentation, some studies focus on transforming heterogeneous feature spaces into homogeneous ones through feature mapping techniques. For instance, certain works [95] propose augmenting feature representations to align disparate feature spaces, enabling more effective model training across diverse data sources. This approach is particularly valuable in federated learning, where data from different clients may exhibit significant variability in feature distributions.

Collectively, these techniques—ranging from traditional data augmentation and Mixup to advanced generative models and feature mapping—play a crucial role in addressing the challenges posed by data heterogeneity in federated learning. By incorporating these strategies, federated learning systems can achieve greater robustness, adaptability, and performance across a wide range of applications. This not only improves the accuracy and reliability of models but also ensures that privacy and data security are maintained, making federated learning a more viable solution for real-world, distributed data scenarios.

6. Personalized Federated Learning

Personalized federated learning is to train personalized models for each client to mitigate the impact of data heterogeneity over the global model.

6.1. Partial Sharing Based Personalized Federated Learning

Partial parameter sharing has emerged as one of the primary strategies for achieving personalized federated learning (PFL), enabling clients to maintain customized models while still benefiting from

collaborative learning. Previous studies in this domain have primarily focused on sharing specific layers of neural networks to achieve personalization. For instance, some works have explored sharing only the batch normalization layers while keeping other layers client-specific, allowing for localized adaptation while maintaining some level of global consistency [10]. More recent research has shifted toward sharing the full feature extractor while customizing only the classifier head, a strategy that has shown promise in balancing global generalization with local personalization [96–99,99–101]. This approach leverages the feature extractor to capture shared knowledge across clients while allowing the classifier head to adapt to local data distributions, thereby achieving a more personalized model.

Several notable methods have been proposed within this framework. For example, Fed-RoD [96] shares the full feature extractor but employs different softmax functions for global and local learning, enabling clients to tailor their models to specific tasks while still benefiting from a shared feature space. Similarly, FedBABU [98] maintains a fixed global classifier during local fine-tuning, ensuring that the global model remains stable while allowing clients to adapt the feature extractor to their unique data. Another innovative approach, kNN-Per [102], combines a global model with local k-nearest neighbors (kNN) classifiers, enhancing personalization by leveraging both global and local decision boundaries. LG-FedAvg [103] takes a different approach by jointly learning the entire network during local updates and aggregating only the top layers based on pre-trained global networks via FedAvg, ensuring that the shared layers remain consistent across clients. FedRep [101] shares the feature extractor by averaging its parameters across clients, while FedProto [104] aligns feature representations among clients to achieve a shared feature space. The latest method, FedPAC [105], further refines this approach by simultaneously averaging feature extractor parameters and aligning feature representations, achieving a more robust and harmonized shared model.

In summary, partial parameter sharing has become a cornerstone of personalized federated learning, with various methods exploring different strategies for sharing and customizing model components.

6.2. Regularization Based Personalization

Regularization methods play a pivotal role in federated learning by imposing constraints on local training to enhance the personalization of local models while maintaining alignment with the global objective. These methods employ various regularizers to guide the optimization process, ensuring that local models do not deviate excessively from the global model, especially in non-IID data settings. Several studies have explored the use of model parameters to construct regularizers, providing explicit guidance for local training. For instance, [106] introduces a dynamic regularization approach that adapts to the local data distribution, while [107] proposes a personalized federated learning framework that leverages model parameter-based regularizers to balance personalization and generalization.

To address the challenges posed by data drift in non-IID environments, some works focus on correcting the update direction for each client. For example, [108] introduces a control variate mechanism to reduce client drift by aligning local updates with the global model, and [109] proposes a method that dynamically adjusts the aggregation weights based on client performance, ensuring more effective updates in heterogeneous settings. These approaches help mitigate the adverse effects of data heterogeneity, enabling more stable and efficient federated learning.

Another line of research emphasizes aligning the representations or prototypes of heterogeneous clients to facilitate better knowledge transfer and reduce communication overhead. Works such as [110] and [111] propose methods to align client-specific prototypes or feature representations, enabling the global model to learn a more robust feature extractor with fewer communication rounds. These techniques not only improve model performance but also reduce the computational and communication costs associated with federated learning.

Recent advancements in regularization methods have focused on leveraging soft labels or statistical information to enhance knowledge sharing among clients. For instance, [112] utilizes soft

labels to regularize local training, enabling clients to benefit from the collective knowledge of the federation without sharing raw data. Similarly, [113] introduces a framework that aligns client models using statistical information, promoting consistency across heterogeneous clients. Additionally, [15] proposes a method that incorporates global statistics into local training, ensuring that local models remain aligned with the global distribution while preserving personalization.

Collectively, these regularization methods address the inherent challenges of federated learning, such as data heterogeneity, client drift, and communication inefficiency. By incorporating parameter-based regularizers, correcting update directions, aligning prototypes, and leveraging soft labels or statistical information, these techniques significantly enhance the personalization and robustness of local models. This, in turn, improves the overall performance and scalability of federated learning systems, making them more suitable for real-world applications with diverse and distributed data sources.

6.3. Layer-Wise Personalization

Considering the distinct representations of different layers in deep neural networks (DNNs), several personalized federated learning methods have been developed by adopting layer-wise aggregation strategies. These approaches recognize that different layers of a DNN capture varying levels of abstraction, from low-level features in shallow layers to high-level semantic representations in deeper layers. To leverage this hierarchical structure, some methods keep the batch normalization layers personalized and refrain from aggregating them on the server. This strategy helps avoid the drift of local features caused by non-IID (non-independent and identically distributed) data distributions across clients, ensuring that each client retains its unique feature normalization statistics [114,115]. By preserving the local characteristics of batch normalization layers, these methods enhance the personalization of models while still enabling collaborative learning.

In addition to personalizing batch normalization layers, many works focus on aggregating only the shallow layers of DNNs, as these layers typically capture general features that are transferable across clients. These methods assign the same aggregation weights to the shallow layers, facilitating the transfer of general knowledge while allowing deeper layers to remain client-specific for personalized adaptation [116,117]. This approach strikes a balance between global generalization and local personalization, enabling clients to benefit from shared knowledge while maintaining the flexibility to adapt to their unique data distributions.

Recently, more advanced techniques have emerged to address the challenges of layer-wise aggregation in heterogeneous federated learning environments. For instance, pFedLA (personalized Federated Learning with Layer-wise Aggregation) considers the varying impacts of different layers and employs a hypernetwork to generate layer-wise aggregation weights for each client. By dynamically adjusting the aggregation weights based on the specific characteristics of each client's data, pFedLA mitigates knowledge transfer conflicts and enhances the effectiveness of collaborative learning in heterogeneous settings [118]. However, one limitation of pFedLA is the significant training effort required to achieve convergence, which can hinder the adaptive transfer of knowledge among clients.

To address this limitation, KAPC (Knowledge-Adaptive Personalized Collaboration) proposes a more efficient approach by directly learning the layer-wise aggregation weights instead of generating them through a large DNN. This method reduces the computational overhead and accelerates convergence while maintaining the ability to adaptively transfer knowledge across clients. By focusing on the direct optimization of aggregation weights, KAPC demonstrates great effectiveness in achieving personalized federated learning with minimal training effort [119,120].

In summary, layer-wise aggregation has become a key strategy for personalized federated learning, enabling clients to share general knowledge while preserving local adaptations. From personalizing batch normalization layers to dynamically generating aggregation weights using hypernetworks, these methods have significantly advanced the field. However, the computational complexity of some

approaches, such as pFedLA, has highlighted the need for more efficient solutions. KAPC addresses this challenge by directly learning aggregation weights, offering a scalable and effective alternative for personalized federated learning in heterogeneous environments. These advancements underscore the importance of balancing computational efficiency with the need for adaptive and personalized knowledge transfer in federated learning systems.

7. Future Direction

7.1. Data Heterogeneity in Federated Multimodal Learning

Multimodal data has broader applications in real-world scenarios, and multimodal learning is the foundational paradigm for learning from multimodal data [121]. Integrating multimodal learning with federated learning represents one of the key future research directions. Therefore, federated multimodal learning [122,123] is recently proposed, which significantly differs from traditional federated learning, primarily reflected in the following aspects. First, Federated multimodal learning needs to handle multiple types of data simultaneously (such as text, images, audio, and video) [124–126], while traditional federated learning typically focuses on a single data type. The heterogeneity of multimodal data is stronger, making feature extraction and fusion more challenging. Federated multimodal learning requires designing complex model architectures to process data from different modalities and studying cross-modal feature fusion mechanisms (such as attention mechanisms and cross-modal alignment), whereas traditional federated learning often employs a single model structure. The distribution differences of multimodal data across different devices or nodes are more pronounced, and there may be alignment issues between modalities (e.g., mismatched text and images). Federated multimodal learning needs to explore cross-modal data alignment and distribution adaptation techniques, while traditional federated learning mainly focuses on distribution differences within a single modality. The communication and computational costs for multimodal data are higher, necessitating research into efficient modal compression [127,128], selective transmission [129,130], and asynchronous training techniques [28,131,132] in federated multimodal learning, whereas traditional federated learning focuses more on optimizing the communication of gradients or model parameters. In summary, federated multimodal learning faces more challenges in data-heterogeneous scenarios but also offers broader application prospects for cross-modal collaborative learning.

7.2. Data Heterogeneity in Federated Learning with Large Language Models

With the rapid development of Large Language Models (LLMs) based on the transformer architecture, LLMs are widely regarded as a key pathway to achieving general artificial intelligence [133–137]. The advancement of these large models requires vast amounts of data, and federated learning stands out as one of the crucial methods for aggregating data from multiple parties. However, the training of federated large models also faces the challenge of data heterogeneity [138]. In the direction of data heterogeneity, research on federated LLMs significantly differs from traditional federated learning. Federated LLMs typically have a massive number of parameters (e.g., GPT, BERT, etc.), while traditional federated learning often focuses on lightweight models [139,140]. The training and communication costs for are large [141–143], especially for LLMs, making efficient training and deployment in heterogeneous data environments a key challenge. Traditional methods for small models may not be suitable [144], and novel communication reduction methods are required. Federated LLM need to handle more complex heterogeneous data (such as text, images, videos, and other multimodal data), whereas traditional federated learning usually deals with single data types. LLM require stronger cross-domain feature extraction and fusion capabilities to address the diversity of data distributions. Federated LLM place greater emphasis on balancing global generalization and local personalization, while traditional federated learning focuses more on global model performance. LLM need to achieve personalized adaptation through techniques like fine-tuning and prompt learning (Prompt Tuning). Federated LLM have higher demands for communication and computational resources, with research

focusing on model compression, asynchronous training, and other techniques, whereas traditional federated learning primarily concerns simple gradient aggregation and synchronous updates. In summary, research on federated LLM in heterogeneous data scenarios is more challenging but also offers greater potential for cross-domain collaborative learning.

8. Conclusion

Federated learning has made significant progress in addressing data heterogeneity issues, primarily through strategies such as personalized models, adaptive aggregation, and knowledge distillation to tackle the challenges posed by non-independent and identically distributed (non-IID) data. This paper first describes methods for mitigating model drift. Then, it reviews model aggregation techniques, including ensemble distillation and other approaches. Next, it summarizes client selection methods. Following that, it discusses data augmentation techniques, such as sharing generated or public data, to alleviate data heterogeneity. Additionally, it explores personalized methods, such as retaining local parameters or dynamically adjusting layer-wise aggregation weights. Finally, it provides an outlook on future research directions for data heterogeneity in federated learning, including federated multimodal learning and federated large model learning. Future research needs to further explore more efficient and secure heterogeneous data processing mechanisms in federated multimodal learning and federated large model learning to enable the broader practical application of federated learning.

References

1. McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B.A. Communication-Efficient Learning of Deep Networks from Decentralized Data. In Proceedings of the Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS, 2017.
2. Tong Xia, Abhirup Ghosh, X.Q.; Mascolo, C. FLear: Addressing Data Scarcity and Label Skew in Federated Learning via Privacy-preserving Feature Augmentation **2024**.
3. Xu, W.; Wang, H.; Lu, Z.; Hua, C.; Cheng, N.; Guo, S. Mobile Collaborative Learning Over Opportunistic Internet of Vehicles. *IEEE Trans. Mob. Comput.* **2024**, *23*, 3187–3199.
4. Guo, P.; Wang, P.; Zhou, J.; Jiang, S.; Patel, V.M. Multi-Institutional Collaborations for Improving Deep Learning-Based Magnetic Resonance Image Reconstruction Using Federated Learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19–25, 2021, pp. 2423–2432.
5. Xu, A.; Li, W.; Guo, P.; Yang, D.; Roth, H.; Hatamizadeh, A.; Zhao, C.; Xu, D.; Huang, H.; Xu, Z. Closing the Generalization Gap of Cross-silo Federated Medical Image Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18–24, 2022, pp. 20834–20843.
6. Li, Y.; Wang, H.; Xu, W.; Xiao, T.; Liu, H.; Tu, M.; Wang, Y.; Yang, X.; Zhang, R.; Yu, S.; et al. Unleashing the Power of Continual Learning on Non-Centralized Devices: A Survey. *CoRR* **2024**, *abs/2412.13840*.
7. Ramaswamy, S.; Mathews, R.; Rao, K.; Beaufays, F. Federated Learning for Emoji Prediction in a Mobile Keyboard. *CoRR* **2019**, *abs/1906.04329*.
8. Yichen Li, Yijing Shan, Y.L.H.W.W.W.Y.W.R.L. Personalized Federated Recommendation for Cold-Start Users via Adaptive Knowledge Fusion. In Proceedings of the The 34th ACM Web Conference (WWW 2025), Sydney, Australia, Apr 28th - May 2nd, 2025; Singh, A.; Zhu, X.J., Eds., 2025.
9. Liu, Y.; Wang, H.; Wang, S.; He, Z.; Xu, W.; Zhu, J.; Yang, F. Disentangle Estimation of Causal Effects from Cross-Silo Data. In Proceedings of the ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2024, pp. 6290–6294. <https://doi.org/10.1109/ICASSP48485.2024.10446298>.
10. Li, X.; Jiang, M.; Zhang, X.; Kamp, M.; Dou, Q. FedBN: Federated Learning on Non-IID Features via Local Batch Normalization. In Proceedings of the 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3–7, 2021.
11. Zhu, J.; Zheng, H.; Xu, W.; Wang, H.; He, Z.; Liu, Y.; Wang, S.; Sun, Q. Harmonizing Global and Local Class Imbalance for Federated Learning. *IEEE Trans. Mob. Comput.* **2025**, *24*, 1120–1131.

12. Li, Y.; Xu, W.; Qi, Y.; Wang, H.; Li, R.; Guo, S. SR-FDIL: Synergistic Replay for Federated Domain-Incremental Learning. *IEEE Trans. Parallel Distributed Syst.* **2024**, *35*, 1879–1890.
13. Wang, H.; Qu, Z.; Zhou, Q.; Zhang, H.; Luo, B.; Xu, W.; Guo, S.; Li, R. A Comprehensive Survey on Training Acceleration for Large Machine Learning Models in IoT. *IEEE Internet Things J.* **2022**, *9*, 939–963.
14. Hu, M.; Zhou, P.; Yue, Z.; Ling, Z.; Huang, Y.; Li, A.; Liu, Y.; Lian, X.; Chen, M. FedCross: Towards Accurate Federated Learning via Multi-Model Cross-Aggregation. In Proceedings of the IEEE International Conference on Data Engineering (ICDE). IEEE, 2024, pp. 2137–2150.
15. Zhang, J.; Guo, S.; Ma, X.; Wang, H.; Xu, W.; Wu, F. Parameterized Knowledge Transfer for Personalized Federated Learning. In Proceedings of the Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual, pp. 10092–10104.
16. Kang, H.; Cha, S.; Shin, J.; Lee, J.; Kang, J. NeFL: Nested Federated Learning for Heterogeneous Clients. *CoRR* **2023**, *abs/2308.07761*. <https://doi.org/10.48550/ARXIV.2308.07761>.
17. Wang, H.; Xu, H.; Li, Y.; Xu, Y.; Li, R.; Zhang, T. FedCDA: Federated Learning with Cross-rounds Divergence-aware Aggregation. In Proceedings of the The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024.
18. Ye, M.; Fang, X.; Du, B.; Yuen, P.C.; Tao, D. Heterogeneous Federated Learning: State-of-the-art and Research Challenges. *ACM Comput. Surv.* **2024**, *56*, 79:1–79:44.
19. Xu, C.; Qu, Y.; Xiang, Y.; Gao, L. Asynchronous federated learning on heterogeneous devices: A survey. *Comput. Sci. Rev.* **2023**, *50*, 100595.
20. Qin, L.; Zhu, T.; Zhou, W.; Yu, P.S. Knowledge Distillation in Federated Learning: a Survey on Long Lasting Challenges and New Solutions. *CoRR* **2024**, *abs/2406.10861*. <https://doi.org/10.48550/ARXIV.2406.10861>.
21. Zhang, J.; Qu, Z.; Chen, C.; Wang, H.; Zhan, Y.; Ye, B.; Guo, S. Edge Learning: The Enabling Technology for Distributed Big Data Analytics in the Edge. *ACM Comput. Surv.* **2022**, *54*, 151:1–151:36.
22. Wang, H.; Li, Y.; Xu, W.; Li, R.; Zhan, Y.; Zeng, Z. DaFKD: Domain-aware Federated Knowledge Distillation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023. IEEE, 2023, pp. 20412–20421.
23. Hu, M.; Cao, Y.; Li, A.; Li, Z.; Liu, C.; Li, T.; Chen, M.; Liu, Y. FedMut: Generalized Federated Learning via Stochastic Mutation. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2024, Vol. 38, pp. 12528–12537.
24. Li, Y.; Li, Q.; Wang, H.; Li, R.; Zhong, W.; Zhang, G. Towards Efficient Replay in Federated Incremental Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024, Seattle, WA, USA, June 16-22, 2024. IEEE, 2024, pp. 12820–12829.
25. Li, Y.; Wang, Y.; Xiao, T.; Wang, H.; Qi, Y.; Li, R. Rehearsal-Free Continual Federated Learning with Synergistic Regularization. *CoRR* **2024**, *abs/2412.13779*. <https://doi.org/10.48550/ARXIV.2412.13779>.
26. Li, T.; Sahu, A.K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; Smith, V. Federated Optimization in Heterogeneous Networks. In Proceedings of the Proceedings of Machine Learning and Systems 2020, MLSys 2020, Austin, TX, USA, March 2-4, 2020, 2020.
27. Acar, D.A.E.; Zhao, Y.; Navarro, R.M.; Mattina, M.; Whatmough, P.N.; Saligrama, V. Federated Learning Based on Dynamic Regularization. In Proceedings of the 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021, 2021.
28. Wang, H.; Qu, Z.; Guo, S.; Wang, N.; Li, R.; Zhuang, W. LOSP: Overlap Synchronization Parallel With Local Compensation for Fast Distributed Training. *IEEE J. Sel. Areas Commun.* **2021**, *39*, 2541–2557.
29. Gong, Y.; Li, Y.; Freris, N.M. FedADMM: A Robust Federated Deep Learning Framework with Adaptivity to System Heterogeneity. In Proceedings of the 38th IEEE International Conference on Data Engineering, ICDE 2022, Kuala Lumpur, Malaysia, May 9-12, 2022, 2022, pp. 2575–2587.
30. Sun, Y.; Shen, L.; Huang, T.; Ding, L.; Tao, D. FedSpeed: Larger Local Interval, Less Communication Round, and Higher Generalization Accuracy. In Proceedings of the The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023.
31. Wang, J.; Liu, Q.; Liang, H.; Joshi, G.; Poor, H.V. Tackling the Objective Inconsistency Problem in Heterogeneous Federated Optimization. In Proceedings of the Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual, 2020.

32. Charles, Z.; Konečný, J. Convergence and Accuracy Trade-Offs in Federated Learning and Meta-Learning. In Proceedings of the The 24th International Conference on Artificial Intelligence and Statistics, AISTATS 2021, April 13-15, 2021, Virtual Event, 2021, pp. 2575–2583.
33. Malinovskiy, G.; Kovalev, D.; Gasanov, E.; Condat, L.; Richtárik, P. From Local SGD to Local Fixed-Point Methods for Federated Learning. In Proceedings of the Proceedings of the 37th International Conference on Machine Learning, ICML, 13-18 July, Virtual Event, 2020, pp. 6692–6701.
34. Jhunhunwala, D.; Wang, S.; Joshi, G. FedExP: Speeding Up Federated Averaging via Extrapolation. In Proceedings of the The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023.
35. Karimireddy, S.P.; Kale, S.; Mohri, M.; Reddi, S.J.; Stich, S.U.; Suresh, A.T. SCAFFOLD: Stochastic Controlled Averaging for Federated Learning. In Proceedings of the Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event, pp. 5132–5143.
36. Wang, H.; Xu, W.; Fan, Y.; Li, R.; Zhou, P. AOCC-FL: Federated Learning with Aligned Overlapping via Calibrated Compensation. In Proceedings of the IEEE INFOCOM 2023 - IEEE Conference on Computer Communications, New York City, NY, USA, May 17-20, 2023. IEEE, 2023, pp. 1–10.
37. Wu, F.; Guo, S.; Wang, H.; Zhang, H.; Qu, Z.; Zhang, J.; Liu, Z. From Deterioration to Acceleration: A Calibration Approach to Rehabilitating Step Asynchronism in Federated Optimization. *IEEE Trans. Parallel Distributed Syst.* **2023**, *34*, 1548–1559.
38. Hu, M.; Xia, Z.; Yan, D.; Yue, Z.; Xia, J.; Huang, Y.; Liu, Y.; Chen, M. GitFL: Uncertainty-Aware Real-Time Asynchronous Federated Learning Using Version Control. In Proceedings of the In Proceedings of IEEE Real-Time Systems Symposium (RTSS). IEEE, 2023, pp. 145–157.
39. Gao, L.; Fu, H.; Li, L.; Chen, Y.; Xu, M.; Xu, C. FedDC: Federated Learning with Non-IID Data via Local Drift Decoupling and Correction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022, pp. 10102–10111.
40. Luo, K.; Li, X.; Lan, Y.; Gao, M. GradMA: A Gradient-Memory-based Accelerated Federated Learning with Alleviated Catastrophic Forgetting. *The IEEE/CVF Conference on Computer Vision and Pattern Recognition, (CVPR 2023), Vancouver, Canada* **2023**.
41. Xu, J.; Wang, S.; Wang, L.; Yao, A.C. FedCM: Federated Learning with Client-level Momentum. *CoRR* **2021**, *abs/2106.10874*.
42. Wang, J.; Tanti, V.; Ballas, N.; Rabbat, M.G. SlowMo: Improving Communication-Efficient Distributed SGD with Slow Momentum. In Proceedings of the 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020, 2020.
43. Kim, G.; Kim, J.; Han, B. Communication-Efficient Federated Learning with Acceleration of Global Momentum. *CoRR* **2022**, *abs/2201.03172*.
44. Reddi, S.J.; Charles, Z.; Zaheer, M.; Garrett, Z.; Rush, K.; Konečný, J.; Kumar, S.; McMahan, H.B. Adaptive Federated Optimization. In Proceedings of the 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021.
45. Li, X.; Zhan, D. FedRS: Federated Learning with Restricted Softmax for Label Distribution Non-IID Data. In Proceedings of the KDD '21: The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, Singapore, August 14-18, 2021, 2021, pp. 995–1005.
46. Li, Q.; He, B.; Song, D. Model-Contrastive Federated Learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021, pp. 10713–10722.
47. Kim, J.; Kim, G.; Han, B. Multi-Level Branched Regularization for Federated Learning. In Proceedings of the International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA, pp. 11058–11073.
48. Zhang, J.; Li, Z.; Li, B.; Xu, J.; Wu, S.; Ding, S.; Wu, C. Federated Learning with Label Distribution Skew via Logits Calibration. In Proceedings of the International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA, pp. 26311–26329.
49. Wang, H.; Zheng, P.; Han, X.; Xu, W.; Li, R.; Zhang, T. FedNLR: Federated Learning with Neuron-wise Learning Rates. In Proceedings of the Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2024, Barcelona, Spain, August 25-29, 2024. ACM, 2024, pp. 3069–3080.

50. Li, S.; Zhou, T.; Tian, X.; Tao, D. Learning to Collaborate in Decentralized Learning of Personalized Models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022, 2022, pp. 9756–9765.
51. Rehman, Y.A.U.; Gao, Y.; de Gusmão, P.P.B.; Alibeigi, M.; Shen, J.; Lane, N.D. L-DAWA: Layer-wise Divergence Aware Weight Aggregation in Federated Self-Supervised Visual Representation Learning. *CoRR* **2023**, *abs/2307.07393*.
52. Zhu, J.; Li, Y.; Wang, H.; Qi, Y.; Li, R. Hypernetwork-driven centralized contrastive learning for federated graph classification. *World Wide Web (WWW)* **2024**, *27*, 56.
53. Xia, Y.; Yang, D.; Li, W.; Myronenko, A.; Xu, D.; Obinata, H.; Mori, H.; An, P.; Harmon, S.A.; Turkbey, E.; et al. Auto-FedAvg: Learnable Federated Averaging for Multi-Institutional Medical Image Segmentation. *CoRR* **2021**, *abs/2104.10195*.
54. Li, Z.; Lin, T.; Shang, X.; Wu, C. Revisiting Weighted Aggregation in Federated Learning with Neural Networks. In Proceedings of the International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA, 2023, pp. 19767–19788.
55. Yu, F.; Zhang, W.; Qin, Z.; Xu, Z.; Wang, D.; Liu, C.; Tian, Z.; Chen, X. Fed2: Feature-Aligned Federated Learning. In Proceedings of the KDD '21: The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, Singapore, August 14-18, 2021, 2021, pp. 2066–2074.
56. Singh, S.P.; Jaggi, M. Model Fusion via Optimal Transport. In Proceedings of the Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual, 2020.
57. Li, X.; Xu, Y.; Song, S.; Li, B.; Li, Y.; Shao, Y.; Zhan, D. Federated Learning with Position-Aware Neurons. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022, pp. 10072–10081.
58. Wang, H.; Jia, Y.; Zhang, M.; Hu, Q.; Ren, H.; Sun, P.; Wen, Y.; Zhang, T. FedDSE: Distribution-aware Sub-model Extraction for Federated Learning over Resource-constrained Devices. In Proceedings of the Proceedings of the ACM on Web Conference 2024, WWW 2024, Singapore, May 13-17, 2024. ACM, 2024, pp. 2902–2913.
59. Wang, H.; Yurochkin, M.; Sun, Y.; Papailiopoulos, D.S.; Khazaeni, Y. Federated Learning with Matched Averaging. In Proceedings of the 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020, 2020.
60. Yurochkin, M.; Agarwal, M.; Ghosh, S.; Greenewald, K.H.; Hoang, T.N.; Khazaeni, Y. Bayesian Nonparametric Federated Learning of Neural Networks. In Proceedings of the ICML. PMLR, 2019, Vol. 97, *Proceedings of Machine Learning Research*, pp. 7252–7261.
61. Liu, C.; Lou, C.; Wang, R.; Xi, A.Y.; Shen, L.; Yan, J. Deep Neural Network Fusion via Graph Matching with Applications to Model Ensemble and Federated Learning. In Proceedings of the International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA, 2022, pp. 13857–13869.
62. Hinton, G.; Vinyals, O.; Dean, J. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531* **2015**.
63. Xu, W.; Wan, H.; Wang, H.; Cheng, N.; Chen, Q.; Zhou, H.; Guo, S. Fast Packet Loss Inferring via Personalized Simulation-Reality Distillation. *IEEE Trans. Mob. Comput.* **2024**, *23*, 3696–3706.
64. Yang, C.; Xie, L.; Qiao, S.; Yuille, A.L. Training deep neural networks in generations: A more tolerant teacher educates better students. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, AAAI, 2019, Vol. 33, pp. 5628–5635.
65. Phuong, M.; Lampert, C.H. Distillation-based training for multi-exit architectures. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 1355–1364.
66. Huo, F.; Xu, W.; Guo, J.; Wang, H.; Guo, S. C²KD: Bridging the Modality Gap for Cross-Modal Knowledge Distillation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024, Seattle, WA, USA, June 16-22, 2024. IEEE, 2024, pp. 16006–16015.
67. Wu, G.; Gong, S. Peer Collaborative Learning for Online Knowledge Distillation. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, AAAI, 2021.
68. Guo, Q.; Wang, X.; Wu, Y.; Yu, Z.; Liang, D.; Hu, X.; Luo, P. Online knowledge distillation via collaborative learning. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2020, pp. 11020–11029.

69. Wang, Y.; Li, Y.; Wang, H.; Zhao, L.; Zhang, X. Better Knowledge Enhancement for Privacy-Preserving Cross-Project Defect Prediction. *CoRR* **2024**, *abs/2412.17317*. <https://doi.org/10.48550/ARXIV.2412.17317>.
70. Bistriz, I.; Mann, A.; Bambos, N. Distributed Distillation for On-Device Learning. In Proceedings of the Proceedings of Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems, NeurIPS, 2020.
71. Lin, T.; Kong, L.; Stich, S.U.; Jaggi, M. Ensemble distillation for robust model fusion in federated learning. *Advances in Neural Information Processing Systems* **2020**, *33*, 2351–2363.
72. Chen, H.; Chao, W. FedBE: Making Bayesian Model Ensemble Applicable to Federated Learning. In Proceedings of the 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021.
73. Zhu, Z.; Hong, J.; Zhou, J. Data-free knowledge distillation for heterogeneous federated learning. In Proceedings of the International Conference on Machine Learning. PMLR, 2021, pp. 12878–12889.
74. Zhang, L.; Shen, L.; Ding, L.; Tao, D.; Duan, L. Fine-tuning Global Model via Data-Free Knowledge Distillation for Non-IID Federated Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022, pp. 10164–10173.
75. Wu, F.; Guo, S.; Qu, Z.; He, S.; Liu, Z.; Gao, J. Anchor Sampling for Federated Learning with Partial Client Participation. In Proceedings of the International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA; Krause, A.; Brunskill, E.; Cho, K.; Engelhardt, B.; Sabato, S.; Scarlett, J., Eds. PMLR, 2023, Vol. 202, *Proceedings of Machine Learning Research*, pp. 37379–37416.
76. Fu, L.; Zhang, H.; Gao, G.; Wang, H.; Zhang, M.; Liu, X. Client Selection in Federated Learning: Principles, Challenges, and Opportunities. *arXiv preprint arXiv:2211.01549* **2022**.
77. Németh, G.D.; Lozano, M.Á.; Quadrianto, N.; Oliver, N. A Snapshot of the Frontiers of Client Selection in Federated Learning. *arXiv preprint arXiv:2210.04607* **2022**.
78. Zhao, J.; Zhang, Y.; Li, R.; Li, Y.; Wang, H.; Yi, X.; Deng, Z. XFed: Improving Explainability in Federated Learning by Intersection Over Union Ratio Extended Client Selection. In Proceedings of the ECAI 2023 - 26th European Conference on Artificial Intelligence, September 30 - October 4, 2023, Kraków, Poland - Including 12th Conference on Prestigious Applications of Intelligent Systems (PAIS 2023). IOS Press, 2023, Vol. 372, *Frontiers in Artificial Intelligence and Applications*, pp. 3099–3106.
79. Nishio, T.; Yonetani, R. Client selection for federated learning with heterogeneous resources in mobile edge. In Proceedings of the ICC 2019-2019 IEEE international conference on communications (ICC). IEEE, 2019, pp. 1–7.
80. Mohammed, I.; Tabatabai, S.; Al-Fuqaha, A.; El Bouanani, F.; Qadir, J.; Qolomany, B.; Guizani, M. Budgeted online selection of candidate IoT clients to participate in federated learning. *IEEE Internet of Things Journal* **2020**, *8*, 5938–5952.
81. Goetz, J.; Malik, K.; Bui, D.; Moon, S.; Liu, H.; Kumar, A. Active federated learning. *arXiv preprint arXiv:1909.12641* **2019**.
82. Li, T.; Sanjabi, M.; Beirami, A.; Smith, V. Fair resource allocation in federated learning. *arXiv preprint arXiv:1905.10497* **2019**.
83. Wang, H.; Li, R.; Li, C.; Zhou, P.; Li, Y.; Xu, W.; Guo, S. Gradient Scheduling With Global Momentum for Asynchronous Federated Learning in Edge Environment. *IEEE Internet Things J.* **2022**, *9*, 18817–18828.
84. Lai, F.; Zhu, X.; Madhyastha, H.V.; Chowdhury, M. Oort: Efficient Federated Learning via Guided Participant Selection. In Proceedings of the OSDI, 2021, pp. 19–35.
85. Katharopoulos, A.; Fleuret, F. Not all samples are created equal: Deep learning with importance sampling. In Proceedings of the International conference on machine learning. PMLR, 2018, pp. 2525–2534.
86. Huang, T.; Lin, W.; Shen, L.; Li, K.; Zomaya, A.Y. Stochastic client selection for federated learning with volatile clients. *IEEE Internet of Things Journal* **2022**, *9*, 20055–20070.
87. Dennis, D.K.; Li, T.; Smith, V. Heterogeneity for the win: One-shot federated clustering. In Proceedings of the International Conference on Machine Learning. PMLR, 2021, pp. 2611–2620.
88. Cho, Y.J.; Wang, J.; Joshi, G. Client selection in federated learning: Convergence analysis and power-of-choice selection strategies. *arXiv preprint arXiv:2010.01243* **2020**.
89. Jiang, A.H.; Wong, D.L.K.; Zhou, G.; Andersen, D.G.; Dean, J.; Ganger, G.R.; Joshi, G.; Kaminsky, M.; Kozuch, M.; Lipton, Z.C.; et al. Accelerating deep learning by focusing on the biggest losers. *arXiv preprint arXiv:1910.00762* **2019**.

90. Shah, V.; Wu, X.; Sanghavi, S. Choosing the sample with lowest loss makes sgd robust. In Proceedings of the International Conference on Artificial Intelligence and Statistics. PMLR, 2020, pp. 2120–2130.
91. Balakrishnan, R.; Li, T.; Zhou, T.; Himayat, N.; Smith, V.; Bilmes, J. Diverse client selection for federated learning via submodular maximization. In Proceedings of the International Conference on Learning Representations.
92. Zhao, J.; Li, R.; Wang, H.; Xu, Z. HotFed: Hot Start through Self-Supervised Learning in Federated Learning. In Proceedings of the 2021 IEEE 23rd Int Conf on High Performance Computing & Communications; 7th Int Conf on Data Science & Systems; 19th Int Conf on Smart City; 7th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys), Haikou, Hainan, China, December 20-22, 2021. IEEE, 2021, pp. 149–156.
93. Huang, W.; Ye, M.; Shi, Z.; Du, B. Generalizable Heterogeneous Federated Cross-Correlation and Instance Similarity Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 712–728.
94. Xia, T.; Ghosh, A.; Qiu, X.; Mascolo, C. FLear: Addressing Data Scarcity and Label Skew in Federated Learning via Privacy-preserving Feature Augmentation. In Proceedings of the Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2024, Barcelona, Spain, August 25-29, 2024; Baeza-Yates, R.; Bonchi, F., Eds. ACM, 2024, pp. 3484–3494.
95. Yoon, T.; Shin, S.; Hwang, S.J.; Yang, E. FedMix: Approximation of Mixup under Mean Augmented Federated Learning. In Proceedings of the 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021, 2021.
96. Chen, H.; Chao, W. On Bridging Generic and Personalized Federated Learning for Image Classification. In Proceedings of the The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022.
97. Wang, H.; Guo, S.; Li, R. OSP: Overlapping Computation and Communication in Parameter Server for Fast Machine Learning. In Proceedings of the Proceedings of the 48th International Conference on Parallel Processing, ICPP 2019, Kyoto, Japan, August 05-08, 2019. ACM, 2019, pp. 82:1–82:10.
98. Oh, J.; Kim, S.; Yun, S. FedBABU: Toward Enhanced Representation for Federated Image Classification. In Proceedings of the The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022.
99. Zhu, G.; Liu, X.; Tang, S.; Niu, J. Aligning Before Aggregating: Enabling Communication Efficient Cross-Domain Federated Learning via Consistent Feature Extraction. *IEEE Trans. Mob. Comput.* **2024**, *23*, 5880–5896.
100. Shen, Y.; Zhou, Y.; Yu, L. CD²-pFed: Cyclic Distillation-guided Channel Decoupling for Model Personalization in Federated Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022. IEEE, 2022, pp. 10031–10040.
101. Collins, L.; Hassani, H.; Mokhtari, A.; Shakkottai, S. Exploiting Shared Representations for Personalized Federated Learning. In Proceedings of the Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event, pp. 2089–2099.
102. Marfoq, O.; Neglia, G.; Vidal, R.; Kameni, L. Personalized Federated Learning through Local Memorization. In Proceedings of the International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA, pp. 15070–15092.
103. Liang, P.P.; Liu, T.; Liu, Z.; Salakhutdinov, R.; Morency, L. Think Locally, Act Globally: Federated Learning with Local and Global Representations. *CoRR* **2020**, *abs/2001.01523*.
104. Yue Tan, Guodong Long, L.L.T.Z.J.J. FedProto: Federated Prototype Learning over Heterogeneous Devices. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, AAAI, 2022.
105. Jian Xu, Xinyi Tong, S.L.H. Personalized Federated Learning with Feature Alignment and Classifier Collaboration. In Proceedings of the The 11th International Conference on Learning Representations, ICLR 2023.
106. Acar, D.A.E.; Zhao, Y.; Navarro, R.M.; Mattina, M.; Whatmough, P.N.; Saligrama, V. Federated learning based on dynamic regularization. *arXiv preprint arXiv:2111.04263* **2021**.
107. Dinh, C.T.; Tran, N.H.; Nguyen, T.D. Personalized Federated Learning with Moreau Envelopes. In Proceedings of the Proceedings of the International Conference on Neural Information Processing Systems (NIPS), 2020, pp. 21394–21405.

108. Karimireddy, S.P.; Kale, S.; Mohri, M.; Reddi, S.; Stich, S.; Suresh, A.T. SCAFFOLD: Stochastic Controlled Averaging for Federated Learning. In Proceedings of the Proceedings of the International Conference on Machine Learning (ICML), 2020, pp. 5132–5143.
109. Zhang, M.; Sapra, K.; Fidler, S.; Yeung, S.; Alvarez, J.M. Personalized federated learning with first order model optimization. *arXiv preprint arXiv:2012.08565* **2020**.
110. Tan, Y.; Long, G.; Liu, L.; Zhou, T.; Lu, Q.; Jiang, J.; Zhang, C. FedProto: Federated Prototype Learning across Heterogeneous Clients. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2022, pp. 8432–8440.
111. Xu, J.; Tong, X.; Huang, S.L. Personalized federated learning with feature alignment and classifier collaboration. *arXiv preprint arXiv:2306.11867* **2023**.
112. Jin, H.; Bai, D.; Yao, D.; Dai, Y.; Gu, L.; Yu, C.; Sun, L. Personalized Edge Intelligence via Federated Self-Knowledge Distillation. *IEEE Transactions on Parallel and Distributed Systems* **2023**, *34*, 567–580.
113. Mendieta, M.; Yang, T.; Wang, P.; Lee, M.; Ding, Z.; Chen, C. Local Learning Matters: Rethinking Data Heterogeneity in Federated Learning. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 8387–8396.
114. Li, X.; Jiang, M.; Zhang, X.; Kamp, M.; Dou, Q. Fedbn: Federated learning on non-iid features via local batch normalization. *arXiv preprint arXiv:2102.07623* **2021**.
115. Mills, J.; Hu, J.; Min, G. Multi-Task Federated Learning for Personalised Deep Neural Networks in Edge Computing. *IEEE Transactions on Parallel and Distributed Systems* **2022**, *33*, 630–641.
116. Oh, J.; Kim, S.; Yun, S.Y. FedBABU: Towards Enhanced Representation for Federated Image Classification. *arXiv preprint arXiv:2106.06042* **2021**.
117. Li, A.; Sun, J.; Li, P.; Pu, Y.; Li, H.; Chen, Y. Hermes: An Efficient Federated Learning Framework for Heterogeneous Mobile Clients. In Proceedings of the Proceedings of the Annual International Conference on Mobile Computing and Networking (MobiCom), 2021, p. 420–437.
118. Ma, X.; Zhang, J.; Guo, S.; Xu, W. Layer-wised Model Aggregation for Personalized Federated Learning. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 10082–10091.
119. Zhi, M.; Bi, Y.; Xu, W.; Wang, H.; Xiang, T. Knowledge-Aware Parameter Coaching for Personalized Federated Learning. In Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2024, February 20–27, 2024, Vancouver, Canada; Wooldridge, M.J.; Dy, J.G.; Natarajan, S., Eds. AAAI Press, 2024, pp. 17069–17077.
120. Zhi, M.; Bi, Y.; Cai, L.; Xu, W.; Wang, H.; Xiang, T.; He, Q. Knowledge-Aware Parameter Coaching for Communication-Efficient Personalized Federated Learning in Mobile Edge Computing. *IEEE Trans. Mob. Comput.* **2025**, *24*, 321–337.
121. Fan, Y.; Xu, W.; Wang, H.; Guo, S. Cross-modal Representation Flattening for Multi-modal Domain Generalization. In Proceedings of the Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10–15, 2024; Globersons, A.; Mackey, L.; Belgrave, D.; Fan, A.; Paquet, U.; Tomczak, J.M.; Zhang, C., Eds., 2024.
122. Fan, Y.; Xu, W.; Wang, H.; Huo, F.; Chen, J.; Guo, S. Overcome Modal Bias in Multi-modal Federated Learning via Balanced Modality Selection. In Proceedings of the Computer Vision - ECCV 2024 - 18th European Conference, Milan, Italy, September 29–October 4, 2024, Proceedings, Part LXXXI; Leonardis, A.; Ricci, E.; Roth, S.; Russakovsky, O.; Sattler, T.; Varol, G., Eds. Springer, 2024, Vol. 15139, *Lecture Notes in Computer Science*, pp. 178–195.
123. Le, H.Q.; Nguyen, M.N.H.; Thwal, C.M.; Qiao, Y.; Zhang, C.; Hong, C.S. FedMEKT: Distillation-based embedding knowledge transfer for multimodal federated learning. *Neural Networks* **2025**, *183*, 107017.
124. Fan, Y.; Xu, W.; Wang, H.; Liu, J.; Guo, S. Detached and Interactive Multimodal Learning. In Proceedings of the Proceedings of the 32nd ACM International Conference on Multimedia, MM 2024, Melbourne, VIC, Australia, 28 October 2024 - 1 November 2024; Cai, J.; Kankanhalli, M.S.; Prabhakaran, B.; Boll, S.; Subramanian, R.; Zheng, L.; Singh, V.K.; César, P.; Xie, L.; Xu, D., Eds. ACM, 2024, pp. 5470–5478.
125. J. Li, H. Xu, S.Z.J.H.W. Multilevel Semantic-Aware Model for AI-Generated Video Quality Assessment. *2025 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2025), Hyderabad, India, April 06 to April 11, 2025*. **2025**.

126. Fan, Y.; Xu, W.; Wang, H.; Wang, J.; Guo, S. PMR: Prototypical Modal Rebalance for Multimodal Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023. IEEE, 2023, pp. 20029–20038.
127. Wu, F.; He, S.; Guo, S.; Qu, Z.; Wang, H.; Zhuang, W.; Zhang, J. Sign bit is enough: a learning synchronization framework for multi-hop all-reduce with ultimate compression. In Proceedings of the DAC '22: 59th ACM/IEEE Design Automation Conference, San Francisco, California, USA, July 10 - 14, 2022; Oshana, R., Ed. ACM, 2022, pp. 193–198.
128. Pan, Z.; Li, Y.; Guan, Z.; Liang, M.; Li, A.; Wang, J.; Kou, F. RFCSC: Communication efficient reinforcement federated learning with dynamic client selection and adaptive gradient compression. *Neurocomputing* **2025**, *612*, 128672.
129. Herzog, A.; Southam, R.; Belarbi, O.; Anwar, S.; Bullo, M.; Carnelli, P.; Khan, A. Selective Updates and Adaptive Masking for Communication-Efficient Federated Learning. *IEEE Trans. Green Commun. Netw.* **2024**, *8*, 852–864.
130. Wang, H.; Qu, Z.; Guo, S.; Gao, X.; Li, R.; Ye, B. Intermittent Pulling With Local Compensation for Communication-Efficient Distributed Learning. *IEEE Trans. Emerg. Top. Comput.* **2022**, *10*, 779–791.
131. Qu, Z.; Guo, S.; Wang, H.; Ye, B.; Wang, Y.; Zomaya, A.Y.; Tang, B. Partial Synchronization to Accelerate Federated Learning Over Relay-Assisted Edge Networks. *IEEE Trans. Mob. Comput.* **2022**, *21*, 4502–4516.
132. Hou, Y.; Li, H.; Guo, Z.; Wu, W.; Liu, R.; You, L. FedIBD: a federated learning framework in asynchronous mode for imbalanced data. *Appl. Intell.* **2025**, *55*, 122.
133. Xu, W., C.J.Z.P.Y.X.T.T.Z.W.W.Q.W.H.F.Y.S.Q.; Shen, X. Deploying Foundation Model Powered Agent Services: A Survey. In Proceedings of the arXiv, 2024.
134. Bao, G.; Zhang, H.; Wang, C.; Yang, L.; Zhang, Y. How Likely Do LLMs with CoT Mimic Human Reasoning? In Proceedings of the Proceedings of the 31st International Conference on Computational Linguistics, COLING 2025, Abu Dhabi, UAE, January 19-24, 2025; Rambow, O.; Wanner, L.; Apidianaki, M.; Al-Khalifa, H.; Eugenio, B.D.; Schockaert, S., Eds. Association for Computational Linguistics, 2025, pp. 7831–7850.
135. Zhang, X.; Du, C.; Pang, T.; Liu, Q.; Gao, W.; Lin, M. Chain of Preference Optimization: Improving Chain-of-Thought Reasoning in LLMs. In Proceedings of the Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024; Globersons, A.; Mackey, L.; Belgrave, D.; Fan, A.; Paquet, U.; Tomczak, J.M.; Zhang, C., Eds., 2024.
136. Huo, F.; Xu, W.; Zhang, Z.; Wang, H.; Chen, Z.; Zhao, P. Self-Introspective Decoding: Alleviating Hallucinations for Large Vision-Language Models. *CoRR* **2024**, *abs/2408.02032*, [2408.02032]. <https://doi.org/10.48550/ARXIV.2408.02032>.
137. Jiang, Y.; Wang, H.; Xie, L.; Zhao, H.; Chao, Z.; Qian, H.; Lui, J.C.S. D-LLM: A Token Adaptive Computing Resource Allocation Strategy for Large Language Models. In Proceedings of the Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024; Globersons, A.; Mackey, L.; Belgrave, D.; Fan, A.; Paquet, U.; Tomczak, J.M.; Zhang, C., Eds., 2024.
138. Feijie Wu, Xiaozhe Liu, H.W.X.W.L.S.J.G. Towards Federated RLHF with Aggregated Client Preference for LLMs. In Proceedings of the in Proc. of the International Conference on Learning Representations (ICLR'25), 2025.
139. Wu, F.; Li, Z.; Li, Y.; Ding, B.; Gao, J. FedBiOT: LLM Local Fine-tuning in Federated Learning without Full Model. In Proceedings of the Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2024, Barcelona, Spain, August 25-29, 2024; Baeza-Yates, R.; Bonchi, F., Eds. ACM, 2024, pp. 3345–3355.
140. Ye, R.; Ge, R.; Zhu, X.; Chai, J.; Du, Y.; Liu, Y.; Wang, Y.; Chen, S. FedLLM-Bench: Realistic Benchmarks for Federated Learning of Large Language Models. In Proceedings of the Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024; Globersons, A.; Mackey, L.; Belgrave, D.; Fan, A.; Paquet, U.; Tomczak, J.M.; Zhang, C., Eds., 2024.
141. Li, S.; Xu, W.; Wang, H.; Tang, X.; Qi, Y.; Xu, S.; Luo, W.; Li, Y.; He, X.; Li, R. FedBAT: Communication-Efficient Federated Learning via Learnable Binarization. In Proceedings of the Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024. OpenReview.net, 2024.

142. Xu, Y.; Jiang, Z.; Xu, H.; Wang, Z.; Qian, C.; Qiao, C. Federated Learning With Client Selection and Gradient Compression in Heterogeneous Edge Systems. *IEEE Trans. Mob. Comput.* **2024**, *23*, 5446–5461.
143. Wang, H.; Guo, S.; Qu, Z.; Li, R.; Liu, Z. Error-Compensated Sparsification for Communication-Efficient Decentralized Training in Edge Environment. *IEEE Trans. Parallel Distributed Syst.* **2022**, *33*, 14–25.
144. Li, S.; Cheng, Y.; Wang, H.; Tang, X.; Xu, S.; Luo, W.; Li, Y.; Liu, D.; He, X.; Li, R. Masked Random Noise for Communication-Efficient Federated Learning. In Proceedings of the Proceedings of the 32nd ACM International Conference on Multimedia, MM 2024, Melbourne, VIC, Australia, 28 October 2024 - 1 November 2024; Cai, J.; Kankanhalli, M.S.; Prabhakaran, B.; Boll, S.; Subramanian, R.; Zheng, L.; Singh, V.K.; César, P.; Xie, L.; Xu, D., Eds. ACM, 2024, pp. 3686–3694.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.