International Journal of Machine Learning and Cybernetics
https://doi.org/10.1007/s13042-024-02119-1

ORIGINAL ARTICLE

Emerging trends in federated learning: from model fusion to federated X learning

Shaoxiong Ji[1] · Yue Tan[2] · Teemu Saravirta[1] · Zhiqin Yang[4] · Yixin Liu[5] · Lauri Vasankari[3] · Shirui Pan[6] · Guodong Long[2] · Anwar Walid[7,8]

Received: 24 November 2023 / Accepted: 25 February 2024

Abstract

Federated learning is a new learning paradigm that decouples data collection and model training via multi-party computation and model aggregation. As a flexible learning setting, federated learning has the potential to integrate with other learning frameworks. We conduct a focused survey of federated learning in conjunction with other learning algorithms. Specifically, we explore various learning algorithms to improve the vanilla federated averaging algorithm and review model fusion methods such as adaptive aggregation, regularization, clustered methods, and Bayesian methods. Following the emerging trends, we also discuss federated learning in the intersection with other learning paradigms, termed federated X learning, where X includes multitask learning, meta-learning, transfer learning, unsupervised learning, and reinforcement learning. In addition to reviewing state-of-the-art studies, this paper also identifies key challenges and applications in this field, while also highlighting promising future directions.

Keywords Federated learning · Model fusion · Learning algorithms

1 Introduction

Vast quantities of data are required for state-of-the-art machine learning algorithms. However, the data cannot be uploaded to a central server or cloud due to sheer volume, privacy, or legislative reasons. Federated learning (FL) [1], also known as collaborative learning, has been the subject of many studies. FL adopts a distributed machine learning architecture with a central server for model aggregation, where clients themselves update the machine learning model. Clients can maintain ownership of their data, i.e., upload only the updated model to the central server and not expose any of their private data.

The federated learning paradigm addresses several challenges. The first challenge is privacy. Local data ownership inherits a basic level of privacy. However, federated learning systems can be vulnerable to adversarial attacks, such as backdoor attack [2], model poisoning [3], and data

✉ Shaoxiong Ji
shaoxiong.ji@helsinki.fi

✉ Yixin Liu
yixin.liu@monash.edu

Yue Tan
yue.tan@student.uts.edu.au

Teemu Saravirta
teemu.saravirta@helsinki.fi

Zhiqin Yang
yangzqccc@buaa.edu.cn

Lauri Vasankari
lauri.vasankari@aalto.fi

Shirui Pan
s.pan@griffith.edu.au

Guodong Long
guodong.long@uts.edu.au

Anwar Walid
aie13@columbia.edu

[1] University of Helsinki, Helsinki, Finland
[2] University of Technology Sydney, Ultimo, Australia
[3] Aalto University, Espoo, Finland
[4] Beihang University, Beijing, China
[5] Monash University, Melbourne, Australia
[6] Griffith University, Gold Coast, Australia
[7] Amazon, New York, USA
[8] Columbia University, New York, USA

Published online: 02 April 2024

poisoning [4]. The second challenge is the communication cost for model uploading and downloading. Improving communication efficiency is a critical issue [5–7]. Centralized network architecture also makes the central server suffer from a heavy communication workload, calling for a decentralized server architecture [8]. The third challenge is statistical heterogeneity. Aggregating clients' models together can result in a non-optimal combined model as client data is often non-IID (independent and identically distributed). Statistical heterogeneity introduces a degree of uncertainty into the learning model. Therefore, adopting the right aggregation and learning techniques is vital for robust implementation. This survey gives a particular focus on how different federated learning solutions address statistical heterogeneity.

The robust model aggregation has recently garnered considerable attention. Traditionally, client contributions are weighted according to their sample quantity, while recent research has introduced adaptive weighting [9, 10], attentive aggregation [11], regularization [12], clustering [13], and Bayesian methods [14]. Many methods generally attempt to derive client characteristics by adjusting the relative weights better. Aggregation in the federated setting has also addressed fairness [15] in taking underrepresented clients and classes better into account.

Statistical heterogeneity, or *non-IID data*, leads to the difficulties of choosing models and performing hyperparameter tuning, as the data resides at clients, out of the reach of a preliminary analysis. The edge clients provide the supervision signal for supervised machine learning models. However, the lack of human annotation or interaction between humans and learning systems induces the *label scarcity* and leads to a more restricted application domain.

Label scarcity is one of the problems emblematic of the federated setting. The inability to access client data and the resulting black-box updates are tackled by carefully selecting the aggregation method and supplementary learning paradigms to fit specific real-world scenarios. As a result of label scarcity, the semi-supervised and unsupervised learning paradigms introduce essential techniques to deal with the uncertainty arising from unlabeled data. Faced with the problem that clients' local models can diverge during multiple epochs of local training, the server can be tasked with selecting the *most reliable* client models of the preceding round, regularizing the aggregation for achieving consistency. Fully unsupervised data can be enhanced via domain adaption, where the aim is to transfer knowledge from a labeled domain to an unlabeled one.

1.1 Taxonomy

To establish critical solutions for problems arising from private and non-IID data, we assess the current leading solutions in model fusion and how other learning paradigms are incorporated into the federated learning scenario. We propose a novel taxonomy of federated learning according to the model fusion principle and the connection to other learning paradigms. The taxonomy scheme, as illustrated in Table 1 with some representative instantiations, is organized as below.

- *Federated Model Fusion*. We categorize the major improvements to the pioneering FedAvg model aggregation algorithm into four subclasses (i.e., adaptive/attentive methods, regularization methods, clustered methods, and Bayesian methods), together with a special focus on fairness (Sect. 3).
- *Federated Learning Paradigms*. We investigate how the various learning paradigms fit into the federated learning setting (Sect. 4). The learning paradigms include some key supervised learning scenarios such as transfer learning, multi-task and meta-learning, and learning algorithms beyond supervised learning such as semi-supervised learning, unsupervised learning, and reinforcement learning.

1.2 Contributions

This survey starts from a novel viewpoint of federated learning by coupling federated learning with different learning algorithms. We propose a new taxonomy and conduct a timely and focused survey of recent advances in solving the heterogeneity challenge. Our survey's distinction compared with other comprehensive surveys is that we focused on the emerging trends of federated model fusion and learning paradigms, which are not intensively discussed in previous surveys. Besides, we connect these recent advances with real-world applications and discuss limitations and future directions in this focused context.

This survey is organized as follows. In Sect. 3, we assess in detail the significant improvements recent research has proposed on top of the pioneering FedAvg model aggregation algorithm [1]. In Sect. 4, we analyze how the various learning paradigms are fitted into the federated learning setting. In Sect. 5, we highlight recent successes in applied federated learning. Finally, in Sect. 6, we outline future research directions specifically from the viewpoint of model fusion and complementary learning paradigms. This paper is a focused survey, assessing only the aforementioned coupled subfields, of which learning paradigms make the learned models more robust, and model fusion brings those models together. For a more wide-ranging survey into federated learning, we recommend readers to refer [107–109].

Table 1 Federated learning with other learning algorithms: categorization, conjunctions, and representative methods

Main area	Subarea	Study
Federated model fusion	Adaptive aggregation	IDA [9], ASTW [10], SmartFL [16] ABAVG [17], FedPA [18]
	Attentive Aggregation	FedAtt [11], FedAttOpt [19], FedMed [20] FedAMP [21], AWFDRL [22] FedMCSA [23], ChannelFed [24]
	Regularization Methods	FedAwS [25], FedProx [26] Mime [27], FedDyn [28], FedMLB [29] BLUR & LUS [30], FedCR [31] FedU & dFedU [32], FedProto [33]
	Clustered Methods	FL+HC [13], IFCA [34], FeSEM [35] FedFast [36], k-FED [37] IFCA & UIFCA [38], FedCE [39]
	Bayesian Methods	FedMA [40], PFNM [14], FedBE [41] pFedBayes [42], NAFI [43]
	Fairness	q-FFL [15], AFL [44], FairFed [45] CFFL [46], F2MF [47]
Learning paradigms	Transfer Learning	FTL [48], FADA [49], FedSteg [50] FLTrELM [51], FedHealth [52], SFHTL [53] FedCrack [54]
	Multi-Task Learning	Mocha [55], Kernelized FMTL [56] CFL [57], CoFED [58], FedEM [59] FedMSplit [60], Spreadgnn [61]
	Meta Learning	FedMeta [62], Per-FedAvg [63] MOML & LocalMOML [64], MetaMF [65]
	Knowledge Distillation	FedMD [66], FedGKT [67], FedFed [68] FedDF [69], FedACK [70], CFeD [71] FedICT [72], FDL-HAD [73], FedFTG [74]
	Semi-Supervised Learning	FedMatch [75], PATE-G [76], SemiFL [77] imFed-Sem [78], FAPL [79],RSCFed [80] CBAFed [81], SUMA [82], FedCVT [83]
	Adversarial Learning	Sync. Strategies [84], FedGAN [85], PATE-G [76] DP-FedAvg-GAN [86], FADA [49], FairVFL [87] FAL [88], DBFAT [89], CalFAT [90], FedRBN [91]
	Unsupervised Learning	FURL [92], FPCA [93], FedCA [94] FADA [49], FedEMA [95], Orchestra [96] L-DAWA [97], FedX [98]
	Reinforcement Learning	FedRL [99], Favor [100], FRD and MixFRD [101] DRL-based Aggregator [102], FedSAM [103] QAvg/PAvg [104], SCCD [105], FedHQL [106]

2 Related survey

Several related surveys have been published in recent years, as summarized in Table 2. This section introduces the existing surveys and highlights our survey's contributions to the literature.

2.1 General survey of federated learning

Yang et al. [107] first defined the concepts of federated learning, introduced federated applications, and discussed data privacy and security aspects. Li et al. [108] systematically reviewed the federated learning building blocks,

Table 2 Comparison of related survey articles about federated learning

Publication	Scope
This survey	Learning algorithms
Jin et al. [114]	Semi-supervised learning
Xu et al. [110]	Healthcare informatics
Lo et al. [115]	Software engineering
Lim et al. [112]	Mobile edge networks
Lyu et al. [111]	Threats
Niknam et al. [113]	Wireless communication
Yang et al. [107]	General
Li et al. [108]	General
Kairouz et al. [109]	General
Li et al. [12]	General

including data partitioning, machine learning model, privacy mechanism, communication architecture, the scale of the federation, and motivation of federation. Kairouz et al. [109] detailed definitions of federated learning system components and different types of federated learning systems variations. Li et al. [12] discussed the core challenges of federated learning in communication efficiency, privacy, and some future research directions

2.2 Domain-specific survey

Other surveys review a specific domain. Xu et al. [110] surveyed the healthcare and medical informatics domain. Lyu et al. [111] discussed the security threats and vulnerability challenges dealing with adversaries in federated learning systems Lim et al. [112] focused on mobile edge networks. Niknam et al. [113] reviewed federated learning in the context of wireless communications, covering the data security and privacy challenges, algorithm challenges, and wireless setting challenges Jin et al. [114] conducted a review on federated semi-supervised learning. Jin et al.'s survey is the most related work to our paper. However, it only concentrates on semi-supervised learning. Our paper fills in its gap by including a wider range of model fusion and learning algorithms.

2.3 Distinction of our survey

Our paper reviews the emerging trends of federated learning from a unique and novel angle, i.e., the learning algorithms used in the federated learning paradigms, including the model fusion algorithms (Sec. 3) and the conjunction of federated learning and other learning paradigms (named as Federated X Learning in Sec. 4). This unique perspective has not been well-discussed in any of the aforementioned surveys. Our survey fills in this gap by reviewing recent publications. Besides, we point out challenges and outlook future directions in this specific category of research on federated learning.

3 Federated model fusion

3.1 Overview

The goal of federated learning is to minimize the empirical risks over local data as

$$\min_{\theta} f(\theta) = \sum_{k=1}^{m} p_k \mathcal{L}_k(\theta) \tag{1}$$

where θ is the learnable parameter of the global model, m is the total number of clients in the FL system, $\mathcal{L}_k$ is the local objective of the k-th client, p_k is the importance weight of the k-th client, and $\sum_k p_k = 1$. The widely applied federated learning algorithm, i.e., Federated Averaging (FedAvg) [1], starts with a random initialization or warmed-up model of clients followed by local training, uploading, server aggregation, and redistribution. The learning objective is configured by setting p_k to be $\frac{n_k}{\sum_k n_k}$. Federated averaging assumes a regularization effect, similar to dropout in neural networks, by randomly selecting a fraction of clients on each communication round. Sampling on each round leads to faster training without a significant drop in accuracy. Li et al. [116] conducted a theoretical analysis on the convergence of FedAvg without strong assumptions and found that the sampling and averaging scheme affects the convergence. Recent studies investigate some significant while less considered problems and explore different possibilities for improving vanilla averaging. To mitigate the client drift caused by heterogeneity in FedAvg, the SCAFFOLD algorithm [117] estimates the client drift as the difference between the update directions of the server model and each client model and adopts stochastically controlled averaging of the correct client drift. Reddi et al. [118] proposed adaptive optimization algorithms such as Adagrad and Adam to improve the standard federated averaging-based optimization with convergence guarantees. Singh et al. [119] adopted optimal transport, which minimizes the transportation cost of neurons, to conduct layer-wise model fusion.

3.2 Adaptive weighting

The adaptive weighting approach calculates adaptive weighted averaging of model parameters as:

$$\theta_{t+1} = \sum_{k=1}^{K} \alpha_k \cdot \theta_t^{(k)}, \tag{2}$$

where $\theta_t^{(k)}$ is current model parameter of k-th client, θ_{t+1} is the updated global model parameter after aggregation, and α_k is the adaptive weighting coefficient. Aiming to train a low variance global model with non-IID robustness, Yeganeh et al. [9] proposed an adaptive weighting approach called Inverse Distance Aggregation (IDA) by extracting meta information from the statistical properties of model parameters. Specifically, the weighting coefficient with inverse distance is calculated as:

$$\alpha_k = \left\|\theta_t - \theta_t^{(k)}\right\|^{-1} \Big/ \left(\sum_{k=1}^{K} \left\|\theta_t - \theta_t^{(k)}\right\|^{-1}\right). \tag{3}$$

Considering the time effect during federated communication, Chen et al. [10] proposed temporally weighted aggregation of the local models on the server as:

$$\theta_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n}\left(\frac{e}{2}\right)^{-(t-t^{(k)})} \theta_t^{(k)}, \tag{4}$$

where e is the natural logarithm, t is the current update round and $t^{(k)}$ is the update round of the newest $\theta^{(k)}$. Apart from the time effect, the accuracy of local models can also serve as an important reference for adaptive weighting. In [17], a novel FL algorithm termed Accuracy Based Averaging (ABAVG) is proposed. It can improve existing aggregation strategies in FL via increasing the convergence speed and better handling non-IID problems. In [16], a small amount of proxy data is used to optimize the aggregation weight of each client. The optimized aggregation leads to an FL system that is robust to both data heterogeneity and malicious clients.

Most works still conduct adaptive weighting among all clients, while [18] proposes an adaptively partial model aggregation strategy where only part of the clients contribute to the aggregated global model, addressing the straggler problem in FL and increasing communication efficiency.

3.3 Attentive aggregation

The federated averaging algorithm takes the instance ratio of the client as the weight to calculate the averaged neural parameters during model fusion [1]. In attentive aggregation, the instance ratio is replaced by adaptive weights as Eq. 5:

$$\theta_{t+1} \leftarrow \theta_t - \epsilon \sum_{k=1}^{m} \alpha_k \nabla \mathcal{L}(\theta_t^{(k)}), \tag{5}$$

where α_k is the attention scores for client model parameters. FedAtt [11] proposes a simple layer-wise attentive aggregation scheme that takes the server model parameter as the query. FedAttOpt [19] enhances the attentive aggregation of FedAtt by the scaled dot product. Like attentive aggregation, FedMed [20] proposes an adaptive aggregation algorithm using Jensen-Shannon divergence as the non-parametric weight estimator. These three attentive approaches use centralized aggregation architecture with only one shared global model for client model fusion. Huang et al. [21] studied pairwise collaboration between clients and proposed FedAMP with attentive message passing among similar personalized cloud models of each client. Wang et al. [22] incorporate the attention-weighted mechanism to federated learning systems to avoid the imbalance of local model quality. Concretely, the attention value is computed according to the average reward, average loss, training data size, etc, increasing the possibility of obtaining a more powerful agent model after aggregation.

The attention-based module is also widely used for personalized federated learning [23, 24]. In [24], the authors design a PFL framework termed ChannelFed that uses an attention module to assign different weights to channels on the client side. After incorporating personalized channel attention, the performance of the local model can be improved and client-specific knowledge can be better captured. In [23], a novel FL framework named federated model components self-attention (FedMCSA) is proposed to facilitate collaboration between clients with similar models. In this way, the personalized FL framework can adaptively update models and handle non-IIDness.

3.4 Regularization methods

We summarize federated learning algorithms with additional regularization terms to client learning objectives or server aggregation formulas. One category is to add local constraints for clients. FedProx [26] adds proximal terms to clients' objectives to regularize local training and ensure convergence in the non-IID setting. After removing the proximal term, FedProx degrades to FedAvg. Another direction is to conduct federated optimization on the server side. Mime [27] adapts conventional centralized optimization algorithms into federated learning and uses momentum to reduce client drift with only global statistics as

$$\mathbf{m}_t = (1-\beta)\nabla f_i\left(\mathbf{x}_{t-1}\right) + \beta \mathbf{m}_{t-1} \tag{6}$$

where $\mathbf{m}_{t-1}$ is a moving average of unbiased gradients computed over multiple clients and β is a trade-off parameter. Federated averaging may lead to class embedding collapse to a single point for embedding-based classifiers.

To tackle the embedding collapse, Yu et al. [25] studied the federated setting where users only have access to a single class, for example, face recognition in the mobile phone. They proposed the FedAwS framework with a geometric regularization and stochastic negative mining over the server optimization to spread class embedding space. To make the local-level objective and global-level objective consistent,

[28] proposes a novel dynamic regularization method, termed FedDyn, for FL. By dynamically adjusting the local optimization objective, FedDyn significantly saves communication costs when training across heterogeneous clients.

Kim et al. [29] aimed to address the inconsistency problem between different local models. It proposes FedMLB, a multi-level branched regularization-based FL framework, that prevents the local representations from being deviated too much by local updates. To alleviate the performance degradation problem after introducing user-level differential privacy guarantees, Cheng et al. [30] incorporated regularization techniques along with sparsification technical design into the local update procedure. To handle the training latency across devices and straggler issues, the authors in Chen et al. [31] presented a novel contrastive regularization-based scheme to accelerate the training process of FL. The proposed FedCR algorithm efficiently reduces the training latency and achieves better performance during the test phase. In [32], the authors proposed a new viewpoint to formulate the federated multi-task learning problem by Laplacian regularization, which can help to capture the relationships across clients. In [33], a prototype-based regularization term is added to the original local loss function to force the local representation center to be close to the global representation center. In this way, a balance between generalization and personalization can be achieved.

3.5 Clustered methods

We formulate clustered methods as algorithms that take additional steps with client clustering before federated aggregation or optimization to improve model fusion. One straightforward strategy is the two-stage approach. To be specific, during the global update procedure, the first step is a clustering process which is then followed by the aggregation process within each cluster. Briggs et al. [13] propose to take an additional hierarchical clustering for client model updates and apply federated averaging for each cluster. Diverting client updates to multiple global models from user groups can help better capture the heterogeneity of non-IID data. Xie et al. [35] proposed multi-center federated learning, where clients belong to a specific cluster, clusters update along with the local model updates, and clients also update their belongings to different clusters. The authors formulated a joint optimization problem with distance-based multi-center loss and proposed the FeSEM algorithm with stochastic expectation maximization (SEM) to solve the optimization. Muhammad et al. [36] proposed an active aggregation method with several update steps in their FedFast framework going beyond average. The authors worked on recommendation systems and improved the conventional federated averaging by maintaining user-embedding clusters. They designed a pipelined updating scheme for item embeddings, client delegate embeddings, and subordinate user embeddings to propagate client updates in the cluster with similar clients.

Ghosh et al. [34] formulated clustered federated learning by partitioning different user groups with the same learning tasks and conducting aggregation within the cluster partition. The authors proposed an Iterative Federated Clustering Algorithm (IFCA) with alternate cluster identity estimation and model optimization to capture the non-IID nature. The authors in [38] further extended IFCA to a more general scenario where the data in the same client may belong to different clusters. Based on IFCA, a new generative model-based clustering algorithm termed UIFCA is developed for unsupervised datasets. Dennis et al. [37] presented a one-shot communication scheme for clustering-based FL. The proposed method *k*-FED can significantly alleviate the problems caused by high communication costs and stragglers. This work also presents an interesting viewpoint that, compared with supervised learning, the statistical heterogeneity in unsupervised settings can bring about benefits to better convergence performance, fair models, etc. Considering the cases where each client can be associated with multiple clusters, Cai et al. [39] proposed to quantify the relationship between clients and clusters to better align clients with corresponding clusters. By introducing clustering ensembles, this work establishes a more comprehensive clustering method for FL and improves the performance of existing clustering FL methods.

3.6 Bayesian methods

Bayesian non-parametric machinery is applied to federated deep learning by matching and combining neurons for model fusion. Yurochkin et al. [14] proposed probabilistic federated neural matching (PFNM) using a Beta Bernoulli Process to model the multi-layer perceptron (MLP) weight parameters. Observing the permutation invariance of fully connected layers, the proposed PFNM algorithm first matches the neurons of neural models of clients to the global neurons. It then aggregates via maximum a posteriori estimation of global neurons. However, the authors only considered simple MLP architectures. FedMA [40] extends PFNM to convolutional and recurrent neural networks by matching and averaging hidden elements, specifically channels for CNNs and hidden units for RNNs. It solves the matched averaging objective by iterative optimization. Through theoretical analysis, Xiao and Cheng [43] found that global information can be omitted by PFNM. To fix this missing global information issue, an algorithm that conducts neural aggregation with full information (NAFI) is developed. NAFI introduces KL divergence-based penalty term to help complete the full information so that the missing information problem can be alleviated.

To obtain a more robust prediction via model aggregation, Chen and Chao [41] leveraged Bayesian techniques to sample high-quality models and aggregate the outputs of these models via Bayesian model ensemble. The proposed algorithm is termed FedBE, which has demonstrated applicability to deep networks and different heterogeneous scenarios. To tackle the model overfitting problem, Zhang et al. [42] proposed pFedBayes, a novel personalized FL method based on Bayesian variational inference, where all network parameters can be represented by probability distributions. Both the local and global models are formulated as Bayesian neural networks. The server aims to minimize the KL divergence between global distribution and local distributions, while the clients aim to minimize the construction error on local private data and the KL divergence with global distribution.

3.7 Fairness

When aggregating the global shared model, FedAvg applies a weighted average concerning the number of samples that participating clients used in their training. However, the model updates can easily skew towards an over-represented subgroup of clients where super-users provide the majority of samples. Mohri et al. [44] suggested that valuing each sample without clear discrimination is inherently risky as it might result in sub-optimal performance for underrepresented clients and sought good-intent fairness to ensure federated training not overfitting to some of the specific clients. Instead of the uniform distribution in classic federated learning, the authors proposed agnostic federated learning (AFL) with minimax fairness, which takes a mixture of distributions into account. However, the overall tradeoff between fairness and performance is still not well explored. Inspired by fair resource allocation in wireless networks, the q-fair federated learning (q-FFL) [15] proposes an optimization algorithm to ensure fair performance, i.e., a more uniform distribution of performance gained in federated clients. The optimization objective (Eq. 7) adjusts the traditional empirical risk objective by tunable performance-fairness tradeoff controlled by q.

$$\min_{\theta} f_q(\theta) = \sum_{k=1}^{m} \frac{p_k}{q+1} \mathcal{L}_k^{q+1}(\theta) \tag{7}$$

The flexible q-FFL also generalizes well to previous methods; specifically, it reduces to FedAvg and AFL when the value of q is set to 0 and ∞, respectively.

To investigate the fairness issue in FL systems, Lyu et al. [46] emphasized collaborative fairness. To be specific, all clients receive the same or similar models, though their contributions differ a lot. The authors proposed a novel framework named Collaborative Fair Federated Learning (CFFL), which can take the contribution of each client into consideration and let each client receive models with performance commensurate with their contributions.

Usually, fairness in FL refers to the individual-wise measurement. In [45], the authors investigated fairness problems in FL from a group-wise perspective. Inspired by group fairness in centralized learning, a novel algorithm termed FairFed is developed for participants to conduct aggregation in a fairness-aware way. FairFed can efficiently mitigate the bias against specific populations while maintaining the privacy of local data.

To achieve fairness for recommender systems, Liu et al. [47] proposed to capture the affiliation feature across different groups by using federated learning as a privacy-preserving tool. Based on the existing federated recommendation backbone [120], it designs fairness-aware federated matrix factorization (F2MF), a solution that deals with the conflict between the global fairness objective and the local federated optimization process. By introducing a loss-based fairness metric into the optimization process, the FL systems potentially improve the fairness of recommendations between different user groups.

4 Federated X learning

The customizability of federated learning objectives leads to possibilities in quickly adapting FL to adversarial, semi-supervised, or reinforcement learning settings, offering flexibility to other learning algorithms in conjunction with federated learning. We term FL's intersection with other learning algorithms as Federated X Learning.

4.1 Federated transfer learning

Transfer learning focuses on transferring knowledge from one particular problem to another, and it has also been integrated into federated learning to construct a model from two datasets with different samples and feature spaces [107, 121]. Liu et al. [48] formulated the Federated Transfer Learning (FTL) to solve the problem that traditional federated learning falters when datasets do not share sufficient common features or samples. In this paper, it assumes existing two domains A and B across different parties and formulate the objective function as:

$$\begin{aligned}\min_{\theta_A,\theta_B} \mathcal{L}(\theta_A,\theta_B) &= \ell_1(y^A, \phi(x^B)) + \gamma \ell_2(\phi(x^A), \phi(x^B)) \\ &+ \frac{\lambda}{2}\|\theta_A\|^2 + \frac{\lambda}{2}\|\theta_B\|^2.\end{aligned} \tag{8}$$

where θ_A and θ_B are the model parameters in these two domains while $\phi(\cdot)$ represents the transformation function that projects data into a unified feature space. ℓ_1 and ℓ_2 are logistic loss and alignment loss, respectively. γ and λ are tuneable hyper-parameters. The authors also enhanced the security with homomorphic encryption and secret sharing. In real-world applications, FedSteg [50] applies federated transfer learning for secure image steganalysis to detect hidden information. Alawad et al. [122] utilized federated transfer learning without sharing vocabulary for privacy-preserving NLP applications for cancer registries.

To deal with the widely existing overlapping data insufficiency problem across clients, Feng et al. [53] proposed a Semi-Supervised Federated Heterogeneous Transfer Learning (SFHTL) framework that leverages unlabeled non-overlapping samples to reduce model overfitting. Compared with existing FTL methods, SFHTL can better expand the training set to improve the performance of the local model.

Federated transfer learning can be widely used in various real-world applications, including intrusion detection [51], smart healthcare [52], crack detection [54], etc. It allows the knowledge learned within one specific domain to be transferred to another different domain, especially when there are no sufficient common features across these domains.

4.2 Federated learning with knowledge distillation

Given the assumption that clients have sufficient computational capacity, federated averaging adopts the same model architecture for clients and the server. FedMD [66] couples transfer learning and knowledge distillation (KD), where the centralized server does not control the architecture of models. It introduces an additional public dataset for knowledge distillation, and each client optimizes their local models on both public and private data, like VHL [123]. They employ a combination of a public noise dataset and local private data to train the local model. Furthermore, it leverages domain adaptation techniques to improve the overall performance of the model. In general, the local objective of federated learning with knowledge distillation is often combined with two items:

$$\min_{\theta_k} \mathcal{L}(\theta_k) = \ell_{task} + \ell_{KD}. \tag{9}$$

where θ_k is the local model of the k-th client and ℓ_{task} is task-specific loss, ℓ_{KD} is often computed by different logits or features from various clients.

Strictly speaking, transfer learning differs from knowledge distillation; however, the FedMD framework puts them under one umbrella. Many technical details are only briefly introduced in the original paper of FedMD. Recently, He et al. [67] utilized knowledge distillation with technical solidity to train computationally affordable CNNs for edge devices via knowledge distillation. The authors proposed the Group Knowledge Transfer (FedGKT) framework that optimizes the client and the server model alternatively with knowledge distillation loss. Specifically, the larger server model takes features from the edge to minimize the gap between periodically transferred ground truth and soft label predicted by the edge model, and the small model distils knowledge from the larger server model by optimizing the KD-loss, ℓ_{KD} in Eq.(9), using private data and soft labels transferred back from the server. However, this framework has a potential risk of privacy breach as the server holds the ground truth, especially when ground truth labels are the user's typing records in the mobile keyboard application. Lin et al. [69] applied knowledge distillation to mitigate privacy risk and cost and proposed a novel ensemble distillation for model fusion that utilizes unlabeled data.

Knowledge distillation continues to demonstrate significant potential in addressing various challenges within FL. FedFed [68] introduces a novel variant of knowledge distillation named feature distillation. The authors propose a method where the data is partitioned into two distinct parts, allowing for the sharing of protected performance-sensitive features to alleviate the data heterogeneity. Zhang et al. [74] addressed this challenge by employing Data-Free Knowledge Distillation and proposed FedFTG. Their approach involves the use of a generator to distil and transfer local knowledge to the global model. To improve communication efficiency, Zhang et al. [73] proposed a method called FDL-HAD. It introduces an adaptive regulation mechanism that determines whether clients need to undergo distillation in each round.

Furthermore, knowledge distillation is also valuable in many other federated X learning paradigms. CFeD [71] addresses the challenge of catastrophic forgetting in continual federated learning through KD. Moreover, multi-task learning is an important scenario in federated learning. Wu et al. [72] specifically designed an algorithm tailored for multi-access edge computing in a real-world scenario, leveraging knowledge distillation as a key component. FedNed [124] solve the noisy clients by a kind of KD, called negative distillation. FedACK [70] applies knowledge distillation in the cross-lingual social bot detection domain, showcasing a novel application that combines knowledge distillation and federated learning. This application demonstrates the potential for knowledge distillation to inspire more useful applications within this emerging field.

4.3 Federated multi-task learning

Federated Multi-Task Learning trains separate models for each client with some shared structure between models, where learning from local datasets at different clients is regarded as a separate task. In contrast to federated transfer learning between two parties, federated multi-task learning involves multiple parties and formulates similar tasks clustered with specific constraints over model weights. It exploits related tasks for more efficient learning to tackle the statistical heterogeneity challenge. In federated multi-task learning, the target is to train multiple related tasks across clients with different objective functions:

$$\min_{\theta_1,\ldots,\theta_K,\Omega} \left\{ \sum_{k=1}^{K}\sum_{i=1}^{n_k} f_k\left(\theta_k, \mathbf{x}_i, y_i\right) + \mathcal{R}(\Theta,\Omega) \right\} \tag{10}$$

where $\Theta = [\theta_1, \theta_2, \ldots, \theta_K] \in \mathbb{R}^{d\times K}$ is a matrix collecting weight vectors of all clients and Ω denotes the relationships of different clients with their corresponding tasks. The Mocha framework [55] trains separate yet related models for each client by solving a primal-dual optimization. It leverages a shared representation across multiple tasks and addresses the challenges of data and system heterogeneity. However, the Mocha framework is limited to regularized linear models. Caldas et al. [56] further studied the theoretical potential of kernelized federated multi-task learning to solve the non-linearity. To solve the suboptimal results, Sattler et al. [57] studied the geometric properties of the federated loss surface. They proposed a federated multi-task framework with non-convex generalization to cluster the client population. [59] studies federated multi-task learning under a general assumption that each local data distribution can be seen as a mixture of distributions. Hence, each client learns personalized mixture weights to obtain its personalized local model. There are two algorithms, termed FedEM and D-FedEM, proposed for the client–server and fully decentralized setting, respectively. The approaches yield models with more accurate results, better generalization ability, and fairer performance across clients.

There is also a branch of works that utilizes a federated multi-task learning framework to deal with data in different formats, including graph data [61] and multimodal data [60]. To benefit cross-silo FL where independent data silos have different tasks, Cao et al. [58] proposed a novel FL method CoFED that utilizes a co-training scheme to leverage unlabeled data in a semi-supervised learning manner. CoFED is compatible with heterogeneous models, tasks, and training processes, making it an effective method for federated multi-task learning.

4.4 Federated meta learning

Federated meta learning aims to train a model that is quickly adapted to new tasks with few training data, where clients serve as a variety of learning tasks. The seminal model-agnostic meta-learning (MAML) framework [125] has been intensively applied to this learning scenario. Several studies connect FL and meta-learning, for example, model updating algorithm with average difference descent [126] inspired by the first-order meta-learning algorithm. However, this study focuses on applications in the social care domain with less consideration in practical settings. Jiang et al. [127] further provided a unified view of federated meta-learning to compare MAML and the first-order approximation method. Inspired by the connection between federated learning and meta-learning, Fallah et al. [63] adapted MAML into the federated framework Per-FedAvg, to learn an initial shared model, leading to fast adaption and personalization for each client. FedMeta [62] proposes a two-stage optimization with a controllable meta updating scheme after model aggregation as:

$$\theta_{t+1}^{meta} = \theta_{t+1} - \eta_{meta}\nabla_{\theta_{t+1}}\mathcal{L}\left(\theta_{t+1};\mathcal{D}_{meta}\right), \tag{11}$$

where $\mathcal{D}_{meta}$ is a small set of meta data on the server and η_{meta} is the meta learning rate.

To better exploit the collaborative filtering information across clients for recommender systems, [65] introduces a federated matrix factorization framework named meta matrix factorization (MetaMF). In MetaMF, a meta network is used to generate private item embeddings and rating prediction models based on the collaborative vector in the server. MetaMF achieves competitive performance despite using a small model scale and embedding size. To address the underdeveloped stochastic optimization in MAML, Wang et al. [64] proposed a memory-based stochastic algorithm that ensures convergence with vanishing error, enabling constant mini-batch sizes and making them suitable for continual learning. Meanwhile, this paper introduces a communication-efficient memory-based MAML algorithm for personalized federated learning in cross-device and cross-silo settings. Lin et al. [128] proposed MetaGater, a federated meta-learning algorithm that holistically trains both the backbone network and channel gating. MetaGater enables efficient subnet selection for resource-constrained applications by leveraging model similarity across learning tasks on different nodes, ensuring the effective capture of important filters for quick adaptation to new tasks with experimental results validating its effectiveness.

4.5 Federated adversarial learning

In this section, we summarize federated adversarial learning in two categories. The first class of methods specifically focuses on Generative Adversarial Networks (GANs), a mainstream adversarial learning paradigm for data generation. The second class of methods, differently, uses the idea of adversarial learning to address the challenges of general federated learning.

GANs consist of two competing models, i.e., a generator and a discriminator. The generator learns to produce samples approximating the underlying ground-truth distribution. The discriminator, usually a binary classifier, tries to distinguish the samples produced by the generator from the real samples. A straightforward combination with FL is to have the

GAN models trained locally on clients and the global model fused with different strategies. Fan and Liu [84] studied the synchronization strategies for aggregating discriminator and generator networks on the server and conducted a series of empirical analyses. Updating clients on each round with both the generator and the discriminator models achieves the best results; however, it is twice as computationally expensive as just syncing the generator. Updating just the generator leads to almost equivalent performance than updating both, whereas updating just the discriminator leads to considerably worse performance, closer to updating neither. Rasouli et al. [85] extended the federated GAN with different applications and proposed the FedGAN framework to use an intermediary for averaging and broadcasting the parameters of generator and discriminator. Furthermore, the authors studied the convergence of distributed GANs by connecting the stochastic approximation and communication-efficient SGD optimization for GAN and federated learning. Augenstein et al. [86] proposed differentially private federated generative models to address the challenges of non-inspectable data scenarios. GANs are adopted to synthesize realistic examples of private data for data labeling inspection at inference time.

Apart from generation models, another type of method aims to leverage adversarial learning to enhance several capabilities of federated learning systems, such as fairness and robustness. To enhance fairness under vertical federated learning scenarios, FairVFL [87] employs adversarial learning to mitigate bias, while incorporating a contrastive adversarial learning method to protect user privacy while effectively improving model fairness. To handle the unfair scenarios with label skewness, Chen et al. [90] proposed CalFAT, a federated adversarial training method that adaptively calibrates logits to balance classes, which addresses the root cause of issues related to skewed labels and non-identical class probabilities. Specifically, it can be formulated as:

$$\min \ell_1\left(\gamma \cdot \theta_k\left(x_{adv}\right), y\right) + \lambda \cdot \ell_2\left(\theta_k\left(x_{adv}\right), \theta_g(x)\right), \tag{12}$$

where θ_k denotes local model while θ_g is global model. Cross-entropy loss is represented as ℓ_1. KL-loss ℓ_2 is used to constrain the logits of the local and global model. γ and λ are hyper-parameters. In order to improve the robustness of federated learning models against adversarial attacks, Li et al. [88] introduce FAL, a novel bi-level approach with min-max optimization for adversarial learning of federated learning. Specifically, FAL incorporates an inner loop for generating adversarial samples during adversarial training and an outer loop for updating local model parameters. Zhang et al. [89] conducted comprehensive evaluations on various attacks and adversarial training methods, revealing negative impacts on test accuracy when directly applying adversarial training in FL. Based on the findings, they further propose DBFAT, a novel algorithm with local re-weighting and global regularization components, demonstrating superior performance in terms of both accuracy and robustness across multiple datasets in both IID and non-IID settings. To address the challenge of adversarial robustness in federated learning with heterogeneous users, Hong et al. [91] introduced a novel strategy: propagating adversarial robustness from rich-resource users to those with limited resources during FL by utilizing batch normalization.

4.6 Federated semi-supervised learning

Annotation capability plays a crucial role in traditional machine learning and deep learning [129, 130]. The quality and quantity of annotations often determine the performance of models. However, the problem of data heterogeneity naturally arises in decentralized federated learning, posing additional challenges.

Label scarcity is a prevalent and widespread issue in federated learning scenarios, which has prompted the development of a novel learning setup known as federated semi-supervised learning (FSSL). This scenario reflects the realistic situation where users may not label all the data on their devices. Papernot et al. [76] explored semi-supervised learning in distributed scenarios. They put forward a semi-supervised approach with a private aggregation of teacher ensembles (PATE), an architecture where each client votes on the correct label. PATE was shown empirically to be particularly beneficial when used in conjunction with GANs. Similar to centralized semi-supervised learning, the majority of FSSL approaches often utilize a two-part loss function on the client devices. This loss function typically consists of a supervised learning component, denoted as $\mathcal{L}_s(\theta)$, and an unsupervised learning component, denoted as $\mathcal{L}_u(\theta)$. Existing FSSL methods have focused on two different scenarios [75]: labels-at-server and labels-at-clients.

In the labels-at-server scenario, the server has the ability to annotate data, while the client is limited to only collecting data without the capacity to annotate it due to a shortage of expert resources. Numerous works have been dedicated to addressing this specific setting. SemiFL [77] tackles this problem through alternate training. This process consists of two key steps: fine-tuning the global model with labeled data and generating pseudo-labels using the global model on the client side. Importantly, the server and client models are trained in parallel to enable efficient collaboration. Jeong et al. [75] proposed a federated matching (FedMatch) framework with inter-client consistency loss to exploit the heterogeneous knowledge learned by multiple client models. The authors showed that learning on both labeled and unlabeled data simultaneously may result in the model forgetting what it had learned from labeled data. To counter this, the authors decomposed the model parameters

θ to two variables $\theta = \psi + \rho$ and utilize a separate updating strategy, where only ψ is updated during unsupervised learning, and similarly, ρ is updated for supervised learning. In a real-world scenario, Jiang et al. [78] addressed the challenge of imbalanced class distributions among unlabeled clients in the context of medical image diagnosis. They proposed a novel scheme called dynamic bank learning, which aims to collect confident samples and subsequently divide them into sub-banks with varying class proportions.

In contrast, the labels-at-clients approach focuses on scenarios where clients lack sufficient capability to label data. In this setting, the server's primary role is to regulate the federated learning process, without involvement in data collection or ownership. Two types of assumptions exist within this approach: partially labeled data at each client, referred to as **P**artially **D**ata Federated Semi-supervised Learning (PD-FSSL), and partially labeled clients themselves, denoted as **P**artially **C**lients Federated Semi-supervised Learning (PC-FSSL). We can formulate the local function in PD-FSSL as:

$$\min_{\boldsymbol{\theta}_{\mathbf{k}}} \mathcal{L}(\boldsymbol{\theta}_{\mathbf{k}}) = \ell_{sup}(\mathbf{x}_e^k, y_e) + \ell_{unsup}(\mathbf{x}_r^k) \tag{13}$$

where $\mathbf{x}_e$ is labeled data while $\mathbf{x}_r$ represents unlabeled data on the k-th client. In PD-FSSL, the limited labeling capability of each client results in only a portion of the data being labeled. Consequently, the private data of each client is divided into a labeled part and an unlabeled part. FedMatch [75] has demonstrated its effectiveness not only in the client-at-server scenario but also in PD-FSSL. Additionally, FAPL [79] focuses on addressing fairness in PD-FSSL. The authors aim to achieve a balance in the total number of active unlabeled samples (AUSs) for different classes across all selected clients in a global round. They accomplish this by globally aligning the numbers of AUSs for different classes, which helps enhance fairness in the learning process. Another variation, PC-FSSL, assumes that some clients possess the resources and ability to label data, while others can only collect data without annotation. RSCFed [80] proposes a sub-consensus framework. In this framework, traditional cross-entropy training is performed on clients with labeled data. For clients without labels, a consistency regularization framework, such as mean-teacher, is utilized. Generally, the global objective function in PC-FSSL can be written as:

$$\min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) = \sum_{a=1}^{A} \lambda_a \mathcal{L}_a(\theta_a) + \sum_{b=1}^{B} \lambda_b \mathcal{L}_b(\theta_b), \tag{14}$$

where the global model θ aims to minimize a function that is affected by two types of clients: fully-labeled clients, denoted as a, and fully-unlabeled clients, denoted as b. Specifically, $\mathcal{L}_a$ represents the supervised task-relevant loss, which differs from $\mathcal{L}_b$. Among the fully-unlabeled clients, $\mathcal{L}_b$ can be a mean-teacher or contrastive loss. Additionally, RSCFed employs data augmentation techniques, similar to conventional semi-supervised learning, to augment the unlabeled data twice, further improving the learning process. Similarly, CBAFed [81] also utilizes augmentation techniques for pseudo-labeling in the PC-FSSL setting. They introduce an adaptive threshold to determine the reliability of the pseudo-labels generated from the unlabeled data. There are still other scenarios in FSSL. SUMA [82] considers a more general setting where each client has a different ratio of labeled data. FedCVT [83] studies FSSL in vertical federated scenarios.

Despite extensive research, FSSL still faces many challenges. The problem of insufficient data labels in practical applications still necessitates further investigation in order to find effective solutions. Furthermore, existing FSSL algorithms often demonstrate limitations in their performance across various settings, which also leaves a lot of room for exploration.

4.7 Federated unsupervised learning

It is more common that local clients host no labeled data, which naturally leads to the learning paradigm of federated unsupervised learning without supervision in the decentralized learning scenario. A straightforward solution is to pretrain unlabeled data to learn useful features and utilize pretrained features in downstream tasks of federated learning systems [92]. There exist two challenges in federated unsupervised learning, i.e., the inconsistency of representation spaces due to data distribution shift and the misalignment of representations due to the lack of unified information among clients.

FedCA [94], based on SimCLR, proposes a federated contrastive averaging algorithm with the dictionary and alignment modules for client representation aggregation and alignment, respectively. Zhuang et al. [95] conducted comprehensive experiments to evaluate the performance of four popular unsupervised methods in FL: MoCo (V1[131] and V2 [132]), BYOL [133], SimCLR [134], and SimSiam [135]. In their study, the authors discovered that Fed-BYOL demonstrates superior performance compared to the other evaluated methods. They also highlighted the importance of the predictor, exponential moving average (EMA), and stop-gradient operations in improving the performance of non-contrastive federated self-supervised learning. Drawing from their extensive experiments, the authors propose a new method called FedEMA. It incorporates a divergence-aware dynamic moving average update to address the challenges associated with non-IID data in the federated setting. FedX [98] also employs the contrastive paradigm by a two-sided knowledge distillation. Additionally, Lubana et al. [96] conducted an evaluation of federated versions of the prevailing unsupervised methods. Furthermore, they introduced a novel clustering-based method called Orchestra, which differs significantly from mainstream unsupervised algorithms.

The local model training utilizes the contrastive loss and the server aggregates models and dictionaries from clients. Recently, many unsupervised learning methods such as Principal Component Analysis (PCA) and unsupervised domain adaptation have been adopted to combine with federated learning. Peng et al. [49] studied the federated unsupervised domain adaptation that aligns the shifted domains under a federated setting with a couple of learning paradigms. Specifically, unsupervised domain adaptation is explored by transferring the labeled source domain to the unlabelled target domain, and adversarial adaptation techniques are also applied. Grammenos et al. [93] proposed the federated PCA algorithm with a differential privacy guarantee. The proposed FPCA method is permutation invariant and robust to straggler or fault clients. In contrast, L-DAWA [97] takes a different approach by proposing a novel aggregation strategy through layer-wise divergence. L-DAWA introduces angular divergence σ_k to represent the aggregation weight of the k-th client:

$$\sigma_k = \frac{\theta_g^r \cdot \theta_k^r}{\|\theta_g^r\| \cdot \|\theta_k^r\|} \tag{15}$$

where θ_g^r is the r-th round global parameters and θ_k^r is the r-th round local model of client k. They aggregate weights at the layer-level by utilizing the measure of angular divergence between the models of individual clients and the global model.

4.8 Federated reinforcement learning

In deep reinforcement learning (DRL), the deep learning model gets rewards for its actions and learns which actions yield higher rewards. Zhuo et al. [99] introduced reinforcement learning to federated learning framework (FedRL), assuming that distributed agents do not share their observations. The proposed FedRL architecture has two local models: a simple neural network, such as a multi-layer perceptron (MLP), and a Q-network that utilizes Q-learning [136] to compute the reward for a given state and action. The authors provided algorithms on how their model works with two clients and suggested that the approach can be extended to many clients using the same approach. In the proposed architecture, the clients update the local parameters of their respective MLPs first and then share the parameters to train these q-networks. Clients work out this parameter exchange in a peer-to-peer fashion. Federated reinforcement learning can improve federated aggregation to address the non-IID challenge, and it also has real-world applications, such as in the Internet of Things (IoT). A control framework called Favor [100] improves client selection with reinforcement learning to choose the best candidate for federated aggregation. The federated reinforcement distillation (FRD) framework [101], together with its improved variant MixFRD with mixup augmentation, utilizes policy distillation for distributed reinforcement learning. In the fusion stage of FRD, only proxy experience replay memory (ProxRM) with locally averaged policies are shared across agents, aiming to preserve privacy. Facing the tradeoff between the aggregator's pricing and the efficiency of edge computing, Zhan et al. [102] investigated the design of an incentive mechanism with DRL to promote edge learning. In FedSAM [103], the authors extended widely used RL methods, such as on-policy TD (Temporal-Difference) [137], off-policy TD [137], and Q-learning [136], to the federated learning. Subsequently, they put forth an algorithm that integrates federated TD-learning and Q-learning and conducted an extensive analysis of the convergence to these federated RL methods. In real-world applications, RL agents often encounter diverse state transitions across different environments, so-called environmental heterogeneity. Jin et al. [104] investigated this novel setting within FedRL and presented a series of diverse variation approaches to address the varying degrees of complexity in heterogeneous environments. SCCD [105] is also an off-policy-based FedRL framework that introduces a student-teacher-student model learning and fusion method. Fan et al. [106] analyzed the existing FedRL setting, introduced a new problem called federated reinforcement learning with heterogeneous and black-box agents (FedRL-HALE), and posed a challenge called the exploration-exploitation dilemma. This dilemma entails the tradeoff that an agent encounters when making decisions between exploring new actions to gather more information or exploiting its current knowledge to maximize performance. Then, they proposed FedHQL, where the local agents update their action-value independently based on Q-learning. The central server plays a crucial role in coordinating the exchange of knowledge between agents by broadcasting, receiving action-value estimates, and selecting actions with the highest UCB (Upper Confidence Bound) [138] value. There are still many ongoing explorations in other areas where FedRL is being applied, including energy management [139, 140], electric vehicle charging and uncharging [141].

5 Challenges and applications

This section highlights the multifaceted nature of federated learning research, addressing challenges related to client heterogeneity, data privacy, model security, and efficient communication, while also exploring its applicability to a wide range of real-world use cases.

5.1 Statistical and model heterogeneity

Variability among clients, referred to as the heterogeneity problem, stands as the principal hurdle in FL. The most common heterogeneity issues are statistical and model heterogeneity. Addressing both these two challenges is important to achieve effective FL with better personalization and generalization ability.

The statistical heterogeneity challenge arises due to the non-IID (non-identically distributed) nature of data, where each client holds a unique subset of data, often reflecting distinct features, patterns, or statistical characteristics. This variability complicates the process of aggregating information from diverse sources to create a global model. Addressing statistical heterogeneity is crucial as it impacts the performance and generalizability of the global model, requiring specialized techniques that account for and mitigate these disparities without compromising data privacy or communication efficiency.

[26] proposes a local regularization approach to refine the local model of each client. Recent research efforts [142–144] focus on training personalized models, amalgamating globally shared insights with personalized elements [19, 145]. Another approach involves providing multiple global models through clustering local models into distinct groups or clusters [34, 57, 146]. Additionally, recent advancements incorporate self-supervised learning techniques during local training to address these heterogeneity challenges [147–149]. For personalized FL, [63] applies meta-training strategies.

Model heterogeneity exists in FL when there are diverse architectures, configurations, or complexities of models utilized by different clients or devices within the same system. This challenge arises because various participants may employ distinct types of machine learning models, differing in depth, structure, optimization techniques, or even hardware capabilities. Addressing model heterogeneity involves strategies to harmonize various model architectures, enabling collaborative learning while accommodating the varying computational capacities and model complexities across different devices or clients.

Knowledge Distillation (KD)-based FL methods [66, 69, 150, 151] usually assume the inclusion of a shared toy dataset in the federated setting, allowing knowledge transfer from a teacher model to student models with differing architectures. Recent studies also explore merging neural architecture search with FL [152–154], aiming to craft customized model architectures tailored to groups of clients with varying hardware capabilities and configurations. [155] introduces a collective learning platform to handle heterogeneous architectures without accessing local training data or architectures. Moreover, functionality-based neural matching across local models aggregates neurons based on similar functionalities, irrespective of architectural differences [40].

5.2 Security and privacy

In the realm of federated learning, the dual concerns of security and data privacy have driven extensive research into the development of privacy-preserving solutions and the identification of novel attack strategies [156–161]. To safeguard data privacy, recent studies have primarily focused on methods for safeguarding model parameters, thereby preventing unauthorized access to client data and its distribution by the global model.

Notable examples include the FLAME framework [156], which employs randomized and encrypted gradient vectors sent to a shuffler to protect client identities, and SplitFed [157], which combines split learning and federated learning to enhance privacy while maintaining performance. From a security perspective, addressing malicious client behavior has been crucial. Robust learning rate techniques have been proposed to minimize the impact of backdoor attacks [158, 159], alongside strategies like introducing data heterogeneity or using a coordinator to train updated weights before aggregation [158, 160]. Additionally, FedInv presents a novel approach by synthesizing a dummy dataset to mitigate Byzantine attacks effectively [160]. However, the Neurotoxin attack serves as a reminder of persistent threats, as it inserts enduring backdoors into federated learning systems by exploiting sparse gradient descent [161], thereby necessitating continuous efforts to enhance security and privacy in federated learning.

Ensemble Federated Learning (EFL) employs multiple global models and label probabilities relative to the ensemble model client number to counteract the influence of malicious clients, as discussed in [162]. Wen et al. [163] investigated attacks on federated learning that allow the central server to produce malicious parameter vectors, compromising privacy in horizontal and vertical FL settings. Proposed defense strategies include gradient clipping and noise addition. Gupta et al. [164] focused on the recovery of text information during the exchange of parameters in FL. To mitigate this risk, they propose a method to freeze the word embeddings of the model. Bietti et al. [165] addressed the tradeoff between privacy and model accuracy by introducing Personalized-Private-SGD (PPSGD) to personalize local models while preserving privacy. Zhang et al. in [166] studied client-level differential privacy (DP) for federated learning, highlighting the superiority of difference clipping. Hu et al. proposed FedSPA in [167], a sparsification-based privacy mechanism. Sun et al. presented Locally Differential Private Federated Learning in [168], focusing on adaptive range perturbation. Yang et al. [68] explored the Gaussian or Laplacian noise to protect shared features with a differential privacy guarantee. Furthermore, a DP protection method called FKGE [169] is utilized to study the embedding of knowledge graphs in a distributed manner.

FLSG [170] generates random Gaussian noise with the same size of gradient and sends the most similar to the server. Rong et al. in [171] explored poisoning attacks on federated recommender systems. Huang et al. in [172] examined gradient inversion attacks and defense mechanisms. Jin et al. introduced CAFE in [173], a method to recover large batch data from gradients. Sun et al. proposed FL-WBC in [174], a defense mechanism against global model poisoning. FedDefender [175] also focuses on the client side to achieve attack tolerance, which consists of local meta update and global distillation. Park et al. presented Sageflow in [176] to handle slow devices and malicious attacks. Finally, Agarwal et al. extended differential privacy using the Skellam mechanism in [177].

Additionally, many other cryptographic methods are widely used to preserve privacy in FL. Chang et al. [178] revisited many technologies in FL and propose 2DMCFE, a functional encryption method to protect privacy under semi-honest security setting. Furthermore, Hijazi et al. [179] also investigated the use of Fully Homomorphic Encryption (FHE) in FL. To mitigate inference attacks, Zhao et al. [180] proposed an effective strategy that leverages computational Diffie-Hellman (CDH) for generating lightweight keys. These research contributions collectively advance the field of federated learning by addressing various privacy and security challenges with diverse strategies and insights.

5.3 Communication efficiency

Communication efficiency is a challenging research direction in federated learning, which typically focuses on reducing the communication overhead between clients and servers, aiming to minimize data transmission and communication rounds. Several approaches have been proposed to enhance this aspect. Gao et al. introduced two communication-efficient distributed SGD methods in [181], which reduce the communication cost by compressing exchanged gradients and combining local SGD with compressed gradients to the momentum technique. Wang et al. proposed FedCAMS in [182], which combines the Federated AMSGrad adaptive gradient method with Max Stabilization and uses error feedback compression to reduce communication costs. GossipFL [183] uses the sparsified model to reduce communication and gossip matrix for efficient utilization of the bandwidth resources. Yi et al. presented the QSFL algorithm in [184], which samples high-qualification clients for model updates and compresses each update to a single segment. Zhu et al. [185] addressed system heterogeneity and communication efficiency in unstable connections with the FedResCuE algorithm, focusing on the self-distillation of prunable neural networks on clients. Yapp et al. introduced the BFEL framework in [186], which employs blockchain technology to reduce communication overhead by decentralizing the aggregation process. Meanwhile, Zhu et al. proposed Delayed Gradient Averaging (DGA) in [187] to mitigate high communication latency by pipelining communication with computation. Another method called FedPM [188] addresses the challenge of high communication costs in federated learning by freezing weights at initial random values and learning to sparsify the random network. Finally, FedProg [189] extends the progressive learning technique from image generation to federated learning, inherently reducing computational and two-way communication costs while preserving model performance.

In addition to the aforementioned solutions for the general federated framework, there are specialized approaches tailored to address communication challenges in specific scenarios. To address the communication limitation of existing federated learning-based contextual bandit algorithms, Li and Wang [190] introduced a communication-efficient framework utilizing generalized linear bandit models with online regression for local updates and offline regression for global updates. Especially aiming to address the communication challenge in minimax federated framework (e.g., GAN), FedGDA-GT [191] combines gradient tacking with federated gradient descent ascent framework, showcasing linear convergence with constant stepsizes to a global-approximation solution. Cui et al. [192] especially focus on the compute efficiency at the mobile-edge cloud computing system. Furthermore, decentralized training and deploying LLM in a federated manner also need more attention [193]. For federated node embedding problems in graph machine learning [194], Pan and Zhu [195] proposed a random-walk-based algorithm featuring a sequence encoder for privacy preservation and a two-hop neighbor predictor, effectively reducing communication costs.

5.4 Real-world applications

Model fusion and federated X learning have yielded remarkable achievements in some real-world applications. In this subsection, we mainly summarize the applications of federated learning in two research fields, i.e., recommendation and healthcare.

Recommendation is a practical real-world scenario. As a pioneering work, FedFast [36] is a novel approach for accelerating federated learning of deep recommendation models. FedFast efficiently samples from a diverse set of participating clients and employs an active aggregation method, enabling users to benefit from lower communication costs and access more accurate models at the early stages of training. Liang et al. [196] proposed FedRec++, a novel lossless federated recommendation method that enhances privacy-aware preference modeling and personalization in federated recommender systems by allocating denoising clients to eliminate noise introduced by virtual ratings, ensuring accurate and privacy-preserving recommendations with minimal additional communication cost. Motivated by a similar target, Cali3F [197] is a personalized federated recommendation system training algorithm, coupled with a clustering-based aggregation method, to address privacy concerns and enhance fairness in recommendation performance across devices. To handle social recommendation scenarios, Liu et al. [198] proposed FeSoG, a graph neural network-based federated learning recommender system. To address the challenges of heterogeneity, personalization, and privacy protection, FeSoG employs relational attention and aggregation for handling diverse data and infers user embeddings using local data to retain personalization. Different from the above works, Yuan et al. [199] mainly focused on user privacy and system robustness in federated recommendation systems and introduced federated recommendation unlearning (FRU) as a solution. FRU allows users to withdraw their data contributions and enhances the recommender's resistance to attacks by removing specific users' influence through historical parameter updates.

Apart from recommender systems, healthcare is another important application of federated learning [110]. For instance, Xu et al. [200] introduced a federated learning approach to address the challenges of privacy in diagnosing depression, proposing a multi-view federated learning framework with multi-source data and later fusion methods to handle inconsistent time series data. Similarly, Che et al. [201] addressed the challenges of data privacy and heterogeneity in medical data by preventing leakage in multi-view scenarios. Aiming at the heterogeneous challenge in smart healthcare, Liu et al. [202] presented CAFL, an effective method for impartially assessing participants' contributions to federated learning model performance without compromising their private data. To address label noise challenges in medical imaging federated learning, FedGP [203] provides reliable pseudo labels through noisy graph purification on the client side and utilizing a graph-guided negative ensemble loss for robust supervision against label noise. To address the weakly supervised problem in medical image segmentation, FedDM [204] tackles local drift with collaborative annotation calibration for label correction and global drift with hierarchical gradient de-conflicting for robust gradient aggregation respectively.

Federated learning also finds applications in various and diverse scenarios, such as image steganalysis [50], open banking [205], and mobile keyboard suggestion [1, 11]. Anticipated are broader applications to be practically implemented within the federated setting.

6 Future directions

In recent years, federated learning has seen drastic growth in terms of the amount of research and the breadth of topics. There is still a need for studies on the following promising directions.

6.1 Label scarcity

Current federated learning heavily relies on the supervision signals from sufficient training labels. However, in most real-world applications, clients may not have sufficient labels or lack interaction between users to provide interactive labels. The label scarcity problem makes federated learning impractical in many scenarios. In this case, a potential research direction is to consider the label deficiency while keeping private data on-device. To achieve this research objective, comprehensive investigations into federated learning incorporating semi-supervised learning, transfer learning, few-shot learning, and meta learning are warranted. This holistic approach not only mitigates the impact of label scarcity but also opens avenues for more versatile and adaptive federated learning models that can better accommodate the intricacies of real-world scenarios.

6.2 On-device personalization

Conventionally, personalization is achieved by additional fine-tuning before inference. Recently, more research has focused on personalization. On-device personalization [206] brings forward multiple possible scenarios where clients would additionally benefit from personalization. Mansour et al. [146] formulated their approaches for personalization, including user clustering, data interpolation, and model interpolation. Model-agnostic meta-learning aims to learn quick adaptations and also brings the potential to personalize to individual devices. The studies of effective formulation and metrics to evaluate personalized performance are missed. The underlying essence of personalization and the connections between global model learning and personalized on-device training should be addressed.

6.3 Unsupervised learning

The majority of current research on federated learning mainly follows the supervised or semi-supervised paradigms. Due to the label deficiency problem in the real-world scenario, unsupervised representation learning can be the future direction in the federated setting and other learning problems. By forgoing the need for explicit labels, the unsupervised federated learning methods can autonomously decipher intricate data patterns across distributed datasets. Potential unsupervised techniques include autoencoders, GANs, and clustering algorithms. These approaches enable federated learning systems to extract meaningful features and/or model data manifold without relying on labeled data, addressing the label scarcity issue in real-world scenarios. Furthermore, federated self-supervised learning can also be a promising avenue for overcoming data scarcity issues in federated settings. By leveraging the inherent structures within the data itself, federated self-supervised learning techniques empower devices to learn from their local data without requiring explicit labels from a central server.

6.4 Collaboration of multiple federated paradigms

Federated Learning, as a novel training paradigm, presents numerous new challenges that require attention. In most scenarios, the collaboration of various techniques within the FL framework is necessary. For instance, knowledge distillation shows promising potential in overcoming many challenges through the transfer of abstract knowledge, such as addressing heterogeneity and facilitating multi-task learning. Additionally, exploring the application of transfer learning for knowledge reuse under federated learning is meaningful. This approach can improve data utilization and effectively reduce repeated training in federated scenarios with label scarcity or reinforcement learning. Therefore, we suggest studying how multiple federated learning paradigms can work together to address both existing and new challenges.

6.5 Comprehensive benchmark

Among the numerous federated learning algorithms in the literature, it is evident that federated learning encompasses various parameters, reflecting diverse scenarios, different data distributions of edge side, and various communication frequencies. However, existing research often evaluates these algorithms in different settings, hindering researchers from seeking suitable methods for their specific tasks. Therefore, the establishment of a unified benchmark becomes imperative. There are also some infrastructures to speed up algorithm implementations like FedML [207, 208]. This endeavor aims to inspire greater research in federated learning while providing comprehensive benchmarks adhering to standardized criteria. These benchmarks may encompass real-world deployment scenarios, algorithm comparisons across diverse data environments, and intriguing evaluations of foundation models combined with federated learning.

6.6 Production-level federated learning

In the world of federated learning, it is crucial to shift focus towards making it work effectively in real-world production-level settings. Researchers should aim to improve how federated learning can be used practically. This means finding ways to make it easily fit into existing systems, handle differences between devices, and cope with limitations in communication. It is also important to handle unique real-world challenges such as data distribution drift, diurnal variations, and cold start problems [12]. To address these challenges, the implementation of federated X learning holds significant promise for providing viable solutions. For instance, federated transfer learning proves effective in managing distribution drift, while federated meta learning serves as a valuable tool for addressing cold start problems. In the future, more advanced federated X learning methods specifically designed for production-level applications are expected.

7 Conclusion

This paper conducts a timely and focused survey about federated learning coupled with different learning algorithms. The flexibility of FL was showcased by presenting a wide range of relevant learning paradigms that can be employed within the FL framework. In particular, the compatibility was addressed from the standpoint of how learning algorithms fit the FL architecture and how they take into account two of the critical problems in federated learning: efficient learning and statistical heterogeneity.

Author contributions SJ, YT, TS, ZY, and YL wrote the main manuscript text. SJ, YT, TS, ZY, and YL collect and categorize the reviewed papers. SJ, YL, LV, SP, GL, and A.W. organized the framework and architecture of this survey paper. All authors reviewed the manuscript.

Funding Open Access funding provided by University of Helsinki (including Helsinki University Central Hospital).

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. McMahan HB, Moore E, Ramage D, Hampson S, et al (2017) Communication-efficient learning of deep networks from decentralized data. In: International Conference on artificial intelligence and statistics, pp 1273–1282
2. Wang H, Sreenivasan K, Rajput S, Vishwakarma H, Agarwal S, Sohn J-Y, Lee K, Papailiopoulos D (2020) Attack of the tails: yes, you really can backdoor federated learning. Adv Neural Inf Process Sys 33:16070–16084
3. Bagdasaryan E, Veit A, Hua Y, Estrin D, Shmatikov V (2020) How to backdoor federated learning. In: International conference on artificial intelligence and statistics, PMLR, pp 2938–2948
4. Tolpegin V, Truex S, Gursoy ME, Liu L (2020) Data poisoning attacks against federated learning systems. In: Computer security–ESORICS 2020: 25th European Symposium on research in computer security, ESORICS 2020, Guildford, UK, September 14–18, 2020, Proceedings, Part I 25, Springer, pp 480–501
5. Konečnỳ J, McMahan HB, Yu FX, Richtárik P, Suresh AT, Bacon D (2016) Federated learning: strategies for improving communication efficiency. arXiv preprint arXiv:1610.05492
6. Ji S, Jiang W, Walid A, Li X (2020) Dynamic sampling and selective masking for communication-efficient federated learning. arXiv preprint arXiv:2003.09603
7. Tan Y, Long G, Ma J, Liu L, Zhou T, Jiang J (2022) Federated learning from pre-trained models: a contrastive learning approach. Adv Neural Inf Process Syst 35:19332–19344
8. He C, Tan C, Tang H, Qiu S, Liu J (2019) Central server free federated learning over single-sided trust social networks. arXiv preprint arXiv:1910.04956
9. Yeganeh Y, Farshad A, Navab N, Albarqouni S (2020) Inverse distance aggregation for federated learning with non-iid data. In: DCL workshop at MICCAI, pp 150–159
10. Chen Y, Sun X, Jin Y (2020) Communication-efficient federated deep learning with layerwise asynchronous model update and temporally weighted aggregation. IEEE Trans Neural Netw Learn Syst 31(10):4229–4238
11. Ji S, Pan S, Long G, Li X, Jiang J, Huang Z (2019) Learning private neural language modeling with attentive aggregation. In: International joint conference on neural network
12. Li T, Sahu AK, Talwalkar A, Smith V (2020) Federated learning: challenges, methods, and future directions. IEEE Sign Process Mag 37(3):50–60
13. Briggs C, Fan Z, Andras P (2020) Federated learning with hierarchical clustering of local updates to improve training on non-iid data. In: International joint conference on neural network
14. Yurochkin M, Agarwal M, Ghosh S, Greenewald K, Hoang N, Khazaeni Y (2019) Bayesian nonparametric federated learning of neural networks. In: International conference on machine learning, pp 7252–7261
15. Li T, Sanjabi M, Beirami A, Smith V (2020) Fair resource allocation in federated learning. In: International conference on learning representations
16. Xie Y, Zhang W, Pi R, Wu F, Chen Q, Xie X, Kim S (2022) Robust federated learning against both data heterogeneity and poisoning attack via aggregation optimization. arXiv preprint
17. Xiao J, Du C, Duan Z, Guo W (2021) A novel server-side aggregation strategy for federated learning in non-iid situations. In: 2021 20th International symposium on parallel and distributed computing (ISPDC), IEEE, pp 17–24
18. Liu J, Wang JH, Rong C, Xu Y, Yu T, Wang J (2021) Fedpa: an adaptively partial model aggregation strategy in federated learning. Comput Netw 199:108468
19. Jiang J, Ji S, Long G (2020) Decentralized knowledge acquisition for mobile internet applications. World Wide Web
20. Wu X, Liang Z, Wang J (2020) FedMed: a federated learning framework for language modeling. Sensors 20(14):4048
21. Huang Y, Chu L, Zhou Z, Wang L, Liu J, Pei J, Zhang Y (2021) Personalized cross-silo federated learning on non-iid data. In: AAAI conference on artificial intelligence
22. Wang X, Li R, Wang C, Li X, Taleb T, Leung VC (2020) Attention-weighted federated deep reinforcement learning for device-to-device assisted heterogeneous collaborative edge caching. IEEE J Sel Areas Commun 39(1):154–169
23. Guo Q, Qi Y, Qi S, Wu D, Li Q (2023) Fedmcsa: personalized federated learning via model components self-attention. Neurocomputing 560:126831
24. Zheng K, Liu X, Zhu G, Wu X, Niu J (2022) ChannelFed: enabling personalized federated learning via localized channel attention. In: GLOBECOM 2022-2022 IEEE global communications conference, IEEE, pp 2987–2992
25. Yu FX, Rawat AS, Menon AK, Kumar S (2020) Federated learning with only positive labels. In: International conference on machine learning
26. Li T, Sahu AK, Zaheer M, Sanjabi M, Talwalkar A, Smith V (2020) Federated optimization in heterogeneous networks. In: Conference on machine learning and systems
27. Karimireddy SP, Jaggi M, Kale S, Mohri M, Reddi SJ, Stich SU, Suresh AT (2020) Mime: Mimicking centralized stochastic algorithms in federated learning. arXiv preprint arXiv:2008.03606
28. Durmus AE, Yue Z, Ramon M, Matthew M, Paul W, Venkatesh S (2021) Federated learning based on dynamic regularization. In: International conference on learning representations
29. Kim J, Kim G, Han B (2022) Multi-level branched regularization for federated learning. In: International conference on machine learning, PMLR, pp 11058–11073
30. Cheng A, Wang P, Zhang XS, Cheng J (2022) Differentially private federated learning with local regularization and sparsification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 10122–10131
31. Chen R, Wan Q, Prakash P, Zhang L, Yuan X, Gong Y, Fu X, Pan M (2023) Workie-talkie: accelerating federated learning by overlapping computing and communications via contrastive regularization. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 16999–17009
32. Dinh CT, Vu TT, Tran NH, Dao MN, Zhang H (2022) A new look and convergence rate of federated multitask learning with Laplacian regularization. IEEE Trans Neural Netw Learn Syst. https://doi.org/10.1109/TNNLS.2022.3224252
33. Tan Y, Long G, LIU L, Zhou T, Lu Q, Jiang J, Zhang C (2022) Fedproto: federated prototype learning across heterogeneous clients. Proc AAAI Conf Artif Intell 36(8):8432–8440. https://doi.org/10.1609/aaai.v36i8.20819
34. Ghosh A, Chung J, Yin D, Ramchandran K (2020) An efficient framework for clustered federated learning. In: Advances in neural information processing systems
35. Long G, Xie M, Shen T, Zhou T, Wang X, Jiang J (2023) Multi-center federated learning: clients clustering for better personalization. World Wide Web 26(1):481–500
36. Muhammad K, Wang Q, O'Reilly-Morgan D, Tragos E, Smyth B, Hurley N, Geraci J, Lawlor A (2020) FedFast: going beyond average for faster training of federated recommender systems. In: SIGKDD, pp 1234–1242
37. Dennis DK, Li T, Smith V (2021) Heterogeneity for the win: one-shot federated clustering. In: International conference on machine learning, PMLR, pp 2611–2620
38. Chung J, Lee K, Ramchandran K (2022) Federated unsupervised clustering with generative models. In: AAAI 2022 international workshop on trustable, verifiable and auditable federated learning

39. Cai L, Chen N, Cao Y, He J, Li Y (2023) FedCE: personalized federated learning method based on clustering ensembles. In: Proceedings of the 31st ACM international conference on multimedia, pp 1625–1633
40. Wang H, Yurochkin M, Sun Y, Papailiopoulos D, Khazaeni Y (2020) Federated learning with matched averaging. In: International conference on learning representations
41. Chen H-Y, Chao W-L (2020) FedBE: making bayesian model ensemble applicable to federated learning. In: International conference on learning representations
42. Zhang X, Li Y, Li W, Guo K, Shao Y (2022) Personalized federated learning via variational bayesian inference. In: International conference on machine learning, PMLR, pp 26293–26310
43. Xiao P, Cheng S (2023) Bayesian federated neural matching that completes full information. Proc AAAI Conf Artif Intell 37:10473–10480
44. Mohri M, Sivek G, Suresh AT (2019) Agnostic federated learning. In: International conference on machine learning
45. Ezzeldin Y.H, Yan S, He C, Ferrara E, Avestimehr A. S (2023) Fairfed: Enabling group fairness in federated learning. Proc AAAI Conf Artif Intell 37:7494–7502
46. Lyu L, Xu X, Wang Q, Yu H (2020) Collaborative fairness in federated learning. Federated Learning: privacy and Incentive, pp. 189–204
47. Liu S, Ge Y, Xu S, Zhang Y, Marian A (2022) Fairness-aware federated matrix factorization. In: Proceedings of the 16th ACM conference on recommender systems, pp 168–178
48. Liu Y, Kang Y, Xing C, Chen T, Yang Q (2020) A secure federated transfer learning framework. IEEE Intell Syst 35:70–82
49. Peng X, Huang Z, Zhu Y, Saenko K (2020) Federated adversarial domain adaptation. In: International conference on learning representations
50. Yang H, He H, Zhang W, Cao X (2020) FedSteg: a federated transfer learning framework for secure image steganalysis. IEEE Trans Netw Sci Eng 8(2):1084–1094
51. Wang K, Li J, Wu W et al (2022) An efficient intrusion detection method based on federated transfer learning and an extreme learning machine with privacy preservation. Secur Commun Netw. https://doi.org/10.1155/2022/2913293
52. Chen Y, Qin X, Wang J, Yu C, Gao W (2020) Fedhealth: a federated transfer learning framework for wearable healthcare. IEEE Intell Syst 35(4):83–93
53. Feng S, Li B, Yu H, Liu Y, Yang Q (2022) Semi-supervised federated heterogeneous transfer learning. Knowl-Based Syst 252:109384
54. Jin X, Bu J, Yu Z, Zhang H, Wang Y (2023) FedCrack: federated transfer learning with unsupervised representation for crack detection. IEEE Trans Intell Transp Syst. https://doi.org/10.1109/TITS.2023.3286439
55. Smith V, Chiang C-K, Sanjabi M, Talwalkar AS (2017) Federated multi-task learning. In: Advances in neural information processing systems, pp 4427– 4437
56. Caldas S, Smith V, Talwalkar A (2018) Federated kernelized multi-task learning. In: Conference on machine learning and systems
57. Sattler F, Müller K-R, Samek W (2020) Clustered federated learning: model-agnostic distributed multitask optimization under privacy constraints. IEEE Trans Neural Netw Learn Syst 32(8):3710–3722
58. Cao X, Li Z, Sun G, Yu H, Guizani M (2023) Cross-silo heterogeneous model federated multitask learning. Knowl-Based Syst 265:110347
59. Marfoq O, Neglia G, Bellet A, Kameni L, Vidal R (2021) Federated multi-task learning under a mixture of distributions. Adv Neural Inf Process Syst 34:15434–15447
60. Chen J, Zhang A (2022) FedMSplit: correlation-adaptive federated multi-task learning across multimodal split networks. In: Proceedings of the 28th ACM SIGKDD Conference on knowledge discovery and data mining, pp 87– 96
61. He C, Ceyani E, Balasubramanian K, Annavaram M, Avestimehr S (2022) SpreadGNN: decentralized multi-task federated learning for graph neural networks on molecular data. Proc AAAI Conf Artif Intell 36:6865–6873
62. Yao X, Huang T, Zhang R-X, Li R, Sun L (2019) Federated learning with unbiased gradient aggregation and controllable meta updating. In: Advances in neural information processing systems workshop
63. Fallah A, Mokhtari A, Ozdaglar A (2020) Personalized federated learning with theoretical guarantees: a model-agnostic meta-learning approach. In: Advances in neural information processing systems
64. Wang B, Yuan Z, Ying Y, Yang T (2023) Memory-based optimization methods for model-agnostic meta-learning and personalized federated learning. J Mach Learn Res 24:1–46
65. Lin Y, Ren P, Chen Z, Ren Z, Yu D, Ma J, Rijke Md, Cheng X (2020) Meta matrix factorization for federated rating predictions. In: SIGIR, pp 981– 990
66. Li D, Wang J (2019) FedMD: heterogenous federated learning via model distillation. In: Advances in neural information processing systems workshop
67. He C, Annavaram M, Avestimehr S (2020) Group knowledge transfer: federated learning of large cnns at the edge. Adv Neural Inf Process Syst 33:14068–14080
68. Yang Z, Zhang Y, Zheng Y, Tian X, Peng H, Liu T, Han B (2023) FedFed: feature distillation against data heterogeneity in federated learning. In: Thirty-seventh conference on neural information processing systems
69. Lin T, Kong L, Stich SU, Jaggi M (2020) Ensemble distillation for robust model fusion in federated learning. Adv Neural Inf Process Syst 33:2351–2363
70. Yang Y, Yang R, Peng H, Li Y, Li T, Liao Y, Zhou P (2023) FedACK: federated adversarial contrastive knowledge distillation for cross-lingual and cross-model social bot detection. Proc ACM Web Conf 2023:1314–1323
71. Ma Y, Xie Z, Wang J, Chen K, Shou L (2022) Continual federated learning based on knowledge distillation. In: Raedt LD (ed) Proceedings of the thirty-first international joint conference on artificial intelligence, IJCAI, vol 22, pp 2182–2188. (Main Track.). https://doi.org/10.24963/ijcai.2022/303
72. Wu Z, Sun S, Wang Y, Liu M, Pan Q, Jiang X, Gao B (2023) FedICT: federated multi-task distillation for multi-access edge computing. IEEE Trans Parallel Distrib Syst. https://doi.org/10.1109/TPDS.2023.3289444
73. Zhang Y, Zhang W, Pu L, Lin T, Yan J (2023) To distill or not to distill: towards fast, accurate and communication efficient federated distillation learning. IEEE Internet of Things J. https://doi.org/10.1109/JIOT.2023.3324666
74. Zhang L, Shen L, Ding L, Tao D, Duan L-Y (2022) 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, pp 10174–10183
75. Jeong W, Yoon J, Yang E, Hwang SJ (2021) Federated semi-supervised learning with inter-client consistency & disjoint learning. In: International conference on learning representations
76. Papernot N, Abadi M, Erlingsson Ú, Goodfellow I, Talwar K (2017) Semi-supervised knowledge transfer for deep learning from private training data. In: International conference on learning representations
77. Diao E, Ding J, Tarokh V (2022) Semifl: semi-supervised federated learning for unlabeled clients with alternate training. Adv Neural Inf Process Syst 35:17871–17884

78. Jiang M, Yang H, Li X, Liu Q, Heng P-A, Dou Q (2022) Dynamic bank learning for semi-supervised federated image diagnosis with class imbalance. In: International conference on medical image computing and computer-assisted intervention, Springer, pp 196–206
79. Wei X-X, Huang H (2023) Balanced federated semi-supervised learning with fairness-aware pseudo-labeling. IEEE Trans Neural Netw Learn Syst. https://doi.org/10.1109/TNNLS.2022.3233093
80. Liang X, Lin Y, Fu H, Zhu L, Li, X (2022) Rscfed: random sampling consensus federated semi-supervised learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 10154–10163
81. Li M, Li Q, Wang Y (2023) Class balanced adaptive pseudo labeling for federated semi-supervised learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 16292–16301
82. Shang X, Huang G, Lu Y, Lou J, Han B, Cheung Y-m, Wang H (2023) Federated semi-supervised learning with annotation heterogeneity. arXiv preprint arXiv:2303.02445
83. Kang Y, Liu Y, Liang X (2022) Fedcvt: semi-supervised vertical federated learning with cross-view training. ACM Trans Intell Syst Technol (TIST) 13(4):1–16
84. Fan C, Liu P (2020) Federated generative adversarial learning. arXiv preprint arXiv:2005.03793
85. Rasouli M, Sun T, Rajagopal R (2020) FedGAN: federated generative adversarial networks for distributed data. arXiv preprint arXiv:2006.07228
86. Augenstein S, McMahan HB, Ramage D, Ramaswamy S, Kairouz P, Chen M, Mathews R, Arcas BA (2020) Generative models for effective ml on private, decentralized datasets. In: International conference on learning representations
87. Qi T, Wu F, Wu C, Lyu L, Xu T, Liao H, Yang Z, Huang Y, Xie X (2022) Fairvfl: a fair vertical federated learning framework with contrastive adversarial learning. Adv Neural Inf Process Syst 35:7852–7865
88. Li X, Song Z, Yang J (2023) Federated adversarial learning: a framework with convergence analysis. In: International conference on machine learning, PMLR, pp 19932–19959
89. Zhang J, Li B, Chen C, Lyu L, Wu S, Ding S, Wu C (2023) Delving into the adversarial robustness of federated learning. In: Proceedings of the AAAI conference on artificial intelligence
90. Chen C, Liu Y, Ma X, Lyu L (2022) Calfat: calibrated federated adversarial training with label skewness. Adv Neural Inf Process Syst 35:3569–3581
91. Hong J, Wang H, Wang Z, Zhou J (2023) Federated robustness propagation: sharing adversarial robustness in heterogeneous federated learning. Proc AAAI Conf Artif Intell 37:7893–7901
92. Bram Bv, Saeed A, Ozcelebi T (2020) Towards federated unsupervised representation learning. In: ACM EdgeSys, pp 31–36
93. Grammenos A, Mendoza Smith R, Crowcroft J, Mascolo C (2020) Federated principal component analysis. In: Advances in neural information processing systems
94. Zhang F, Kuang K, Chen L, You Z, Shen T, Xiao J, Zhang Y, Wu C, Wu F, Zhuang Y et al (2023) Federated unsupervised representation learning. Front Inf Technol Electron Eng 24(8):1181–1193
95. Zhuang W, Wen Y, Zhang S (2022) Divergence-aware federated self-supervised learning. In: International conference on learning representations
96. Lubana E, Tang CI, Kawsar F, Dick R, Mathur A (2022) Orchestra: unsupervised federated learning via globally consistent clustering. In: International conference on machine learning, PMLR, pp 14461–14484
97. Rehman YAU, Gao Y, Gusmao PPB, Alibeigi M, Shen J, Lane ND (2023) L-DAWA: layer-wise divergence aware weight aggregation in federated self-supervised visual representation learning. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 16464–16473
98. Han S, Park S, Wu F, Kim S, Wu C, Xie X, Cha M(2022) Fedx: unsupervised federated learning with cross knowledge distillation. In: European conference on computer vision, Springer, pp 691–707
99. Zhuo HH, Feng W, Xu Q, Yang Q, Lin Y (2019) Federated deep reinforcement learning. arXiv preprint arXiv:1901.08277
100. Wang H, Kaplan Z, Niu D, Li B (2020) Optimizing federated learning on non-IID data with reinforcement learning. In: IEEE international conference on computer communications, IEEE, pp 1698–1707
101. Cha H, Park J, Kim H, Bennis M, Kim S-L (2020) Proxy experience replay: federated distillation for distributed reinforcement learning. IEEE Intell Syst 35(4):94–101
102. Zhan Y, Zhang J (2020) An incentive mechanism design for efficient edge learning by deep reinforcement learning approach. In: IEEE international conference on computer communications, IEEE, pp 2489–2498
103. Khodadadian S, Sharma P, Joshi G, Maguluri ST (2022) Federated reinforcement learning: linear speedup under markovian sampling. In: International conference on machine learning, PMLR, pp 10997–11057
104. Jin H, Peng Y, Yang W, Wang S, Zhang Z (2022) Federated reinforcement learning with environment heterogeneity. In: International conference on artificial intelligence and statistics, PMLR, pp 18–37
105. Mai W, Yao J, Chen G, Zhang Y, Cheung Y-M, Han B (2023) Server-client collaborative distillation for federated reinforcement learning. ACM Trans Knowl Discov Data 18(1):1–22
106. Fan FX, Ma Y, Dai Z, Tan C, Low BKH (2023) Fedhql: federated heterogeneous q-learning. In: Proceedings of the 2023 international conference on autonomous agents and multiagent systems, pp 2810–2812
107. Yang Q, Liu Y, Chen T, Tong Y (2019) Federated machine learning: concept and applications. ACM Trans Intell Syst Technol 10(2):12
108. Li Q, Wen Z, He B (2019) Federated learning systems: vision, hype and reality for data privacy and protection. arXiv preprint arXiv:1907.09693
109. Kairouz P, McMahan HB, Avent B, Bellet A, Bennis M, Bhagoji AN, Bonawitz K, Charles Z, Cormode G, Cummings R, et al (2019) Advances and open problems in federated learning. arXiv preprint arXiv:1912.04977
110. Xu J, Wang F (2019) Federated learning for healthcare informatics. arXiv preprint arXiv:1911.06270
111. Lyu L, Yu H, Yang Q (2020) Threats to federated learning: a survey. arXiv preprint arXiv:2003.02133
112. Lim WYB, Luong NC, Hoang DT, Jiao Y, Liang Y-C, Yang Q, Niyato D, Miao C (2020) Federated learning in mobile edge networks: a comprehensive survey. IEEE Commun Surv Tutor 22(3):2031–2063
113. Niknam S, Dhillon HS, Reed JH (2020) Federated learning for wireless communications: motivation, opportunities, and challenges. IEEE Commun Mag 58(6):46–51
114. Jin Y, Wei X, Liu Y, Yang Q (2020) A survey towards federated semi-supervised learning. arXiv preprint arXiv:2002.11545
115. Lo SK, Lu Q, Wang C, Paik H, Zhu L (2020) A systematic literature review on federated machine learning: from a software engineering perspective. arXiv preprint arXiv:2007.11354
116. Li X, Huang K, Yang W, Wang S, Zhang Z (2020) On the convergence of Fedavg on non-iid data. In: International conference on learning representations
117. Karimireddy SP, Kale S, Mohri M, Reddi SJ, Stich SU, Suresh AT (2020) Scaffold: stochastic controlled averaging for federated

learning. In: International conference on machine learning, pp 5132–5143

118. Reddi S, Charles Z, Zaheer M, Garrett Z, Rush K, Konečnỳ J, Kumar S, McMahan HB (2021) Adaptive federated optimization. In: International conference on learning representations
119. Singh SP, Jaggi M (2020) Model fusion via optimal transport. Adv Neural Inf Process Syst 33:22045–22055
120. Chai D, Wang L, Chen K, Yang Q (2020) Secure federated matrix factorization. IEEE Intell Syst 36(5):11–20
121. Tan Y, Chen C, Zhuang W, Dong X, Lyu L, Long G (2023) Is heterogeneity notorious? taming heterogeneity to handle test-time shift in federated learning. In: Thirty-seventh conference on neural information processing systems
122. Alawad M, Yoon H-J, Gao S, Mumphrey B, Wu X-C, Durbin EB, Jeong JC, Hands I, Rust D, Coyle L et al (2020) Privacy-preserving deep learning nlp models for cancer registries. IEEE Trans Emerg Top Comput 9(3):1219–1230
123. Tang Z, Zhang Y, Shi S, He X, Han B, Chu X (2022) Virtual homogeneity learning: defending against data heterogeneity in federated learning. In: International conference on machine learning, PMLR, pp 21111–21132
124. Lu Y, Chen L, Zhang Y, Zhang Y, Han B, Cheung Y-m, Wang H (2023) Federated learning with extremely noisy clients via negative distillation. arXiv preprint arXiv:2312.12703
125. Finn C, Abbeel P, Levine S (2017) Model-agnostic meta-learning for fast adaptation of deep networks. In: International conference on machine learning, pp 1126– 1135
126. Ji S, Long G, Pan S, Zhu T, Jiang J, Wang S, Li X(2019) Knowledge transferring via model aggregation for online social care. arXiv preprint arXiv:1905.07665
127. Jiang Y, Konečnỳ J, Rush K, Kannan S (2019) Improving federated learning personalization via model agnostic meta learning. In: Advances in neural information processing systems workshop
128. Lin S, Yang L, He Z, Fan D, Zhang J (2021) Metagater: fast learning of conditional channel gated networks via federated meta-learning. In: 2021 IEEE 18th international conference on mobile Ad Hoc and smart systems (MASS), IEEE, pp 164– 172
129. Sohn K, Berthelot D, Carlini N, Zhang Z, Zhang H, Raffel CA, Cubuk ED, Kurakin A, Li C-L (2020) Fixmatch: simplifying semi-supervised learning with consistency and confidence. Adv Neural Inf Process Syst 33:596–608
130. Yang X, Song Z, King I, Xu Z (2022) A survey on deep semi-supervised learning. IEEE Trans Knowl Data Eng 35(9):8934–8954
131. He K, Fan H, Wu Y, Xie S, Girshick R (2020) Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 9729–9738
132. Chen X, Fan H, Girshick R, He K (2020) Improved baselines with momentum contrastive learning. arXiv preprint arXiv:2003.04297
133. Grill J-B, Strub F, Altché F, Tallec C, Richemond P, Buchatskaya E, Doersch C, Avila Pires B, Guo Z, Gheshlaghi Azar M et al (2020) Bootstrap your own latent-a new approach to self-supervised learning. Adv Neural Inf Process Syst 33:21271–21284
134. Chen T, Kornblith S, Norouzi M, Hinton G (2020) A simple framework for contrastive learning of visual representations. In: International conference on machine learning, PMLR, pp 1597–1607
135. Chen X, He K (2021) Exploring simple siamese representation learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 15750–15758
136. Watkins CJ, Dayan P (1992) Q-learning. Mach Learn 8:279–292
137. Sutton RS (1988) Learning to predict by the methods of temporal differences. Mach Learn 3:9–44
138. Lattimore T, Szepesvári C (2020) Bandit algorithms
139. Khalatbarisoltani A, Boulon L, Hu X (2023) Integrating model predictive control with federated reinforcement learning for decentralized energy management of fuel cell vehicles. IEEE Trans Intell Transp Syst. https://doi.org/10.1109/TITS.2023.3303991
140. Qiu D, Xue J, Zhang T, Wang J, Sun M (2023) Federated reinforcement learning for smart building joint peer-to-peer energy and carbon allowance trading. Appl Energy 333:120526
141. Zhang Z, Jiang Y, Shi Y, Shi Y, Chen W (2022) Federated reinforcement learning for real-time electric vehicle charging and discharging control. In: 2022 IEEE Globecom workshops (GC Wkshps), IEEE, pp 1717–1722
142. Arivazhagan MG, Aggarwal V, Singh AK, Choudhary S (2019) Federated learning with personalization layers. arXiv: 1912.00818 [cs.LG]
143. Liang PP, Liu T, Ziyin L, Salakhutdinov R, Morency L-P (2020) Think locally, act globally: federated learning with local and global representations. Adv Neural Inf Process Syst
144. Deng Y, Kamani MM, Mahdavi M (2020) Adaptive personalized federated learning. arXiv:2003:13461
145. Tan AZ, Yu H, Cui L, Yang Q (2021) Towards personalized federated learning. arXiv preprint arXiv:2103.00710
146. Mansour Y, Mohri M, Ro J, Suresh AT (2020) Three approaches for personalization with applications to federated learning. arXiv preprint arXiv:2002.10619
147. Li Q, He B, Song D (2021) Model-contrastive federated learning. arXiv: 2103.16257 [cs.LG]
148. Liu Y, Pan S, Jin M, Zhou C, Xia F, Yu PS (2021) Graph self-supervised learning: a survey. arXiv preprint arXiv:2103.00111
149. Yang Y, Guan Z, Li J, Zhao W, Cui J, Wang Q (2021) Interpretable and efficient heterogeneous graph convolutional network. IEEE Trans Knowl Data Eng 35(2):1637–1650
150. Jeong E, Oh S, Kim H, Park J, Bennis M, Kim S-L (2018) Communication-efficient on-device machine learning: Federated distillation and augmentation under non-IID private data. In: Advances in neural information processing systems
151. Long G, Shen T, Tan Y, Gerrard L, Clarke A, Jiang J (2021) Federated learning for privacy-preserving open innovation future on digital health. In: Humanity driven AI: productivity, well-being, sustainability and partnership, pp 113–133
152. Zhu H, Zhang H, Jin Y (2020) From federated learning to federated neural architecture search: a survey. Complex Intell Syst 7(2):639–657
153. He C, Annavaram M, Avestimehr S (2020) FedNAS: federated deep learning via neural architecture search. In: Proceedings of the IEEE conference on computer vision and pattern recognition
154. Singh I, Zhou H, Yang K, Ding M, Lin B, Xie P (2020) Differentially-private federated neural architecture search. In: FL-international conference on machine learning workshop
155. Hoang M, Hoang N, Low BKH, Kingsford C (2019) Collective model fusion for multiple black-box experts. In: International conference on machine learning, PMLR, pp 2742–2750
156. Liu R, Cao Y, Chen H, Guo R, Yoshikawa M (2021) Flame: differentially private federated learning in the shuffle model. Proc AAAI Conf Artif Intell 35(10):8688–8696. https://doi.org/10.1609/aaai.v35i10.17053
157. Thapa C, Mahawaga Arachchige PC, Camtepe S, Sun L (2022) Splitfed: when federated learning meets split learning. Proc AAAI Conf Artif Intell 36(8):8485–8493. https://doi.org/10.1609/aaai.v36i8.20825
158. Zawad S, Ali A, Chen P-Y, Anwar A, Zhou Y, Baracaldo N, Tian Y, Yan F (2021) Curse or redemption? how data heterogeneity affects the robustness of federated learning. Proc AAAI Conf

Artif Intell 35(12):10807–10814. https://doi.org/10.1609/aaai.v35i12.17291

159. Ozdayi MS, Kantarcioglu M, Gel YR (2021) Defending against backdoors in federated learning with robust learning rate. Proc AAAI Conf Artif Intell 35(10):9268–9276. https://doi.org/10.1609/aaai.v35i10.17118
160. Zhao B, Sun P, Wang T, Jiang K (2022) Fedinv: byzantine-robust federated learning by inversing local model updates. Proc AAAI Conf Artif Intell 36(8):9171–9179. https://doi.org/10.1609/aaai.v36i8.20903
161. Zhang Z, Panda A, Song L, Yang Y, Mahoney M, Mittal P, Kannan R, Gonzalez J (2022) Neurotoxin: durable backdoors in federated learning. In: Proceedings of the 39th international conference on machine learning, vol 162, pp 26429–26446
162. Cao X, Jia J, Gong NZ (2021) Provably secure federated learning against malicious clients. Proc AAAI Conf Artif Intell 35(8):6885–6893. https://doi.org/10.1609/aaai.v35i8.16849
163. Wen Y, Geiping JA, Fowl L, Goldblum M, Goldstein T (2022) In: Chaudhuri K, Jegelka S, Song L, Szepesvari C, Niu G, Sabato S (eds) Proceedings of the 39th international conference on machine learning. Proceedings of machine learning research. Fishing for user data in large-batch federated learning via gradient magnification, vol 162, pp 23668–23684. https://proceedings.mlr.press/v162/wen22a.html
164. Gupta S, Huang Y, Zhong Z, Gao T, Li K, Chen D (2022) Recovering private text in federated learning of language models. Adv Neural Inf Process Syst 35:8130–8143
165. Bietti A, Wei C-Y, Dudik M, Langford J, Wu S (2022) Personalization improves privacy-accuracy tradeoffs in federated learning. In: Chaudhuri K, Jegelka S, Song L, Szepesvari C, Niu G, Sabato S (eds) Proceedings of the 39th international conference on machine learning. Proceedings of machine learning research, vol 162, pp 1945–1962. https://proceedings.mlr.press/v162/bietti22a.html
166. Zhang X, Chen X, Hong M, Wu S, Yi J (2022) Understanding clipping for federated learning: convergence and client-level differential privacy. In: Chaudhuri K, Jegelka S, Song L, Szepesvari C, Niu G, Sabato S (eds) Proceedings of the 39th international conference on machine learning. Proceedings of machine learning research, vol 162, pp 26048–26067. https://proceedings.mlr.press/v162/zhang22b.html
167. Hu R, Gong Y, Guo Y (2021) Federated learning with sparsification-amplified privacy and adaptive optimization. In: Zhou Z-H (ed) Proceedings of the thirtieth international joint conference on artificial intelligence, IJCAI-21, pp 1463–1469 . Main Track. https://doi.org/10.24963/ijcai.2021/202
168. Sun L, Qian J, Chen X (2021) LDP-FL: practical private aggregation in federated learning with local differential privacy. In: Zhou Z-H (ed) Proceedings of the Thirtieth international joint conference on artificial intelligence, IJCAI-21, pp 1571–1578 . Main Track. https://doi.org/10.24963/ijcai.2021/217
169. Peng H, Li H, Song Y, Zheng V, Li J (2021) Differentially private federated knowledge graphs embedding. In: Proceedings of the 30th ACM international conference on information & knowledge management, pp 1416–1425
170. Fan K, Hong J, Li W, Zhao X, Li H, Yang Y (2023) Flsg: a novel defense strategy against inference attacks in vertical federated learning. IEEE Internet of Things J. https://doi.org/10.1109/JIOT.2023.3302792
171. Rong D, He Q, Chen J (2022) Poisoning deep learning based recommender model in federated learning scenarios. In: Raedt LD (ed) Proceedings of the thirty-first international joint conference on artificial intelligence, IJCAI-22, pp 2204– 2210 . Main Track. https://doi.org/10.24963/ijcai.2022/306
172. Huang Y, Gupta S, Song Z, Li K, Arora S (2021) Evaluating gradient inversion attacks and defenses in federated learning. In: Ranzato M, Beygelzimer A, Dauphin Y, Liang PS, Vaughan JW (eds) Advances in neural information processing systems, vol 34, pp 7232–7241. https://proceedings.neurips.cc/paper/2021/file/3b3fff6463464959dcd1b68d0320f781-Paper.pdf
173. Jin X, Chen P-Y, Hsu C-Y, Yu C-M, Chen T (2021) CAFE: catastrophic data leakage in vertical federated learning. In: Ranzato M, Beygelzimer A, Dauphin Y, Liang PS, Vaughan JW (eds) Advances in neural information processing systems, vol 34, pp 994–1006. https://proceedings.neurips.cc/paper/2021/file/08040837089cdf46631a10aca5258e16-Paper.pdf
174. Sun J, Li A, DiValentin L, Hassanzadeh A, Chen Y, Li H (2021) FL-WBC: enhancing robustness against model poisoning attacks in federated learning from a client perspective. In: Ranzato M, Beygelzimer A, Dauphin Y, Liang PS, Vaughan JW (eds) Advances in neural information processing systems, vol 34, pp 12613– 12624. https://proceedings.neurips.cc/paper/2021/file/692baebec3bb4b53d7ebc3b9fabac31b-Paper.pdf
175. Park S, Han S, Wu F, Kim S, Zhu B, Xie X, Cha M (2023) Feddefender: client-side attack-tolerant federated learning. In: Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining, pp 1850–1861
176. Park J, Han D-J, Choi M, Moon J (2021) Sageflow: robust federated learning against both stragglers and adversaries. In: Ranzato M, Beygelzimer A, Dauphin Y, Liang PS, Vaughan JW (eds) Advances in neural information processing systems, vol 34, pp 840–851. https://proceedings.neurips.cc/paper/2021/file/076a8133735eb5d7552dc195b125a454-Paper.pdf
177. Agarwal N, Kairouz P, Liu Z (2021) The skellam mechanism for differentially private federated learning. In: Ranzato M, Beygelzimer A, Dauphin Y, Liang PS, Vaughan JW (eds) Advances in neural information processing systems, vol 34, pp 5052–5064. https://proceedings.neurips.cc/paper/2021/file/285baacbdf8fda1de94b19282acd23e2-Paper.pdf
178. Chang Y, Zhang K, Gong J, Qian H (2023) Privacy-preserving federated learning via functional encryption, revisited. IEEE Trans Inf Forens Secur 18:1855–1869
179. Hijazi NM, Aloqaily M, Guizani M, Ouni B, Karray F (2023) Secure federated learning with fully homomorphic encryption for IoT communications. IEEE Internet of Things J. https://doi.org/10.1109/JIOT.2023.3302065
180. Zhao P, Cao Z, Jiang J, Gao F (2022) Practical private aggregation in federated learning against inference attack. IEEE Internet Things J 10(1):318–329
181. Gao H, Xu A, Huang H (2021) On the convergence of communication-efficient local sgd for federated learning. Proc AAAI Conf Artif Intel 35(9):7510–7518. https://doi.org/10.1609/aaai.v35i9.16920
182. Wang Y, Lin L, Chen J (2022) Communication-efficient adaptive federated learning. In: Chaudhuri K, Jegelka S, Song L, Szepesvari C, Niu G, Sabato S (eds) Proceedings of the 39th international conference on machine learning. Proceedings of machine learning research, vol 162, pp 22802–22838 . https://proceedings.mlr.press/v162/wang22o.html
183. Tang Z, Shi S, Li B, Chu X (2022) Gossipfl: a decentralized federated learning framework with sparsified and adaptive communication. IEEE Trans Parallel Distrib Syst 34(3):909–922
184. Yi L, Gang W, Xiaoguang L (2022) QSFL: a two-level uplink communication optimization framework for federated learning. In: Chaudhuri K, Jegelka S, Song L, Szepesvari C, Niu G, Sabato S (eds) Proceedings of the 39th international conference on machine learning. Proceedings of machine learning research, vol 162, pp 25501–25513. https://proceedings.mlr.press/v162/yi22a.html

185. Zhu Z, Hong J, Drew S, Zhou J (2022) Resilient and communication efficient learning for heterogeneous federated systems. Proc Mach Learn Res 162:27504
186. Yapp AZH, Koh HSN, Lai YT, Kang J, Li X, Ng JS, Jiang H, Lim WYB, Xiong Z, Niyato D (2021) Communication-efficient and scalable decentralized federated edge learning. In: Zhou Z-H (ed) Proceedings of the thirtieth international joint conference on artificial intelligence, IJCAI-21, pp 5032– 5035 . https://doi.org/10.24963/ijcai.2021/720 . Demo Track. https://doi.org/10.24963/ijcai.2021/720
187. Zhu L, Lin H, Lu Y, Lin Y, Han S (2021) Delayed gradient averaging: tolerate the communication latency for federated learning. In: Ranzato M, Beygelzimer A, Dauphin Y, Liang PS, Vaughan JW (eds) Advances in neural information processing systems, vol 34, pp 29995–30007. https://proceedings.neurips.cc/paper/2021/file/fc03d48253286a798f5116ec00e99b2b-Paper.pdf
188. Isik B, Pase F, Gunduz D, Weissman T, Michele Z: Sparse random networks for communication-efficient federated learning. In: The Eleventh international conference on learning representations (2022)
189. Wang H-P, Stich S, He Y, Fritz M (2022) Progfed: effective, communication, and computation efficient federated learning by progressive training. In: International conference on machine learning, PMLR, pp 23034–23054
190. Li C, Wang H (2022) Communication efficient federated learning for generalized linear bandits. Adv Neural Inf Process Syst 35:38411–38423
191. Sun Z, Wei E (2022) A communication-efficient algorithm with linear convergence for federated minimax learning. Adv Neural Inf Process Syst 35:6060–6073
192. Cui Y, Cao K, Zhou J, Wei T (2022) Optimizing training efficiency and cost of hierarchical federated learning in heterogeneous mobile-edge cloud computing. IEEE Tran Comput-Aid Des Integr Circuits Syst. https://doi.org/10.1109/TCAD.2022.3205551
193. Tang Z, Wang Y, He X, Zhang L, Pan X, Wang Q, Zeng R, Zhao K, Shi S, He B, et al (2023) Fusionai: decentralized training and deploying llms with massive consumer-level gpus. arXiv preprint arXiv:2309.01172
194. Tan Y, Liu Y, Long G, Jiang J, Lu Q, Zhang C (2023) Federated learning on non-iid graphs via structural knowledge sharing. Proc AAAI Conf Artif Intel 37:9953–9961
195. Pan Q, Zhu Y (2022) Fedwalk: communication efficient federated unsupervised node embedding with differential privacy. In: Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining, pp 1317–1326
196. Liang F, Pan W, Ming Z (2021) Fedrec++: Lossless federated recommendation with explicit feedback. Proc AAAI Conf Artif Intel 35:4224–4231
197. Zhu Z, Si S, Wang J, Xiao J (2022) Cali3f: calibrated fast fair federated recommendation system. In: 2022 international joint conference on neural networks (IJCNN), IEEE, pp 1–8
198. Liu Z, Yang L, Fan Z, Peng H, Yu PS (2022) Federated social recommendation with graph neural network. ACM Trans Intell Syst Technol (TIST) 13(4):1–24
199. Yuan W, Yin H, Wu F, Zhang S, He T, Wang H (2023) Federated unlearning for on-device recommendation. In: Proceedings of the sixteenth ACM international conference on web search and data mining, pp 393–401
200. Xu X, Peng H, Bhuiyan MZA, Hao Z, Liu L, Sun L, He L (2021) Privacy-preserving federated depression detection from multisource mobile health data. IEEE Trans Industr Inf 18(7):4788–4797
201. Che S, Kong Z, Peng H, Sun L, Leow A, Chen Y, He L (2022) Federated multi-view learning for private medical data integration and analysis. ACM Trans Intell Syst Technol (TIST) 13(4):1–23
202. Liu Z, Chen Y, Zhao Y, Yu H, Liu Y, Bao R, Jiang J, Nie Z, Xu Q, Yang Q (2022) Contribution-aware federated learning for smart healthcare. Proc AAAI Conf Artif Intel 36:12396–12404
203. Chen Z, Li W, Xing X, Yuan Y (2023) Medical federated learning with joint graph purification for noisy label learning. Med Image Anal 90:102976
204. Zhu M, Chen Z, Yuan Y (2023) FedDM: federated weakly supervised segmentation via annotation calibration and gradient deconflicting. IEEE Trans Med Imag. https://doi.org/10.1109/TMI.2023.3235757
205. Long G, Tan Y, Jiang J, Zhang C (2020) Federated learning for open banking. In: Federated learning: privacy and incentive, pp 240–254
206. Wang K, Mathews R, Kiddon C, Eichner H, Beaufays F, Ramage D (2019) Federated evaluation of on-device personalization. arXiv preprint arXiv:1910.10252
207. He C, Li S, So J, Zhang M, Wang H, Wang X, Vepakomma P, Singh A, Qiu H, Shen L, Zhao P, Kang Y, Liu Y, Raskar R, Yang Q, Annavaram M, Avestimehr S (2020) FedML: a research library and benchmark for federated machine learning. arXiv preprint arXiv:2007.13518
208. Tang Z, Chu X, Ran RY, Lee S, Shi S, Zhang Y, Wang Y, Liang AQ, Avestimehr S, He C (2023) Fedml parrot: a scalable federated learning system via heterogeneity-aware scheduling on sequential and hierarchical training. arXiv preprint arXiv:2303.01778

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.