FedMental: Topology-Aware Federated Prototype Learning for Polymorphic Multimodal Psychiatry

Haoyu Li[1,*], He Li[1,*], Wenke Huang[2], Yujing Rao[1], Xiaofen Zong[3,†], and Mang Ye[1,4,†]

[1] School of Computer Science, Wuhan University, Wuhan, China
[2] Nanyang Technological University, Singapore
[3] Renmin Hospital of Wuhan University, Wuhan, China
[4] Taikang Center for Life and Medical Sciences, Wuhan University, Wuhan, China

Abstract. Psychiatric diagnosis is moving from subjective clinical observation toward objective multimodal assessment using audio, visual, and textual signals. Federated learning offers a natural framework for multi-center collaboration without centralizing sensitive patient data, but psychiatric data introduce a form of non-IID heterogeneity that is not captured by standard label or domain skew: patients with the same diagnosis can exhibit highly polymorphic manifestations, forming multi-peak within-class distributions. Existing prototype-based FL methods usually aggregate client or class features into coarse centroids, which can collapse distinct symptom modes into semantically ambiguous prototypes. We propose **FedMental**, a topology-aware federated prototype framework for multi-center multimodal psychiatry. FedMental first preserves subsampled sample-level embeddings before premature client-mean aggregation, then applies **Topology-Aware Prototype Refinement (TAPR)** on the server to discover class-view latent modes through parameter-free clustering. On the client side, **Intra-View Mode Alignment (IVMA)** softly anchors local representations to matched mode prototypes, while **Cross-View Semantic Contrast (CVSC)** promotes view-invariant semantics across modalities and centers. Experiments on the real-world Psycare psychiatric benchmark and two standard heterogeneous FL benchmarks show that FedMental consistently improves over strong FL baselines, achieving +8.83% over FedAvg on Psycare, +11.42% on Office-Caltech, and +4.41% on Digits.

Keywords: Federated Learning · Multimodal Psychiatry

1 Introduction

Psychiatric diagnosis has predominantly relied on standardized clinical scales and the subjective observation of practitioners [3,30]. While foundational, these traditional heuristic approaches are inherently susceptible to recall bias and

† Corresponding Author
{zongxiaofen, yemang}@whu.edu.cn

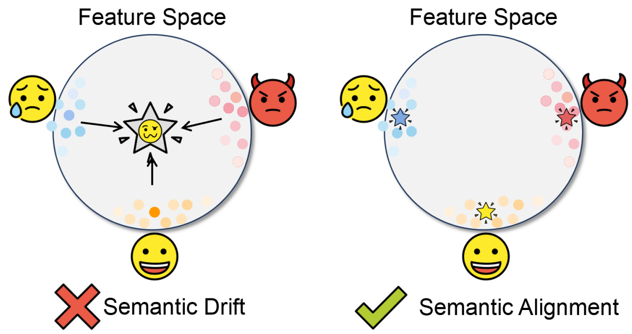


Fig. 1: Topology-Awareness vs. Coarse Aggregation. Psychiatric disorders natively exhibit multi-peak feature distributions due to polymorphic manifestations. For instance, according to DSM-5 criteria, Major Depressive Disorder (MDD) can manifest divergently within the same diagnostic category. **Left:** Traditional coarse-grained aggregation averages these clinically conflicting signals into a single generic chimera prototype. This mathematically neutralizes critical diagnostic features (e.g., averaging agitation and retardation), causing severe semantic drift. **Right:** Our proposed topology-aware approach preserves these intrinsic latent semantic modes. By maintaining multiple mode prototypes, it successfully preserves the true manifold structure and enables precise, manifestation-specific semantic alignment.

inter-rater variability [48]. Consequently, psychiatric assessment is undergoing a critical paradigm shift toward objective, data-driven analysis. Clinical signals derived from patient interviews, such as spoken semantics, acoustic prosody, and visual facial expressions, serve as objective and complementary biomarkers that are essential for accurate diagnosis [4]. Recent advances in artificial intelligence have demonstrated remarkable potential in automating this process, offering scalable screening tools that rival human expert performance when trained on large-scale datasets [54]. However, the highly sensitive nature of mental health data imposes strict privacy barriers. Clinical records are naturally fragmented across isolated medical centers, governed by rigorous regulations like HIPAA [55] and GDPR [13] that strictly preclude centralized data aggregation.

Federated Learning (FL) [40, 42] has emerged as a compelling solution, enabling collaborative model training while keeping raw patient data entirely localized. Yet, deploying FL in real-world psychiatry faces a unique and severe hurdle: *extreme data heterogeneity*, often referred to as the non-IID data quagmire [25], originating from disparate data *views*. In our framework, we define a view to encompass both varied sensory inputs (e.g., audio, video, text) and distinct hospital acquisition protocols. Unlike standard object recognition tasks where classes are relatively compact, psychiatric disorders exhibit profound polymorphic manifestations [2, 14]. Patients sharing the same diagnosis often display vastly different symptom clusters. Some individuals present as agitated and anxious, whereas others appear retarded and melancholic. In the high-dimensional feature space, this translates to a complex multi-peak topological structure, where a single class naturally splits into multiple disjoint *latent semantic modes*.

Existing prototype-based FL methods [27,51] suffer from a fundamental theoretical flaw when confronting this specific heterogeneity: **coarse-grained mean aggregation**. By mathematically averaging diverse, cross-view representations into a single global centroid, they implicitly impose a rigid unimodal Gaussian assumption upon the feature space. This oversimplification forcefully collapses distinct, well-separated symptom subtypes into a distorted chimera representation (as visualized in Fig. 1). Consequently, local models are repeatedly subjected to conflicting gradient signals during optimization, leading to severe representation collapse and blurred decision boundaries. Such semantic drift is particularly destructive in multi-center psychiatry, where preserving subtle, fine-grained behavioral nuances is critical for accurate differential diagnosis.

To bridge the critical gap between coarse algorithmic assumptions and fine-grained clinical reality, we propose **FedMental**, a topology-aware federated prototype framework for multi-center multi-view psychiatric diagnosis. Respecting mental health data's intrinsic polymorphism, our server-side **Topology-Aware Prototype Refinement (TAPR)** uses parameter-free unsupervised clustering to dynamically identify and preserve multi-peak structures for each class–view pair, without rigid prior assumptions about underlying subtype counts. This effectively ensures the global prototype library faithfully reflects symptom structural diversity, preventing critical diagnostic information dilution.

To leverage fine-grained global prototypes under strict privacy constraints, FedMental introduces a dual-objective alignment strategy balancing both view-specific stability and cross-view generalization. First, **Intra-View Mode Alignment (IVMA)** uses dynamic soft-assignment to anchor local features to their matched semantic mode (not a generic class mean), stabilizing local training against center-specific shifts. Second, inspired by advanced self-supervised learning [8,23], **Cross-View Semantic Contrast (CVSC)** uses an InfoNCE objective to bridge cross-view semantic gaps, enforcing positive cross-view pair consistency and widening inter-class margins via negative repulsion. This synergistic closed-loop design lets local models learn more view-invariant representations while preserving fine-grained sensitivity to clinical heterogeneity.

Our main contributions are summarized as follows:

1. We identify **polymorphic manifestation** as a critical bottleneck in psychiatric FL, revealing that traditional coarse-grained aggregation collapses multi-peak symptom structures into distorted chimera prototypes and causes severe representation collapse.
2. We propose **FedMental**, a topology-aware federated prototype framework, featuring a server-side **TAPR** module to preserve global topological diversity via parameter-free clustering, and a dual-objective client strategy (**IVMA** & **CVSC**) to dynamically match local features to fine-grained semantic modes.
3. We validate FedMental on the real-world, multi-center Psycare benchmark [46], achieving substantial gains over strong FL baselines. Supplementary vision

benchmarks and additional analyses on privacy, efficiency, TAPR stability, and client scalability are provided in the supplementary material.

2 Related Work

2.1 Psychiatric Heterogeneity and Mode-Aware Modeling

Clinical heterogeneity in psychiatry typically presents as polymorphic manifestations within a single diagnosis, indicating the existence of distinct latent modes [2]. The DSM-5 [3] formally categorizes these intra-class distinctions via *specifiers*, which help characterize heterogeneous clinical presentations and support treatment-oriented assessment [14]. For instance, MDD spans drastically different behavioral profiles [14], further complicated by comorbidities that form complex symptom topologies [53]. Similarly, machine learning has revealed stable latent modes in schizophrenia with distinct neural correlates [7]. Leveraging these insights, multimodal learning increasingly captures these modes through objective biomarkers [4], utilizing acoustic [10] and facial cues [11] to map distinct symptom clusters.

This inherent polymorphism poses formidable challenges for FL. Distributed training across distinct data views induces severe non-IID shifts [25, 34]. Critically, these mode-driven multi-peak structures cause traditional coarse global averaging to yield clinically ambiguous and mathematically distorted prototypes. Such vulnerabilities strongly motivate our topology-aware learning mechanism, designed to faithfully preserve fine-grained latent modes during privacy-preserving multi-center collaboration.

2.2 Heterogeneous and Clustered FL

The proliferation of Internet of Things devices and edge computing paradigms has catalyzed Heterogeneous FL to tackle critical statistical (Non-IID) and systemic challenges [28,32,61]. Traditional optimization algorithms like FedAvg [40] suffer from severe weight divergence and unacceptably slow convergence when exposed to extreme data heterogeneity. To mitigate client drift, regularization-based methods such as FedProx [36], SCAFFOLD [29], FedDyn [1], and FedDC [16] have been proposed to constrain local updates by applying proximal penalty terms or explicit local drift decoupling mechanisms [56].

Another prominent line of research to handle data heterogeneity involves Clustered FL [5, 12, 19] and Federated Domain Generalization [33, 38, 44, 59], which group clients or align distributions to train robust models. Concurrently, methods utilizing local normalization [31,37], ensemble distillation [21], and generalization adjustments [58, 64] aim to mitigate domain-specific feature skew. While these approaches successfully improve local adaptation on edge devices, they fundamentally model client data distributions as skewed *single-peak Gaussian models* or assume a shared domain context. Consequently, they fail to explicitly address the complex intra-class structural divergence (i.e., multi-peak distributions) prevalent in multi-center clinical datasets, leaving the intrinsic latent mode variations largely underexplored and vulnerable to severe feature collapse. These methods primarily address label skew, domain shift, or client clustering, whereas FedMental targets within-class multi-peak structure induced by psychiatric polymorphism; we discuss them as complementary heterogeneous-FL settings rather than direct substitutes for topology-aware prototype refinement.

2.3 Prototype-based FL

Prototype learning [15,39,50,62], which represents entire categories via semantic centroids, has recently emerged as a communication-efficient and architecture-agnostic solution for HFL. FedProto [51] pioneered this paradigm by exchanging lightweight class prototypes between the server and clients instead of full neural network weights. This strategy seamlessly bypasses the issue of model architecture heterogeneity. Building upon this foundational concept, subsequent frameworks such as FedProc [41], FedPLVM [57], FedTGP [63], and semantic anchor mechanisms [45,65,66] have significantly improved prototype robustness through discrepancy-aware aggregation strategies.

Despite these substantial advancements, capturing the true multi-peak topology of highly non-IID data remains a critical challenge. Most methods rely on coarse-grained global averaging, which forcefully collapses diverse, polymorphic features into distorted chimera prototypes. Recognizing this limitation, recent SOTA work like FPL [27] introduced unsupervised clustering [6,22] (e.g., FINCH [49]) to construct cluster prototypes that preserve domain knowledge. However, FPL exhibits a fundamental theoretical vulnerability: it clusters *client-level mean prototypes*. In clinical reality, a single medical center often houses highly polymorphic patients (e.g., agitated and melancholic depression variants). Averaging these distinct patient features locally instantly obliterates the intra-client topological diversity before it ever reaches the server. Furthermore, FPL ultimately relies on an averaged unbiased prototype for consistency regularization, effectively reverting to a rigid unimodal Gaussian assumption.

In stark contrast, FedMental fundamentally redefines fine-grained prototype construction. Rather than prematurely averaging features at the client level, FedMental uploads privacy-preserving, *subsampled sample-level features* grouped by class–view pairs. Our Topology-Aware Prototype Refinement (TAPR) then performs parameter-free FINCH clustering directly on this rich global feature manifold. This architectural distinction ensures that the true multi-peak latent modes—whether inter-client or intra-client—are faithfully discovered and preserved without semantic dilution, making it particularly suited for the profound heterogeneity inherent in cross-view psychiatric data.

2.4 Contrastive Alignment in FL

Originating from self-supervised visual representation learning [8, 17, 23], contrastive learning has been widely adopted in FL to refine feature representations, with a core objective of maximizing agreement between positive pairs

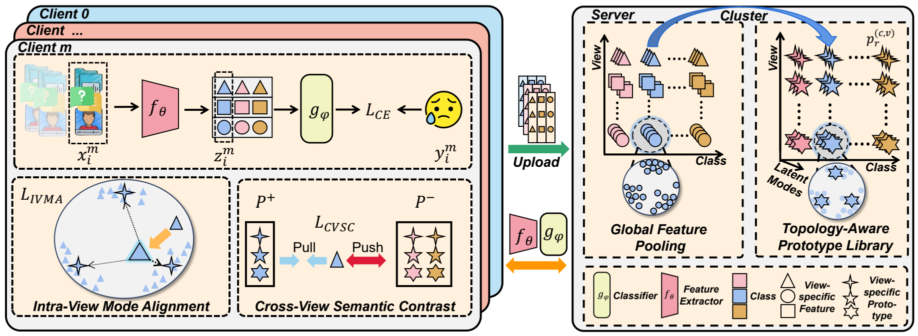


Fig. 2: The overall framework of FedMental. Left (Client): Each client extracts view-specific features and optimizes them via three complementary objectives: task loss ($\mathcal{L}_{CE}$), Intra-View Mode Alignment (IVMA) which softly anchors features to matched prototypes, and Cross-View Semantic Contrast (CVSC) which aligns cross-view positive pairs while repelling negative ones. **Right (Server):** The server performs Topology-Aware Prototype Refinement (TAPR), leveraging parameter-free clustering to discover and preserve multiple latent semantic modes (represented as stars) in the global prototype library.

while minimizing it for negative pairs. Advanced works have refined this framework via debiased sampling [9] and mitigating diminished semantics [18]. In FL, MOON [35] and pre-trained contrastive frameworks [52] introduce model-level contrastive objectives to pull local representations toward the global consensus, while pushing them away from historical local outputs.

However, standard contrastive FL methods are inherently view-agnostic: they force indiscriminate alignment across domains/modalities solely based on class labels, ignoring the data's underlying geometric structure and severely disrupting the multi-peak topology of local feature spaces. To address this, FedMental proposes a dual-objective alignment strategy for topology-aware learning. First, Intra-View Mode Alignment (IVMA) uses dynamic soft assignment to anchor local features to their matched mode prototypes, ensuring view-specific stability without manifold distortion. Second, Cross-View Semantic Contrast (CVSC) leverages an InfoNCE-based mechanism [43] to extract cross-view invariant semantics. This design elegantly balances view-specific nuances and cross-view generalization, establishing a new paradigm for topology-aware FL.

3 Methodology

As illustrated in Fig. 2 and summarized in Algorithm 1, **FedMental** establishes a robust closed-loop paradigm between topology-aware prototype refinement on the server and fine-grained representation alignment on the clients. For clarity, a comprehensive summary of the key notations is provided in the supplementary material.

Algorithm 1 FedMental: Topology-Aware Federated Prototypes

Input: Comm. epochs E, local rounds T, number of clients M.
Local dataset $\mathcal{D}_m$ and local model θ_m for client m.
Output: The final global model θ^E and prototype library $\mathbf{P}$.
Server Execution:
for $e = 1, 2, \ldots, E$ **do**
  for $m = 1, 2, \ldots, M$ **in parallel do**
    $\theta_m^e, \mathbf{Z}_m \leftarrow \text{ClientUpdate}(\theta^{e-1}, \mathbf{P})$
  end for
  Conduct Topology-Aware Prototype Refinement (TAPR) via Eq. (3)-Eq. (5).
  / Global model update */*
  $\theta^e \leftarrow \sum_{m=1}^{M} \frac{|\mathcal{D}_m|}{|\mathcal{D}|} \theta_m^e$
end for
Client Execution:
ClientUpdate$(\theta^{e-1}, \mathbf{P})$:
$\theta_m^e \leftarrow \theta^{e-1}$
for $t = 1, 2, \ldots, T$ **do**
  for batch $(x_i, y_i) \in \mathcal{D}_m$ **do**
    Update θ_m^e using $\mathbf{P}$ by minimizing Eq. (10).
  end for
end for
$\mathbf{Z}_m^{(c,v)} \leftarrow \text{Subsample}(\{z_i^{(v)} \mid (x_i^m, y_i^m) \in \mathcal{D}_m,\ y_i^m = c,\ v \in \mathcal{V}_m\}, \gamma)$
return $\theta_m^e, \mathbf{Z}_m$

3.1 Problem Definition & Preliminaries

We consider a FL setup comprising a central server and M clients. Unlike traditional FL which assumes data are Independently and Identically Distributed (IID), we address the highly challenging *Cross-View Heterogeneity* scenario. Here, a *view* ($v \in \mathcal{V}$) is broadly defined to encompass varying data modalities (e.g., audio, video, text) or distinct domain centers (e.g., hospitals). Each client m holds a real-world local dataset $\mathcal{D}_m = \{(x_i^m, y_i^m)\}_{i=1}^{N_m}$, where $y_i^m \in \mathcal{Y} = \{1, \ldots, C\}$. The data of each client is naturally constrained to a specific view v_m, or a subset of views in multi-view settings.

Clients share a unified model composed of feature extractors f_θ and a classifier g_φ. For a sample x_i^m, its extracted representation in view v is denoted as $z_i^{(v)} \in \mathbb{R}^d$. Traditional FL optimizes the local model by minimizing the standard Cross-Entropy loss:

$$\mathcal{L}_{CE} = -\frac{1}{|\mathcal{D}_m|} \sum_{(x_i^m, y_i^m) \in \mathcal{D}_m} \log p(y_i^m \mid x_i^m), \tag{1}$$

where $p(y_i^m \mid x_i^m)$ is the predicted probability of the true class by g_φ. However, the massive feature distribution divergence caused by the View Gap renders direct parameter aggregation ineffective, causing severe semantic conflict and representation collapse.

3.2 Topology-Aware Global Prototype Construction

In standard prototype-based FL, global prototypes are constructed by taking the arithmetic mean of all local features of the same class, implicitly assuming a unimodal distribution. However, when distributions vary significantly across views, data from the exact same category exhibit drastically different multi-peak patterns. Averaging these diverse representations into a single global centroid induces *semantic dilution*, rendering the prototype a distorted chimera that fails to represent any specific latent mode accurately. To circumvent this, we propose Topology-Aware Global Prototypes, retaining a distinct set of fine-grained *mode prototypes* (represented as stars in Fig. 2) for each *class–view pair* to preserve the multi-peak structural topology. It mainly contains two parts.

Local Feature Collection. At the final local training epoch, client m extracts per-view embeddings from $\mathbf{z}_i^m = \{z_i^{(v)}\}_{v\in\mathcal{V}_m}$ using the shared view-specific encoder. To reduce communication overhead, the client randomly subsamples a fraction of these high-level embeddings with a retention ratio γ:

$$\mathbf{Z}_m^{(c,v)} = \{z_i^{(v)} \mid (x_i^m, y_i^m) \in \mathcal{D}_m,\ y_i^m = c\}_{\text{subsample}(\gamma)}, \quad v \in \mathcal{V}_m \tag{2}$$

Although raw data remain local, uploaded embeddings may still pose feature-level leakage risks; we therefore treat subsampling as a communication-control mechanism rather than a formal privacy guarantee, and report leakage diagnostics in the supplementary material. The client then uploads the subsampled feature collection $\mathbf{Z}_m = \{\mathbf{Z}_m^{(c,v)}\}$ alongside the model parameters to the server.

Topology-Aware Prototype Refinement (TAPR). Upon receiving feature sets from all active clients, the server aggregates them by (class, view) pairs into a global pool $\mathcal{Z}_{c,v} = \bigcup_{m=1}^{M} \mathbf{Z}_m^{(c,v)}$.

A fundamental challenge in privacy-preserving FL is that the server is blind to the raw data, making it impossible to pre-determine the true number of latent clusters (e.g., hyperparameter K in K-Means). Therefore, instead of a crude average, FedMental employs FINCH (First Integer Neighbor Clustering Hierarchy), a parameter-free unsupervised clustering algorithm, to automatically discover the underlying multi-peak structures. FINCH partitions the pool $\mathcal{Z}_{c,v}$ into an adaptive number of $R_{c,v}$ distinct clusters based on cosine similarities:

$$\{\mathcal{G}_1^{(c,v)}, \mathcal{G}_2^{(c,v)}, \dots, \mathcal{G}_{R_{c,v}}^{(c,v)}\} = \text{FINCH}(\mathcal{Z}_{c,v}). \tag{3}$$

The server then computes the centroid of each cluster to generate a set of Latent Mode Prototypes:

$$p_r^{(c,v)} = \frac{1}{|\mathcal{G}_r^{(c,v)}|} \sum_{z\in\mathcal{G}_r^{(c,v)}} z, \quad \forall r \in \{1, \dots, R_{c,v}\} \tag{4}$$

$$\mathbf{P}^{(c,v)} = \{p_r^{(c,v)}\}_{r=1}^{R_{c,v}} \tag{5}$$

This tailored TAPR mechanism dynamically identifies distinct sub-distributions, ensuring the global prototype library $\mathbf{P}$ precisely preserves the intrinsic multi-peak topological structures (the rightmost panel in Fig. 2) without prior structural assumptions. Moreover, intra-cluster averaging naturally filters out client outliers and noise.

Why TAPR avoids semantic collapse. When intra-class features follow a mixture-like distribution, a single class mean can lie between latent modes and fail to represent any clinically meaningful subtype. TAPR therefore delays centroid formation until after parameter-free topology discovery on the global feature pool, so averaging is performed within discovered modes rather than across them. Round-wise stabilization of cluster counts and prototype drift is analyzed in the supplementary material.

3.3 Dual Fine-grained View Alignment Strategy

With a global prototype library rich in latent mode semantics, forcing view-agnostic alignment to all prototypes indiscriminately would destroy local manifolds. FedMental instead introduces a tailored dual-alignment strategy to simultaneously consolidate view-specific local stability and bridge cross-view gaps.

Intra-View Mode Alignment (IVMA). To ensure feature stability, local features should primarily align with the global standard of their own view. Since $\mathbf{P}^{(c,v)}$ contains multiple mode prototypes, forcing alignment to a random one is highly suboptimal. Thus, we design a Soft-assignment mechanism.

For a local feature $z_i^{(v)}$ with label c, we compute its similarity distribution over all mode prototypes $p_r^{(c,v)}$ within the *same class and view*:

$$\alpha_{i,r} = \frac{\exp\left(\cos(z_i^{(v)}, p_r^{(c,v)})/\tau_{agg}\right)}{\sum_{j=1}^{R_{c,v}} \exp\left(\cos(z_i^{(v)}, p_j^{(c,v)})/\tau_{agg}\right)} \tag{6}$$

where τ_{agg} controls the assignment sharpness. As visualized in the IVMA module of Fig. 2 (denoted by dashed arrows $\alpha_1, \alpha_2 \dots$), we construct a dynamically weighted Mixture Prototype $\tilde{p}_i^{(c,v)}$ that acts as the optimal anchor for the current sample:

$$\tilde{p}_i^{(c,v)} = \sum_{r=1}^{R_{c,v}} \alpha_{i,r} p_r^{(c,v)} \tag{7}$$

Finally, we minimize the cosine alignment loss to anchor the feature within its corresponding latent mode manifold:

$$\mathcal{L}_{IVMA} = 1 - \cos\left(z_i^{(v)}, \tilde{p}_i^{(c,v)}\right) \tag{8}$$

IVMA allows local features to stand firm within their inherent mode space, heavily stabilizing local training against non-IID data skew.

Cross-View Semantic Contrast (CVSC). Focusing solely on intra-view alignment confines the model, failing to address cross-view generalization. To encourage the extraction of view-invariant representations, we leverage an InfoNCE-based contrastive mechanism.

For a feature $z_i^{(v)}$ (label c), we construct specific positive and negative prototype sets (as depicted in the CVSC module of Fig. 2):

Positive Set $\mathcal{P}^+$. All mode prototypes belonging to the *same class* but from *different views* ($\bigcup_{v' \neq v} \mathbf{P}^{(c,v')}$).

Negative Set $\mathcal{P}^-$. All prototypes belonging to *different classes* across *all views* ($\bigcup_{c' \neq c} \bigcup_{v'} \mathbf{P}^{(c',v')}$).

The Cross-View Semantic Contrast loss is defined as:

$$\mathcal{L}_{CVSC} = -\log \frac{\sum_{p \in \mathcal{P}^+} \exp(\cos(z_i^{(v)}, p)/\tau_{cross})}{\sum_{p \in \mathcal{P}^+ \cup \mathcal{P}^-} \exp(\cos(z_i^{(v)}, p)/\tau_{cross})}, \tag{9}$$

where τ_{cross} is the contrastive temperature. CVSC compels the local model to look beyond style and view boundaries to distill the universal essence of the class, while repelling negative clusters to widen the inter-class decision margin.

3.4 Overall Optimization Objective

In summary, the total objective for local clients combines task supervision, intra-view stabilization, and cross-view contrast:

$$\mathcal{L} = \mathcal{L}_{CE} + \lambda_1 \mathcal{L}_{IVMA} + \lambda_2 \mathcal{L}_{CVSC}, \tag{10}$$

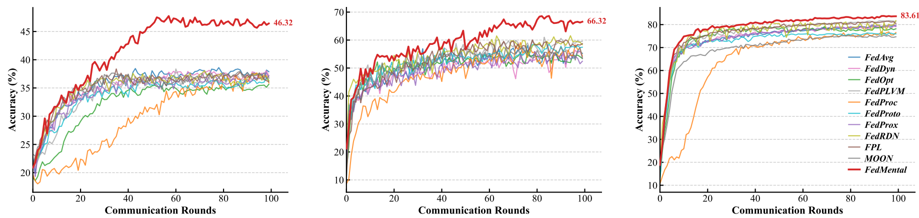
where λ_1 and λ_2 are balancing hyperparameters. This elegant dual design establishes a paradigm that simultaneously respects View Specificity (via IVMA) and enforces Semantic Generalizability (via CVSC), achieving superior robustness under extreme data heterogeneity.

4 Experiments

4.1 Experimental Setup

Datasets We evaluate FedMental on three benchmarks: *Psycare* [46], a real-world multi-center multi-view psychiatric dataset; *Digits* [27], a standard multi-domain digit recognition benchmark with 4 domains (MNIST, USPS, SVHN, SYN) and 10 classes; *Office-Caltech* [20], a classic multi-domain object recognition benchmark with 4 domains (Caltech, Amazon, Webcam, DSLR). Psycare comprises QA segments from psychiatric patients, exhibits extreme cross-view heterogeneity and polymorphic manifestations, and involves aligned tri-view sequences.

Baselines We compare FedMental against a comprehensive suite of SOTA FL methods: standard/regularized FL (FedAvg [40], FedProx [36], FedOpt [47], FedDyn [1]); contrastive FL (MOON [35]); and prototype-based FL (FedProto [51], FPL [27], FedProc [41], FedPLVM [57], FedRDN [60]).



(a) Psycare **(b)** Office-Caltech **(c)** Digits

Fig. 3: Testing accuracy curves across communication rounds. FedMental (the topmost curve) consistently demonstrates superior convergence speed and striking stability compared to other SOTA baselines across all benchmarks.

Implementation Details. We implement FedMental in PyTorch. For vision benchmarks, we use ResNet-10 [24] and follow partition protocols in [26, 35]: 20 clients for Digits, distributed as 3 for MNIST, 7 for USPS, 6 for SVHN, and 4 for SYN; and 10 clients for Office-Caltech, allocated as 3 for Caltech, 2 for Amazon, 1 for Webcam, and 4 for DSLR. For Psycare, we design a multi-branch feature extractor to align audio, visual, and text views, and simulate 7 clients with non-overlapping view combinations, namely 3 single-view, 3 dual-view, and 1 tri-view. Across all experiments, we set $E = 100$ communication epochs, a batch size of 64, and an SGD optimizer with a momentum of 0.9 and a weight decay of 10^{-5}, along with a learning rate of 0.01 for vision tasks and 0.001 for Psycare. We report average accuracy over the last 5 communication rounds across 3 independent runs, and we specifically utilize 5-fold cross-validation for Psycare. Additional analyses on communication/runtime cost, privacy leakage, TAPR stability, client scalability, and quantitative topology metrics are deferred to the supplementary material.

4.2 Performance Comparison with SOTA Methods

As detailed in Tab. 1 and Tab. 2, FedMental consistently improves over strong baselines across highly heterogeneous scenarios.

Primary Psychiatric Benchmark (Psycare). Psycare poses a formidable challenge due to the profound intra-class heterogeneity inherent in psychiatric diagnoses [2]. Methods relying on rigid unimodal assumptions suffer from severe semantic drift here and yield sub-optimal performance. FedMental achieves the best overall accuracy of 46.32% by mining and aligning topology-aware latent modes, yielding a +8.83% absolute improvement over FedAvg.

Supplementary Vision Benchmarks. FedMental also improves on the vision benchmarks (Fig. 3). On Digits, FedMental reaches 83.61%, +4.41% over FedAvg. On Office-Caltech, FedMental attains 66.32%, delivering a substantial +11.42% gain over FedAvg and outperforming the prototype baseline FedPLVM by nearly 7%.

Table 1: Comparison with SOTA methods on Digits and Office-Caltech. AVG denotes average accuracy across all domains. Best results are **bold**, second best underlined.

Methods	Digits						Office Caltech					
	MNIST	USPS	SVHN	SYN	AVG	Δ	Caltech	Amazon	Webcam	DSLR	AVG	Δ
FedAvg [40]	96.08	89.80	86.41	44.52	79.20	–	57.59	71.58	53.10	37.33	54.90	–
FedProx [36]	95.64	89.83	86.60	46.67	79.68	+0.48	60.89	73.16	47.93	30.67	53.16	−1.74
MOON [35]	92.85	90.45	86.76	29.12	74.79	−4.41	62.86	77.68	44.49	35.33	55.09	+0.19
FedDyn [1]	95.04	89.58	86.85	44.55	79.00	−0.20	62.59	74.74	54.48	28.67	55.12	+0.22
FedOPT [47]	95.54	92.72	80.57	42.89	77.93	−1.27	59.20	54.00	61.03	42.67	54.22	−0.68
FedProc [41]	94.48	86.37	87.60	36.13	76.14	−3.06	61.52	76.95	50.00	36.00	56.12	+1.22
FedProto [51]	96.35	89.86	77.69	40.02	75.98	−3.22	61.43	78.31	54.48	36.00	57.56	+2.66
FPL [27]	96.19	93.15	90.43	44.77	81.14	+1.94	61.86	82.05	59.69	29.67	58.32	+3.42
FedPLVM [57]	95.96	91.77	89.45	48.90	81.52	+2.32	67.74	81.14	54.77	33.67	59.33	+4.43
FedRDN [60]	86.38	93.36	91.18	48.55	79.87	+0.67	64.88	77.26	55.59	39.07	59.20	+4.30
FedMental	**97.00**	**94.72**	**92.32**	**50.41**	**83.61**	+4.41	**68.35**	**82.84**	**62.41**	**51.67**	**66.32**	**+11.42**

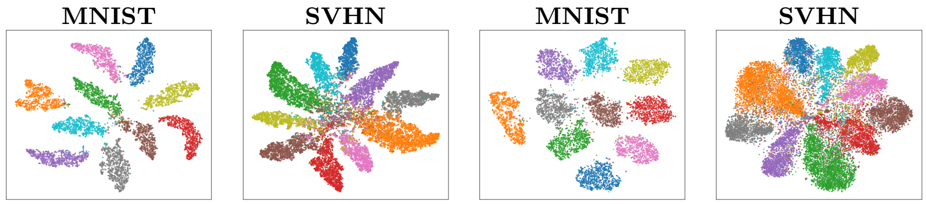
Table 2: Comparison with SOTA methods on Psycare. AVG denotes the per-class average accuracy. Best in **bold** and second with underline.

Methods	Classes					Overall	
	HC	BD1	BD2	MDD	SZ	AVG	Δ
FedAvg	79.57	21.63	29.82	39.81	16.61	37.49	–
FedProx	77.78	24.28	28.97	37.73	15.81	36.91	−0.58
MOON	80.91	17.65	30.45	43.35	13.87	37.25	−0.24
FedDyn	77.25	23.49	31.00	39.59	16.19	37.50	+0.01
FedOPT	77.84	21.18	28.61	34.80	14.75	35.44	−2.05
FedProc	75.87	23.59	31.12	39.03	14.31	36.78	−0.70
FedProto	75.36	20.14	28.10	40.18	17.33	36.22	−1.27
FPL	76.20	**33.25**	22.84	36.61	14.86	36.75	−0.74
FedPLVM	75.89	30.12	26.53	32.81	17.76	36.62	−0.87
FedRDN	77.47	24.08	26.75	40.49	15.45	36.85	−0.64
FedMental	**85.48**	28.19	**39.24**	**54.65**	**24.05**	**46.32**	**+8.83**

4.3 Prototype Representation Analysis

To visually validate FedMental's capacity to enhance semantic discriminability while disentangling polymorphic manifolds, we project the learned feature spaces and prototypes of FedMental alongside FedRDN using t-SNE (Fig. 4). We select MNIST and SVHN to represent extreme domain shift.

For FedRDN, we observe that its feature clusters are relatively loose on MNIST with noticeable inter-class overlap. This semantic collapse is severely exacerbated on SVHN, where features scatter erratically, resulting in heavily blurred decision boundaries. In contrast, FedMental consistently generates uniform, tight, and well-separated latent modes across both disparate domains. Even under the extreme background noise of SVHN, each class forms a compact cluster with distinct inter-class margins, devoid of cross-class entanglement.



(a) Ours (b) Ours (c) FedRDN (d) FedRDN

Fig. 4: t-SNE visualization of the learned feature topologies. FedMental effectively preserves intra-class compactness and clear inter-class margins despite severe domain shifts, outperforming the baseline FedRDN with blurred decision boundaries.

Table 3: Ablation study on the Digits and Office-Caltech datasets. Δ denotes the absolute accuracy improvement over the Base configuration.

Components		**Digits**						**Office-Caltech**					
IVMA	CVSC	MNIST	USPS	SVHN	SYN	**AVG**	Δ	Caltech	Amazon	Webcam	DSLR	**AVG**	Δ
×	×	96.08	89.80	86.41	44.52	79.20	–	57.59	71.58	53.10	37.33	54.90	–
✓	×	96.64	93.71	88.42	42.22	80.25	+1.05	63.93	78.63	53.45	42.00	59.50	+4.60
×	✓	96.18	91.35	89.16	45.96	80.66	+1.46	64.30	74.61	53.10	43.31	58.83	+3.93
✓	✓	**97.00**	**94.72**	**92.32**	**50.41**	**83.61**	**+4.41**	**68.35**	**82.84**	**62.41**	**51.67**	**66.32**	**+11.42**

These visualizations suggest that replacing coarse averaging with topology-aware latent mode preservation helps prevent semantic dilution. Because t-SNE is only a qualitative diagnostic, we additionally report intra-class compactness, inter-class separation, and cross-view alignment metrics in the supplementary material.

4.4 Ablation Study on Key Components

To isolate and quantify the individual contributions of our dual-alignment strategy, we systematically ablate the Intra-View Mode Alignment (IVMA) and Cross-View Semantic Contrast (CVSC) modules.

As reported in Tab. 3, the minimalist baseline yields sub-optimal performance. Applying only IVMA consistently boosts performance, validating that soft-assigning local features to their precise topological modes successfully anchors view-specific semantics and stabilizes local updates. Similarly, introducing only CVSC explicitly forces cross-view generalization via semantic contrast, yielding notable gains independently.

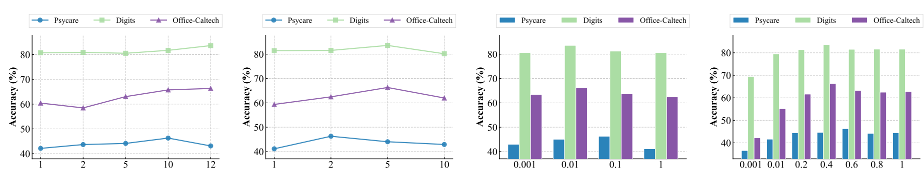
Crucially, the synergistic combination of both components achieves the highest accuracy across the board. On Office-Caltech, this dual-objective pushes accuracy to 66.32%, achieving a significant **+11.42%** improvement over the baseline. We further ablate these components on the Psycare clinical benchmark (Tab. 4), where the full framework secures the best performance, conclusively validating its necessity under real-world, mode-driven heterogeneity.

Table 4: Ablation study on the Psycare benchmark. Combining both topological alignment objectives optimally resolves psychiatric intra-class heterogeneity.

Components		Classes					Overall	
IVMA	CVSC	HC	BD1	BD2	MDD	SZ	AVG	Δ
×	×	79.57	21.63	29.82	39.81	16.61	37.49	–
✓	×	81.71	23.77	31.96	41.95	18.75	39.63	+2.14
×	✓	82.83	27.11	31.71	37.30	21.88	40.17	+2.68
✓	✓	**85.48**	**28.19**	**39.24**	**54.65**	**24.05**	**46.32**	**+8.83**

4.5 Hyperparameter Analysis

We analyze the sensitivity of FedMental to loss weights (λ_1, λ_2) and temperature factors $(\tau_{agg}, \tau_{cross})$ across all benchmarks (Fig. 5). Increasing λ_1 consistently improves performance and peaks around $[10.0, 12.0]$, confirming that strong topological anchoring stabilizes local training. λ_2 follows an inverted-U trend, with moderate values in $[2.0, 5.0]$ extracting view-invariant semantics without over-smoothing view-specific manifolds. For temperatures, $\tau_{agg} \in [0.01, 0.1]$ balances soft assignment against coarse averaging, while $\tau_{cross} \in [0.4, 0.6]$ avoids both representation collapse and gradient instability.



(a) λ_1 **(b)** λ_2 **(c)** τ_{agg} **(d)** τ_{cross}

Fig. 5: Sensitivity analysis of FedMental across different benchmarks. We evaluate the impact of loss weights (λ_1, λ_2) and temperature scaling factors $(\tau_{agg}, \tau_{cross})$ on the model performance. The results demonstrate the robustness of our method within a reasonable hyperparameter range while validating the necessity of our dual-objective design.

5 Conclusion

Conventional psychiatric diagnosis has suffered from subjectivity and inter-rater variability, while data-driven AI solutions are hindered by strict privacy regulations and fragmented clinical data. In this paper, we address this psychiatry-critical challenge in multi-view FL: *within-diagnosis* heterogeneity yields multi-peak topological mode structures, and coarse prototype averaging collapses them into semantically diluted chimera prototypes. To this end, we propose FedMental, a topology-aware federated prototype framework that refines class-view prototypes via TAPR and optimizes clients with a dual objective to balance view-specific stability and cross-view generalization. This framework provides a FL path that keeps raw data local for multi-center collaborative psychiatric screening. Experiments on a real-world multi-center multi-view psychiatric dataset demonstrate substantial gains over strong federated baselines, while supplementary vision benchmarks further support consistent gains from topology-aware mode-preserving aggregation under the evaluated heterogeneous settings.

Acknowledgements

This work was partially supported by the National Key Research and Development Program of China (2024YFC3308400, 2026YFE0202100). It was also supported by the National Natural Science Foundation of China (62361166629, 62576254, 62506269, 623B2080). Additional support was provided by the Key Research and Development Program of Wuhan (2025061202030423). The Major Project of Science and Technology Innovation of Hubei Province (2024BCA003) also supported this work. This work was also supported by the WHU-Kingsoft Joint Lab. The supercomputing system at the Supercomputing Center of Wuhan University supported the numerical calculations in this paper.

References

1. Acar, D.A.E., Zhao, Y., Navarro, R.M., Mattina, M., Whatmough, P.N., Saligrama, V.: Federated learning based on dynamic regularization. In: Int. Conf. Learn. Represent. (2021)
2. Allsopp, K., Read, J., Corcoran, R., Kinderman, P.: Heterogeneity in psychiatric diagnostic classification. Psychiatry research (2019)
3. Association, A.P.: Diagnostic and statistical manual of mental disorders. American Psychiatric Association Publishing (2022)
4. Baltrušaitis, T., Ahuja, C., Morency, L.P.: Multimodal machine learning: A survey and taxonomy. IEEE Trans. Pattern Anal. Mach. Intell. (2018)
5. Briggs, C., Fan, Z., Andras, P.: Federated learning with hierarchical clustering of local updates to improve training on non-iid data. In: 2020 international joint conference on neural networks (IJCNN) (2020)
6. Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., Joulin, A.: Unsupervised learning of visual features by contrasting cluster assignments. Adv. Neural Inform. Process. Syst. (2020)
7. Chen, J., Patil, K.R., Weis, S., Sim, K., Nickl-Jockschat, T., Zhou, J., Aleman, A., Sommer, I.E., Liemburg, E.J., Hoffstaedter, F., et al.: Neurobiological divergence of the positive and negative schizophrenia subtypes identified on a new factor structure of psychopathology using non-negative factorization: an international machine learning study. Biological psychiatry (2020)
8. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: Int. Conf. Mach. Learn. (2020)
9. Chuang, C.Y., Robinson, J., Lin, Y.C., Torralba, A., Jegelka, S.: Debiased contrastive learning. Adv. Neural Inform. Process. Syst. (2020)

10. Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., Quatieri, T.F.: A review of depression and suicide risk assessment using speech analysis. Speech communication (2015)
11. Dibeklioğlu, H., Hammal, Z., Cohn, J.F.: Dynamic multimodal measurement of depression severity using deep autoencoding. IEEE journal of biomedical and health informatics (2017)
12. Duan, M., Liu, D., Ji, X., Liu, R., Liang, L., Chen, X., Tan, Y.: Fedgroup: Efficient federated learning via decomposed similarity-based clustering. In: 2021 IEEE Intl Conf on parallel & distributed processing with applications, big data & cloud computing, sustainable computing & communications, social computing & networking (ISPA/BDCloud/SocialCom/SustainCom) (2021)
13. European Parliament, Council of the European Union: Regulation (eu) 2016/679 (general data protection regulation). Official Journal of the European Union, L 119 (2016), 27 April 2016, pp. 1–88
14. Fiorillo, A., Albert, U., Dell'Osso, B., Pompili, M., Sani, G., Sampogna, G.: The clinical utility and relevance in clinical practice of dsm-5 specifiers for major depressive disorder: A delphi expert consensus study. Comprehensive Psychiatry (2024)
15. Gao, J., Ma, X., Xu, C.: Learning transferable conceptual prototypes for interpretable unsupervised domain adaptation. IEEE Transactions on Image Processing (2024)
16. Gao, L., Fu, H., Li, L., Chen, Y., Xu, M., Xu, C.Z.: Feddc: Federated learning with non-iid data via local drift decoupling and correction. In: IEEE Conf. Comput. Vis. Pattern Recog. (2022)
17. Ge, C., Wang, J., Tong, Z., Chen, S., Song, Y., Luo, P.: Soft neighbors are positive supporters in contrastive visual representation learning. In: Int. Conf. Learn. Represent. (2023)
18. Ge, S., Mishra, S., Li, C.L., Wang, H., Jacobs, D.: Robust contrastive learning using negative samples with diminished semantics. Adv. Neural Inform. Process. Syst. (2021)
19. Ghosh, A., Chung, J., Yin, D., Ramchandran, K.: An efficient framework for clustered federated learning. Adv. Neural Inform. Process. Syst. (2020)
20. Gong, B., Shi, Y., Sha, F., Grauman, K.: Geodesic flow kernel for unsupervised domain adaptation. In: IEEE Conf. Comput. Vis. Pattern Recog. (2012)
21. Gong, X., Sharma, A., Karanam, S., Wu, Z., Chen, T., Doermann, D., Innanje, A.: Preserving privacy in federated learning with ensemble cross-domain knowledge distillation. In: AAAI (2022)
22. Guo, Y., Xu, M., Li, J., Ni, B., Zhu, X., Sun, Z., Xu, Y.: Hcsc: Hierarchical contrastive selective coding. In: IEEE Conf. Comput. Vis. Pattern Recog. (2022)
23. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: IEEE Conf. Comput. Vis. Pattern Recog. (2020)
24. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conf. Comput. Vis. Pattern Recog. (2016)
25. Hsieh, K., Phanishayee, A., Mutlu, O., Gibbons, P.: The non-iid data quagmire of decentralized machine learning. In: Int. Conf. Mach. Learn. (2020)
26. Huang, W., Ye, M., Du, B.: Learn from others and be yourself in heterogeneous federated learning. In: IEEE Conf. Comput. Vis. Pattern Recog. (2022)
27. Huang, W., Ye, M., Shi, Z., Li, H., Du, B.: Rethinking federated learning with domain shift: A prototype view. In: IEEE Conf. Comput. Vis. Pattern Recog. (2023)

28. Huang, W., Ye, M., Shi, Z., Wan, G., Li, H., Du, B., Yang, Q.: Federated learning for generalization, robustness, fairness: A survey and benchmark. IEEE Transactions on Pattern Analysis and Machine Intelligence (2024)
29. Karimireddy, S.P., Kale, S., Mohri, M., Reddi, S., Stich, S., Suresh, A.T.: Scaffold: Stochastic controlled averaging for federated learning. In: Int. Conf. Mach. Learn. (2020)
30. Kroenke, K., Spitzer, R.L., Williams, J.B.: The phq-9: validity of a brief depression severity measure. Journal of general internal medicine (2001)
31. Le, K., Ho, L., Do, C., Le-Phuoc, D., Wong, K.S.: Efficiently assemble normalization layers and regularization for federated domain generalization. In: IEEE Conf. Comput. Vis. Pattern Recog. (2024)
32. Lee, J., Solat, F., Kim, T.Y., Poor, H.V.: Federated learning-empowered mobile network management for 5g and beyond networks: From access to core. IEEE Communications Surveys & Tutorials (2024)
33. Li, D., Yang, Y., Song, Y.Z., Hospedales, T.M.: Deeper, broader and artier domain generalization. In: Int. Conf. Comput. Vis. (2017)
34. Li, Q., Diao, Y., Chen, Q., He, B.: Federated learning on non-iid data silos: An experimental study. In: 2022 IEEE 38th international conference on data engineering (ICDE) (2022)
35. Li, Q., He, B., Song, D.: Model-contrastive federated learning. In: IEEE Conf. Comput. Vis. Pattern Recog. (2021)
36. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. Proc. Mach. Learn. Syst. (2020)
37. Li, X., Jiang, M., Zhang, X., Kamp, M., Dou, Q.: Fedbn: Federated learning on non-iid features via local batch normalization. In: Int. Conf. Learn. Represent. (2021)
38. Liu, Q., Chen, C., Qin, J., Dou, Q., Heng, P.A.: Feddg: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space. In: IEEE Conf. Comput. Vis. Pattern Recog. (2021)
39. Ma, S., Zhu, F., Zhang, X.Y., Liu, C.L.: Protogcd: Unified and unbiased prototype learning for generalized category discovery. IEEE Transactions on Pattern Analysis and Machine Intelligence (2025)
40. McMahan, B., Moore, E., Ramage, D., Hampson, S., Arcas, B.A.y.: Communication-Efficient Learning of Deep Networks from Decentralized Data. In: International Conference on Artificial Intelligence and Statistics (2017)
41. Mu, X., Shen, Y., Cheng, K., Geng, X., Fu, J., Zhang, T., Zhang, Z.: Fedproc: Prototypical contrastive federated learning on non-iid data. Future Generation Computer Systems (2023)
42. Nguyen, D.C., Pham, Q.V., Pathirana, P.N., Ding, M., Seneviratne, A., Lin, Z., Dobre, O., Hwang, W.J.: Federated learning for smart healthcare: A survey. ACM Comput. Surv. (2022)
43. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)
44. Peng, X., Bai, Q., Xia, X., Huang, Z., Saenko, K., Wang, B.: Moment matching for multi-source domain adaptation. In: Int. Conf. Comput. Vis. (2019)
45. Qiao, Y., Munir, M.S., Adhikary, A., Le, H.Q., Raha, A.D., Zhang, C., Hong, C.S.: Mp-fedcl: Multiprototype federated contrastive learning for edge intelligence. IEEE Internet of Things journal (2023)
46. Rao, Y., Li, H., Li, X., Li, M., Zong, X., Ye, M.: Psycare: A multi-dimensional multimodal benchmark for major psychiatric disorder recognition. Harvard Dataverse, V3 (2025)

47. Reddi, S., Charles, Z., Zaheer, M., Garrett, Z., Rush, K., Konečný, J., Kumar, S., McMahan, H.B.: Adaptive federated optimization. In: Int. Conf. Learn. Represent. (2021)
48. Rush, A.J., Trivedi, M.H., Wisniewski, S.R., Nierenberg, A.A., Stewart, J.W., Warden, D., Niederehe, G., Thase, M.E., Lavori, P.W., Lebowitz, B.D., et al.: Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a star* d report. American Journal of Psychiatry (2006)
49. Sarfraz, S., Sharma, V., Stiefelhagen, R.: Efficient parameter-free clustering using first neighbor relations. In: IEEE Conf. Comput. Vis. Pattern Recog. (2019)
50. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. Adv. Neural Inform. Process. Syst. (2017)
51. Tan, Y., Long, G., Liu, L., Zhou, T., Lu, Q., Jiang, J., Zhang, C.: Fedproto: Federated prototype learning across heterogeneous clients. In: AAAI (2022)
52. Tan, Y., Long, G., Ma, J., Liu, L., Zhou, T., Jiang, J.: Federated learning from pre-trained models: A contrastive learning approach. Adv. Neural Inform. Process. Syst. (2022)
53. Tong, P., Bo, P., Shi, Y., Dong, L., Sun, T., Gao, X., Yang, Y.: Clinical traits of patients with major depressive disorder with comorbid borderline personality disorder based on propensity score matching. Depression and anxiety (2020)
54. Topol, E.J.: High-performance medicine: the convergence of human and artificial intelligence. Nature medicine (2019)
55. United States Congress: Health insurance portability and accountability act of 1996. Public Law 104-191 (1996), enacted August 21, 1996
56. Wang, H., Yurochkin, M., Sun, Y., Papailiopoulos, D., Khazaeni, Y.: Federated learning with matched averaging. In: Int. Conf. Learn. Represent. (2020)
57. Wang, L., Bian, J., Zhang, L., Chen, C., Xu, J.: Taming cross-domain representation variance in federated prototype learning with heterogeneous data domains. In: Adv. Neural Inform. Process. Syst. (2024)
58. Wang, Z., Wang, Z., Wang, Z., Fan, X., Wang, C.: Federated learning with domain shift eraser. In: IEEE Conf. Comput. Vis. Pattern Recog. (2025)
59. Wu, G., Gong, S.: Collaborative optimization and aggregation for decentralized domain generalization and adaptation. In: Int. Conf. Comput. Vis. (2021)
60. Yan, Y., Fu, H., Li, Y., Xie, J., Ma, J., Yang, G., Zhu, L.: A simple data augmentation for feature distribution skewed federated learning. In: IEEE Conf. Comput. Vis. Pattern Recog. (2025)
61. Ye, M., Fang, X., Du, B., Yuen, P.C., Tao, D.: Heterogeneous federated learning: State-of-the-art and research challenges. ACM Computing Surveys (2023)
62. Zhang, B., Li, X., Ye, Y., Huang, Z., Zhang, L.: Prototype completion with primitive knowledge for few-shot learning. In: IEEE Conf. Comput. Vis. Pattern Recog. (2021)
63. Zhang, J., Liu, Y., Hua, Y., Cao, J.: Fedtgp: Trainable global prototypes with adaptive-margin-enhanced contrastive learning for data and model heterogeneity in federated learning. In: AAAI (2024)
64. Zhang, R., Xu, Q., Yao, J., Zhang, Y., Tian, Q., Wang, Y.: Federated domain generalization with generalization adjustment. In: IEEE Conf. Comput. Vis. Pattern Recog. (2023)
65. Zhou, T., Zhang, J., Tsang, D.H.: Fedfa: Federated learning with feature anchors to align features and classifiers for heterogeneous data. IEEE Transactions on Mobile Computing (2023)

66. Zhou, Y., Qu, X., You, C., Zhou, J., Tang, J., Zheng, X., Cai, C., Wu, Y.: Fedsa: A unified representation learning via semantic anchors for prototype-based federated learning. In: AAAI (2025)