Domain Sensitive Federated Learning with Fisher-Informed Pruning

Supplementary Material

7. Algorithm

The overall workflow of the proposed FedFIP framework, including domain-sensitive pruning, prototype-guided regularization, and structure-aware aggregation, is detailed in Algorithm 1. In each communication iteration, clients estimate and upload per-domain importance. The server derives a global pruning mask and updates domain prototypes. Clients then reactivate important channels, perform local training with regularization, and upload submodels for structure-aware aggregation.

Algorithm 1 Pipeline of FedFIP

Input: Global sparsity ratio ρ, local importance threshold ϕ, prototype update coefficient μ, contrastive regularization weight λ, learning rate η, number of communication iterations T, local epochs E.
Initialize: Global model $\boldsymbol{w}^0$, global pruning mask $M^g \leftarrow \mathbf{1}$, domain prototypes $\{p_d\} \leftarrow \mathbf{0}$.
Output: Final global model $\boldsymbol{w}^T$.
for each iteration $t = 0, 1, \ldots, T-1$ **do**
  for each client i **in parallel do**
    Estimate per-domain Fisher importance $F_{i,j}^k$ for all domains k by Eq. (2)
    Compress per-domain importance and upload $\tilde{\boldsymbol{F}}_i$ to server
  end for
  Server computes global pruning mask $\boldsymbol{M}^{(g)}$ over $F_j^{(g)}$ by Eq. (3)
  Server collects $\boldsymbol{V}_i^{(k)}$ from clients and aggregates to construct the prototype $\tilde{\mathbf{p}}^{(k)}$ by Eq. (5)
  Server updates domain prototypes by Eq. (6)
  for each client i **in parallel do**
    Apply $\boldsymbol{M}^{(g)}$ and reactivate channels where $\widetilde{F}_{i,j} > \phi$
    for each epoch $e = 0, 1, \ldots, E-1$ **do**
      Compute loss by Eq. (8)
    end for
    Upload model $\boldsymbol{w}_i^t = \{\boldsymbol{w}_{i,j}^t\}, \forall j \in C_{\text{shared}}$ to server
  end for
  Server aggregates the received model parameters by Eq. (9)
end for

8. Additional Experiments

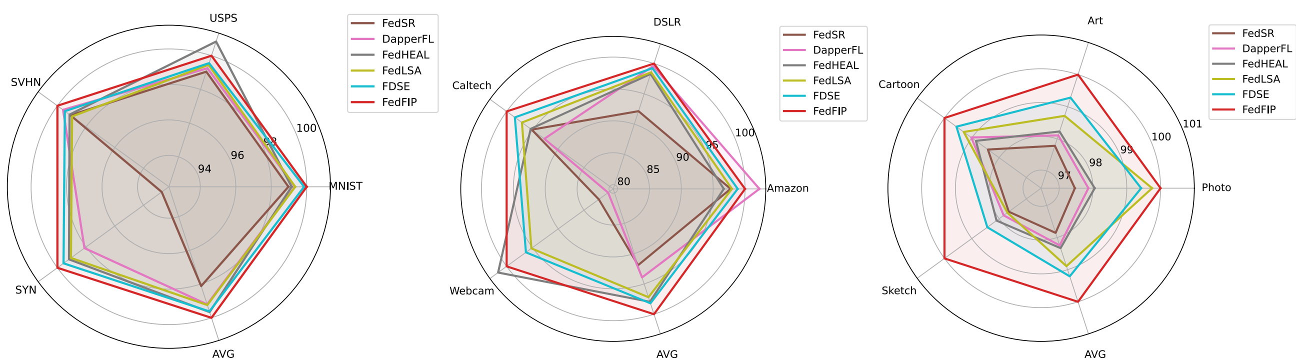
8.1. Introduction to Comparison Baselines

The comparison FL baselines used in this work are introduced as follows:

- FedAvg-AISTATS 2017 [28]: FedAvg is a classical and widely adopted FL framework. In each communication iteration, selected clients perform local training on their private data and upload model updates to the central server. The server then performs weighted averaging over the received local updates to produce the global model.
- FedProx-MLSys 2020 [23]: FedProx extends FedAvg by incorporating a proximal term into the local objective to mitigate client drift, penalizing deviations from the global model and thereby improving robustness to data heterogeneity.
- MOON-CVPR 2021 [21]: MOON tackles client drift in non-IID settings by incorporating contrastive learning into local training. In addition to minimizing the task-specific loss, each client uses contrastive objectives to align its current local model with the global model.
- FedBABU-ICLR 2022 [31]: FedBABU is a personalized FL framework that freezes the global feature extractor and only fine-tunes the classifier locally, enhancing client adaptability while maintaining a stable shared representation under data heterogeneity.
- FedFA-TMC 2023 [47]: FedFA addresses data heterogeneity by introducing feature anchors to unify the feature space used by all clients, ensuring that local feature extractors and classifiers are aligned during training.

- FEDSR-NeurIPS 2022 [30]: FedSR improves domain generalization by learning simple, invariant representations without sharing data or features across clients. It employs two regularizers: the L2-norm to implicitly align marginal distributions and the conditional mutual information term to suppress domain-specific, label-irrelevant cues, which mitigates spurious correlations and enhances generalization to unseen domains.
- DAPPERFL-NeurIPS 2024 [16]: DapperFL is a domain-adaptive FL framework that applies global structured pruning to reduce model complexity and communication overhead, while introducing a regularization strategy to encourage the encoder to capture domain-invariant representations, thereby improving generalization under domain-skewed data.
- FEDHEAL-CVPR 2024 [5]: FedHEAL addresses domain skew by combining local consistency regularization—enforcing prediction consistency between augmented data views—with a fairness-aware aggregation strategy that reweights client contributions based on local training loss to ensure equitable global model updates.
- FEDLSA-AAAI 2025 [8]: FedLSA learns a set of semantic anchors aided by the global semantic aware classifier, with margins enhanced through a separation loss. On the client side, representations are projected into a hyperspherical space and optimized with a von Mises–Fisher contrastive objective to ensure uniformity and intra-class compactness.
- FDSE-CVPR 2025 [39]: FDSE mitigates domain shift by decomposing each layer into a Domain-agnostic Feature Extractor (DFE) and a Domain-specific Skew Eraser (DSE), enabling iterative feature extraction and deskewing. A consistency regularizer aligns DSE statistics with global ones, while aggregation merges DFE modules via consensus-maximizing updates and personalizes DSE modules through similarity-aware attention.

8.2. Experimental Results



(a) Digits (b) Office-Caltech (c) PACS

Figure 4. Relative Accuracy of Baselines Against FEDFIP Across Different Datasets.

Table 8. Performance Comparison with Different Client Numbers.

Client Number	Digits		Office-Caltech		PACS	
	AVG	STD	AVG	STD	AVG	STD
$N = 20$	76.48	22.06	64.44	5.99	82.02	8.36
$N = 40$	76.89	21.84	64.73	5.66	82.31	8.19
$N = 60$	77.43	20.67	65.18	5.74	82.73	8.02
$N = 80$	78.02	21.15	65.97	5.01	83.07	7.89
$N = 100$	78.30	20.51	66.05	4.92	83.34	7.76

8.2.1. Relative Accuracy Comparison

Figure 4 depicts the relative performance of baselines that address domain generalization against FEDFIP, showing how their accuracies compare as a percentage of FEDFIP across different domains. Across Digits, Office-Caltech, and PACS, the polygons of competing methods are generally enclosed within FEDFIP's contour, indicating lower relative accuracy on most domains. Although a few baselines occasionally extend slightly outside FEDFIP's boundary on certain individual domains, suggesting a momentary advantage, these cases are isolated and do not persist across the domains. Importantly, on the AVG axis of all three datasets, the FEDFIP polygon consistently forms the outermost envelope, while baseline curves remain inside, demonstrating that even when brief domain-wise improvements occur, they do not translate into stronger overall performance. This consistent pattern highlights that FEDFIP delivers the most robust and stable results under the domain skew scenario.

8.2.2. Scalability Study

To assess the scalability of FEDFIP, we vary the number of participating clients $N = \{20, 40, 60, 80, 100\}$ and provide the results on Digits, Office-Caltech, and PACS datasets in Table 8. Note that as N increases, the average accuracy steadily improves and the standard deviation slightly decreases. This performance gain can be attributed to the design of FEDFIP: more clients introduce richer and more diverse domain distributions, which enhance the estimation of domain-wise Fisher importance and prototype representations. As a result, the global pruning mask becomes more accurate, and the structure-aware regularization more effectively aligns client models, leading to better generalization and improved model stability at scale.