

FSANL: Symmetric Active Negative Loss for Robust Federated Learning Under Heterogeneous Label Noise

Girum Fitihamlak Ejigu[†], Avi Deb Raha[†], Mrityunjay Gain[‡], Choong Seon Hong^{†*}

[†]Department of Computer Science and Engineering, Kyung Hee University, Yongin, Republic of Korea.

[‡]Department of Artificial Intelligence, Kyung Hee University, Yongin, Republic of Korea.

Email: girumfitex@khu.ac.kr, avi@khu.ac.kr, gain@khu.ac.kr, cshong@khu.ac.kr

Abstract—Federated learning (FL) enables collaborative model training across decentralized clients while preserving data privacy. However, real-world FL deployments often face two major challenges: data heterogeneity and label noise. The former arises due to non-IID client distributions, while the latter results from user-generated annotations, sensor failures, or distribution shifts, conditions that severely degrade the performance of standard FL algorithms. Existing approaches typically rely on multi-stage training, client selection strategies, or assume the presence of clean clients, which may not hold in realistic settings. In this work, we propose Federated Symmetric Active Negative Loss (FSANL), a simple yet effective single-stage framework designed to improve robustness under heterogeneous label noise without assuming any clean clients. FSANL integrates normalized and symmetric cross-entropy losses with an active negative suppression term, along with global L1 regularization to counteract client drift. Extensive experiments on CIFAR-10 demonstrate that FSANL outperforms existing baselines, including FedCorr and FedProx, under various non-IID and noisy label scenarios.

Index Terms—Federated learning, noisy label learning, non-IID, robust loss

I. INTRODUCTION

Federated Learning (FL) has emerged as a decentralized paradigm that enables multiple clients to collaboratively train a shared model while preserving local data privacy [1]. This decentralized framework is well-suited for privacy-sensitive domains such as healthcare, finance, and mobile applications [2], where raw data cannot be centralized. However, FL faces two major challenges in practical deployments: (i) statistical heterogeneity, where client data distributions are non-independent and identically distributed (non-IID) [3], and (ii) label noise, arising from user-generated data, annotation errors, or sensor corruption [4].

To address the issue of non-IID data, prior works have introduced regularization-based strategies [5], model alignment techniques such as MOON [6], and normalization adaptations like Population Normalization (PN) [7]. Although these

approaches effectively mitigate client drift, they generally assume label correctness, which does not hold in noisy real-world environments.

Recent efforts in noise-robust FL include FedCorr [8] and FedClean [9], which represent state-of-the-art solutions for label corruption. FedCorr utilizes dimensionality scores to identify clean clients and then corrects labels based on sample losses. However, it relies on the assumption that some clients are noise-free, which may not hold under widespread corruption. FedClean removes this dependency by introducing a two-stage framework: local GMM-based noise estimation followed by server-side refinement. While effective, such approaches increase training complexity and communication costs. Moreover, multi-stage training schemes like relaxed contrastive learning [10] further add auxiliary components, limiting scalability.

In this work, we propose Federated Symmetric Active Negative Loss (FSANL), a streamlined *single-stage* framework that is robust to both non-IID data and heterogeneous label noise. FSANL integrates Symmetric Cross Entropy [11], Active Negative Loss [12], and global model regularization into a unified loss formulation that requires no clean client identification, auxiliary networks, or multi-phase optimization.

We evaluate FSANL under heterogeneous label noise settings on CIFAR-10, where all clients are noisy but differ in noise intensity. Our method consistently outperforms recent baselines, including FedCorr, in highly noisy environments.

Our contributions are summarized as follows:

- We propose FSANL, a single-stage federated framework that is robust to non-IID distributions and client-level heterogeneous label noise.
- FSANL introduces a unified loss function combining symmetric and active negative loss components with L1-based global model regularization.
- Extensive experiments on CIFAR-10 demonstrate the effectiveness and scalability of FSANL compared to state-of-the-art baselines.

II. RELATED WORK

A. Federated Learning under Non-IID Settings

Federated Learning (FL) enables decentralized clients to collaboratively train machine learning models while preserving

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (No. RS-2023-00207816), (No. RS-2024-00352423), and supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2019-0-01287, Evolvable Deep Learning Model Generation Platform for Edge Computing) and supported by IITP- ITRC(Information Technology Research Center) grant funded by the Korea government(Ministry of Science and ICT)(IITP-2025-RS-2023-00258649) *Dr. CS Hong is the corresponding author.

data privacy. The foundational method, Federated Averaging (FedAvg) [1], performs multiple local updates on client devices and averages the model parameters at the server. However, its performance is severely affected under non-independent and identically distributed (non-IID) data due to client drift.

To mitigate this, FedProx [5] introduces a proximal term to stabilize local updates. MOON [6] further addresses heterogeneity by enforcing representation consistency through model contrastive learning, encouraging local models to remain aligned with the global trajectory. Another direction involves modifying the normalization process: FedBN [13] retains local batch normalization layers to address distributional shifts, while Population Normalization (PN) [7] replaces batch-derived statistics with trainable population-level statistics to ensure consistent inference regardless of batch size or participation ratio.

Other approaches include representation-based regularization [14], seeking flat minima [15], and personalization strategies [16], all aiming to improve generalization under heterogeneous client distributions. However, these methods typically assume label correctness and do not address robustness to noisy supervision, which is increasingly common in practical federated scenarios.

B. Noise-Robust Federated Learning

Label noise is an emerging challenge in FL, especially in settings with user-generated annotations, sensor errors, or domain mismatch. To enhance noise robustness, several strategies have been proposed. Mixup [17], initially developed for centralized learning, has been used directly [18] or adapted to FL [19] as a simple and effective regularization method.

More advanced federated methods specifically targeting label noise include FedCorr [8] and FedClean [9], considered state-of-the-art in this domain. FedCorr constructs a client-level dimensionality score and identifies clean clients and corrects noisy labels in three stages. However, its reliance on the presence of clean clients limits its applicability in uniformly corrupted environments. FedClean addresses this limitation by incorporating a two-stage framework: it estimates label noise using Gaussian Mixture Models (GMM) and then refines model predictions through server-side correction. Despite its effectiveness, this approach increases communication and computational complexity.

More recently, a relaxed contrastive learning approach [10] and representation regularization methods [20] have been proposed, which integrate a relaxed contrastive regularizer or a representation regularization with robust loss functions to address label noises. These methods typically involve additional projection heads and a sharpness-aware minimization technique.

In contrast, our proposed method, Federated Symmetric Active Negative Loss (FSANL), introduces a streamlined, *single-stage* noise-robust framework. By integrating Symmetric Cross-Entropy [11], Active Negative Loss [12], and global model regularization into a unified loss, FSANL enhances

robustness without requiring clean client identification or auxiliary networks. Furthermore, FSANL operates under standard FedAvg aggregation and accommodates heterogeneous label noise across clients, offering a practical and scalable solution for realistic federated settings.

III. METHOD

A. Problem Formulation

Federated learning (FL) enables decentralized clients to collaboratively train a global model without sharing raw data. Let $\mathcal{K} = \{1, 2, \dots, K\}$ denote the set of K clients. Each client $k \in \mathcal{K}$ holds a private dataset $\mathcal{D}_k = \{(\mathbf{x}_{k,i}, y_{k,i})\}_{i=1}^{n_k}$, where $\mathbf{x}_{k,i} \in \mathbb{R}^d$ represents an input sample and $y_{k,i} \in \{1, \dots, C\}$ is its class label. The union of all local datasets forms the global dataset $\mathcal{D} = \bigcup_{k=1}^K \mathcal{D}_k$ with a total of $N = \sum_{k=1}^K n_k$ samples.

The objective of FL is to learn a global model $\mathbf{w} \in \mathbb{R}^p$ that minimizes the empirical risk over the global data distribution:

$$\min_{\mathbf{w}} \mathbb{E}_{(\mathbf{x}, y) \sim \mathcal{D}} [\ell(f(\mathbf{x}; \mathbf{w}), y)], \quad (1)$$

where $f(\cdot; \mathbf{w})$ is a deep neural network parameterized by $\mathbf{w}$ and $\ell(\cdot, \cdot)$ is a supervised loss function.

Each client performs local updates on $\mathcal{D}_k$ using its local model $\mathbf{w}_k^{(t)}$ over E local epochs. After local training, the updated parameters are aggregated on the server via weighted averaging:

$$\mathbf{w}^{(t+1)} = \sum_{k \in \mathcal{S}^{(t)}} \frac{n_k}{\sum_{j \in \mathcal{S}^{(t)}} n_j} \mathbf{w}_k^{(t+1)}, \quad (2)$$

where $\mathcal{S}^{(t)} \subseteq \mathcal{K}$ denotes the subset of clients participating in round t .

However, two major challenges arise in realistic FL settings: (i) *statistical heterogeneity* due to non-IID client data, often simulated by Dirichlet distributions; and (ii) *label noise*, where $y_{k,i}$ may be corrupted. These issues degrade generalization and stability of the global model. Our method aims to address both challenges in a unified framework.

B. Federated Symmetric Active Negative Loss (FSANL)

We propose Federated Symmetric Active Negative Loss (FSANL), a principled loss formulation designed to enhance robustness to both label noise and client heterogeneity. FSANL draws from recent advances in symmetric loss learning [11] and active negative loss functions [12], and integrates them with a global regularization mechanism.

The FSANL loss on client k is defined as:

$$\mathcal{L}_{\text{FSANL}}(\mathbf{w}_k; \mathcal{D}_k, \mathbf{w}) = \mathcal{L}_{\text{SANL}}(\mathbf{w}_k; \mathcal{D}_k) + \lambda \|\mathbf{w}_k - \mathbf{w}\|_1, \quad (3)$$

where $\mathbf{w}$ is the global model broadcast to clients, and λ controls the strength of the L1 regularization. This term encourages proximity between local and global models, alleviating client drift in non-IID settings.

1) *Symmetric Active Negative Loss (SANL)*: The core of FSANL is the SANL loss, which integrates three loss components:

$$\mathcal{L}_{\text{SANL}} = \alpha \mathcal{L}_{\text{NCE}} + \gamma \mathcal{L}_{\text{NNCE}} + \delta \mathcal{L}_{\text{SCE}}, \quad (4)$$

where $\mathcal{L}_{\text{NCE}}$ is Normalized Cross-Entropy, $\mathcal{L}_{\text{NNCE}}$ is Normalized Negative Cross-Entropy (from ANL), and $\mathcal{L}_{\text{SCE}}$ is Symmetric Cross-Entropy [11]. The hyperparameters α , γ , and δ determine their relative contributions.

For a batch $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^B$ with predictions $\mathbf{p}_i = \text{softmax}(f(\mathbf{x}_i; \mathbf{w}_k))$ and one-hot label vectors $\mathbf{y}_i$, the loss terms are:

a) *Normalized Cross Entropy (NCE)*:

$$\mathcal{L}_{\text{NCE}} = -\frac{1}{B} \sum_{i=1}^B \frac{\mathbf{y}_i^\top \log \mathbf{p}_i}{-\sum_{c=1}^C \log p_{i,c}}, \quad (5)$$

This term emphasizes learning on confident predictions while normalizing for distribution sharpness.

b) *Normalized Negative Cross Entropy (NNCE)*:

$$\mathcal{L}_{\text{NNCE}} = \frac{1}{B} \sum_{i=1}^B \left(1 - \frac{\mathbf{y}_i^\top \log \mathbf{q}_i}{-\sum_{c=1}^C \log q_{i,c}} \right), \quad (6)$$

where $\mathbf{q}_i = \max(\mathbf{p}_i, \epsilon)$ ensures numerical stability. This component penalizes overconfidence in non-target classes, increasing robustness to noisy labels.

c) *Symmetric Cross Entropy (SCE)*:

$$\mathcal{L}_{\text{SCE}} = \frac{1}{B} \sum_{i=1}^B (\eta \cdot \text{CE}(\mathbf{y}_i, \mathbf{p}_i) + \zeta \cdot \text{RCE}(\mathbf{p}_i, \mathbf{y}_i)), \quad (7)$$

where CE and RCE denote the forward and reverse cross-entropy terms, with η and ζ controlling their balance. Inspired by [11], RCE mitigates underfitting on hard classes and overfitting on noisy samples.

2) *Global Model Regularization*: To address model inconsistency caused by statistical heterogeneity, we introduce a regularization term that penalizes the L1 norm difference between local and global models:

$$\mathcal{L}_{\text{L1}}(\mathbf{w}_k, \mathbf{w}) = \|\mathbf{w}_k - \mathbf{w}\|_1 = \sum_{j=1}^p |\mathbf{w}_{k,j} - \mathbf{w}_j|, \quad (8)$$

This enforces consistency and encourages sparse deviations, acting as a communication-efficient prior similar in spirit to techniques like FedProx [5].

3) *Local Update and Server Aggregation*: Each client minimizes the FSANL loss over E local steps using SGD:

$$\mathbf{w}_k^{(t+1)} \leftarrow \mathbf{w}_k^{(t)} - \eta \nabla_{\mathbf{w}_k} \mathcal{L}_{\text{FSANL}}(\mathbf{w}_k^{(t)}), \quad (9)$$

where η is the learning rate. After local updates, the server aggregates:

$$\mathbf{w}^{(t+1)} = \sum_{k \in \mathcal{S}^{(t)}} \frac{n_k}{\sum_{j \in \mathcal{S}^{(t)}} n_j} \mathbf{w}_k^{(t+1)}. \quad (10)$$

4) *Mixup Augmentation*: To further mitigate the adverse impact of label noise, FSANL incorporates Mixup augmentation [17], which linearly interpolates both inputs and labels from pairs of training samples. Specifically, given two samples $(\mathbf{x}_i, \mathbf{y}_i)$ and $(\mathbf{x}_j, \mathbf{y}_j)$, the augmented sample is constructed as:

$$\tilde{\mathbf{x}} = \Lambda \mathbf{x}_i + (1 - \Lambda) \mathbf{x}_j, \quad \tilde{\mathbf{y}} = \Lambda \mathbf{y}_i + (1 - \Lambda) \mathbf{y}_j, \quad (11)$$

where $\Lambda \sim \text{Beta}(\alpha_{\text{mix}}, \alpha_{\text{mix}})$ and $\alpha_{\text{mix}} > 0$ is a hyperparameter controlling the interpolation strength. This strategy promotes smoother decision boundaries and reduces the risk of overfitting to noisy labels.

C. Summary

The FSANL framework integrates multiple robustness components in a single-stage federated learning pipeline. It builds upon three synergistic components: (i) the Symmetric Active Negative Loss (SANL), which combines normalized cross-entropy, active negative suppression, and symmetric cross-entropy to enhance resistance to label noise; (ii) L1-based regularization with the global model to minimize model drift under non-IID conditions; and (iii) Mixup augmentation to further improve generalization. Clients perform local updates by minimizing the FSANL objective, and the global model is aggregated using the standard FedAvg strategy as shown in Algorithm 1. This unified approach provides a computationally efficient, robust alternative to multi-stage or correction-based federated techniques.

IV. EXPERIMENT

A. Implementation Details

We conduct experiments on the CIFAR-10 dataset [21] to evaluate the performance of FSANL under heterogeneous label noise and non-IID data settings. Following the setup of FedCorr [8], we simulate label corruption using a heterogeneous noise model parameterized by (ρ, τ) , where ρ denotes the percentage of clients with noisy labels, and τ specifies the minimum label noise ratio among corrupted clients. In this setting, each noisy client has a noise ratio sampled uniformly from $[\tau, 1.0]$, while the remaining $(1 - \rho)$ fraction of clients retain clean labels.

To introduce data heterogeneity, we adopt the widely used Dirichlet-based partitioning with parameter $\beta = 10.0$, which ensures moderate statistical skew among clients.

We use a ResNet-18 [22] model and simulate 100 clients with 10 randomly selected clients participating in each round. The models are trained using the SGD optimizer with momentum 0.5. We use a learning rate of 0.03. Each selected client performs five local training epochs per round, and the federated training proceeds for 700 communication rounds.

We compare FSANL against several representative baselines. FedAvg [1] serves as the standard averaging approach in FL. FedAvg + Mixup augments the FedAvg baseline with Mixup regularization [17] during local training. FedProx [5] incorporates a proximal term to stabilize local updates; we use a proximal coefficient of 5. FedCorr [8], considered a state-of-the-art method for label noise robustness, follows its original

Algorithm 1 Federated Learning with Federated Symmetric Active Negative Loss (FSANL)**Server executes:**

```

1: Initialize global model  $\mathbf{w}^{(0)}$ 
2: for each communication round  $t = 1, 2, \dots, T$  do
3:   for each client  $k \in \mathcal{K}$  in parallel do
4:      $\mathbf{w}_k^{(t+1)} \leftarrow \text{ClientUpdate}(\mathbf{w}^{(t)}, k)$ 
5:   end for
6:    $\mathbf{w}^{(t+1)} = \sum_{k \in \mathcal{K}} \frac{n_k}{N} \mathbf{w}_k^{(t+1)}$  (FedAvg)
7: end for

```

ClientUpdate($\mathbf{w}^{(t)}, k$):

```

1: Initialize local model:  $\mathbf{w}_k^{(t)} \leftarrow \mathbf{w}^{(t)}$ 
2: for each local epoch  $e = 1, 2, \dots, E$  do
3:   for each minibatch  $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^B$  from  $\mathcal{D}_k$  do
4:     Sample Mixup pair  $(\mathbf{x}_j, \mathbf{y}_j)$ 
5:     Sample  $\Lambda \sim \text{Beta}(\alpha_{\text{mix}}, \alpha_{\text{mix}})$ 
6:     Mixup augmented sample:
7:      $\tilde{\mathbf{x}} = \Lambda \mathbf{x}_i + (1 - \Lambda) \mathbf{x}_j$ 
8:      $\tilde{\mathbf{y}} = \Lambda \mathbf{y}_i + (1 - \Lambda) \mathbf{y}_j$ 
9:     Compute prediction:  $\mathbf{p}_i = \text{softmax}(f(\tilde{\mathbf{x}}; \mathbf{w}_k^{(t)}))$ 
10:    Compute SANL loss  $\mathcal{L}_{\text{SANL}}$  using Eqns. (5)–(8)
11:    Compute regularization:  $\mathcal{L}_{\text{L1}} = \|\mathbf{w}_k^{(t)} - \mathbf{w}^{(t)}\|_1$ 
12:    Total loss:  $\mathcal{L}_{\text{FSANL}} = \mathcal{L}_{\text{SANL}} + \lambda \mathcal{L}_{\text{L1}}$ 
13:    Update model:  $\mathbf{w}_k^{(t)} \leftarrow \mathbf{w}_k^{(t)} - \eta \nabla_{\mathbf{w}_k^{(t)}} \mathcal{L}_{\text{FSANL}}$ 
14:   end for
15: end for
16: return  $\mathbf{w}_k^{(t)}$ 

```

TABLE I: Test accuracy (%) on CIFAR-10 under IID setting with varying heterogeneous noise conditions.

Method	$\rho = 0.0$	$\rho = 0.5$	$\rho = 1.0$	$\rho = 1.0$
	$\tau = 0.0$	$\tau = 0.3$	$\tau = 0.2$	$\tau = 0.5$
FedAvg	92.89	85.21	61.05	46.21
FedProx	88.55	82.51	63.30	47.13
FedAvg + Mixup	94.37	89.50	71.76	57.03
FedCorr	93.08	92.95	84.21	70.61
FSANL	90.49	87.33	84.69	74.97

three-stage configuration with 5 rounds for the initial warm-up phase, 300 rounds for the noise estimation stage, and 350 rounds for server-side correction.

Our method FSANL is configured with $\alpha = 1$, $\gamma = 1$, and $\delta = 0.5$ to balance the contributions of the Normalized Cross Entropy, Normalized Negative Cross Entropy, and Symmetric Cross Entropy components. Within the SCE loss, we set $\eta = 0.1$ and $\zeta = 1.0$ for the forward and reverse cross-entropy terms, respectively. The global L1 regularization is applied with coefficient $\lambda = 0.1$. Mixup augmentation is used in FSANL and FedCorr, and is implemented using the same Beta distribution configuration described in FedCorr. All methods are trained using the same data partitions, client sampling strategy, and noise injection procedure to ensure a fair and controlled comparison.

TABLE II: Test accuracy (%) on CIFAR-10 under non-IID setting with varying heterogeneous noise conditions.

Method	$\rho = 0.5$	$\rho = 1.0$	$\rho = 1.0$
	$\tau = 0.3$	$\tau = 0.2$	$\tau = 0.5$
FedAvg	83.29	57.50	44.94
FedProx	80.39	57.07	38.00
FedAvg + Mixup	88.00	65.61	52.87
FedCorr	91.24	79.24	55.28
FSANL	86.15	81.08	72.41

B. Results**C. Results and Analysis**

Tables I and II present the classification performance of various federated learning methods on CIFAR-10 under both IID and non-IID data partitions, across a range of heterogeneous label noise settings parameterized by (ρ, τ) . The parameter ρ represents the fraction of clients with corrupted labels, while τ defines the minimum noise ratio within those noisy clients. Together, these parameters simulate a realistic federated environment where all clients may be noisy, but to varying extents.

Under the IID setting (Table I), our proposed method FSANL demonstrates competitive or superior performance in scenarios where all clients are noisy ($\rho = 1.0$) and especially under high noise intensities ($\tau = 0.5$). In the most challenging case with full corruption and high noise ratio ($\rho = 1.0, \tau = 0.5$), FSANL achieves 74.97% accuracy, significantly outperforming all baselines including FedCorr (70.61%) and FedAvg + Mixup (57.03%). This highlights FSANL's ability to operate effectively in uniformly noisy environments without assuming the existence of clean clients. Similarly, in the setting with moderate noise ($\rho = 1.0, \tau = 0.2$), FSANL maintains its edge, achieving the highest accuracy (84.69%) among all compared methods.

In contrast, FedCorr shows exceptional performance when a subset of clients are clean, particularly in the $\rho = 0.5, \tau = 0.3$ configuration, where it achieves 92.95%, a result consistent with its design that leverages clean-client identification for noise correction. However, its performance degrades more noticeably in fully noisy setups compared to FSANL, which remains stable and robust.

Figure 1 illustrates the accuracy trends of all methods over 700 communication rounds in a fully noisy IID scenario ($\rho = 1.0, \tau = 0.5$). Most baseline methods, including FedAvg, FedProx, and FedAvg + Mixup, show signs of overfitting or instability as training progresses. In contrast, FSANL and FedCorr demonstrate more stable and sustained improvements in accuracy. Notably, FSANL continues to improve steadily throughout training, highlighting its robustness and superior generalization in highly noisy federated settings.

The non-IID results (Table II) further demonstrate FSANL's resilience to the dual challenges of statistical heterogeneity and label noise. While FedCorr performs best when $\rho = 0.5$, FSANL achieves the top accuracy in both fully noisy scenarios: 81.08% under $\rho = 1.0, \tau = 0.2$, and 72.41% under $\rho = 1.0, \tau = 0.5$. These results affirm the importance of

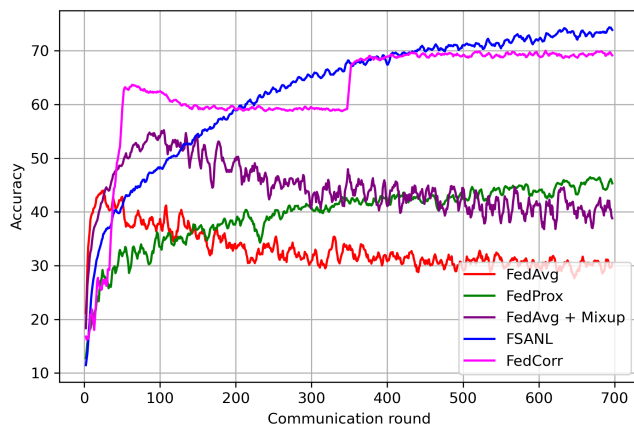


Fig. 1: Test accuracy on CIFAR-10 under IID setting with $\rho = 1.0$, $\tau = 0.5$ across 700 communication rounds.

FSANL's integrated robustness design, which does not rely on clean data assumptions and is well-suited to decentralized environments where both noise and heterogeneity are prevalent.

In summary, FSANL provides a robust and scalable alternative to state-of-the-art methods. While FedCorr excels when clean clients are available, FSANL surpasses all baselines in uniformly noisy and high-noise federated environments, both IID and non-IID, confirming its practical applicability for real-world federated learning systems.

V. CONCLUSION

In this work, we presented FSANL, a single-stage noise-robust federated learning framework designed to address heterogeneous label noise and client drift under non-IID conditions. By unifying symmetric and active negative losses with global regularization, FSANL achieved strong performance without clean client assumptions. In future work, we plan to extend our evaluation to additional datasets and model architectures.

REFERENCES

- [1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.
- [2] T. Nevratkai, A. Iliadou, G. Ntolkeras, I. Sfakianakis, L. Lazaridis, G. Maraslidis, N. Asimopoulos, and G. F. Fragulis, "A survey on federated learning applications in healthcare, finance, and data privacy/data security," in *AIP Conference Proceedings*, vol. 2909, no. 1. AIP Publishing LLC, 2023, p. 120015.
- [3] X. Ma, J. Zhu, Z. Lin, S. Chen, and Y. Qin, "A state-of-the-art survey on solving non-iid data in federated learning," *Future Generation Computer Systems*, vol. 135, pp. 244–258, 2022.
- [4] R. Pu, L. Yu, S. Zhan, G. Xu, F. Zhou, C. X. Ling, and B. Wang, "Fedelr: When federated learning meets learning with noisy labels," *Neural Networks*, vol. 187, p. 107275, 2025.
- [5] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proceedings of Machine learning and systems*, vol. 2, pp. 429–450, 2020.
- [6] Q. Li, B. He, and D. Song, "Model-contrastive federated learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 10713–10722.

- [7] Z. Wang, F. Yi, P. Gong, C. He, C. Jin, and W. Zhang, "Population normalization for federated learning," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 10214–10223.
- [8] J. Xu, Z. Chen, T. Q. Quek, and K. F. E. Chong, "Fedcorr: Multi-stage federated learning for label noise correction," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10184–10193.
- [9] X. Jiang and J. Zhang, "Fedclean: A general robust label noise correction for federated learning," in *Forty-second International Conference on Machine Learning*, 2025.
- [10] G. F. Ejigu, A. Adhikary, and C. S. Hong, "Relaxed contrastive learning for robust federated models with noisy labels and limited clients," in *2025 27th International Conference on Advanced Communications Technology (ICACT)*. IEEE, 2025, pp. 1–6.
- [11] Y. Wang, X. Ma, Z. Chen, Y. Luo, J. Yi, and J. Bailey, "Symmetric cross entropy for robust learning with noisy labels," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 322–330.
- [12] X. Ye, X. Li, T. Liu, Y. Sun, W. Tong *et al.*, "Active negative loss functions for learning with noisy labels," *Advances in Neural Information Processing Systems*, vol. 36, pp. 6917–6940, 2023.
- [13] X. Li, M. Jiang, X. Zhang, M. Kamp, and Q. Dou, "Fedbn: Federated learning on non-iid features via local batch normalization," *arXiv preprint arXiv:2102.07623*, 2021.
- [14] Y. L. Tun, C. M. Thwal, Y. M. Park, S.-B. Park, and C. S. Hong, "Federated learning with intermediate representation regularization," in *2023 IEEE International conference on big data and smart computing (BigComp)*. IEEE, 2023, pp. 56–63.
- [15] D. Caldarola, B. Caputo, and M. Ciccone, "Improving generalization in federated learning by seeking flat minima," in *European Conference on Computer Vision*. Springer, 2022, pp. 654–672.
- [16] A. Z. Tan, H. Yu, L. Cui, and Q. Yang, "Towards personalized federated learning," *IEEE transactions on neural networks and learning systems*, vol. 34, no. 12, pp. 9587–9603, 2022.
- [17] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.
- [18] G. F. Ejigu, S. H. Hong, and C. S. Hong, "Robust federated learning with local mixed co-teaching," in *2023 International Conference on Information Networking (ICOIN)*. IEEE, 2023, pp. 277–281.
- [19] T. Yoon, S. Shin, S. J. Hwang, and E. Yang, "Fedmix: Approximation of mixup under mean augmented federated learning," *arXiv preprint arXiv:2107.00233*, 2021.
- [20] G. F. Ejigu, K. Kim, and C. S. Hong, "Mitigating label noise in federated learning with regularized features and robust loss," *IEEE Transactions on Artificial Intelligence*, 2025.
- [21] A. Krizhevsky, G. Hinton *et al.*, "Learning multiple layers of features from tiny images," 2009.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.