HYDRA-FL: Hybrid Knowledge Distillation for Robust and Accurate Federated Learning NeurIPS 2024

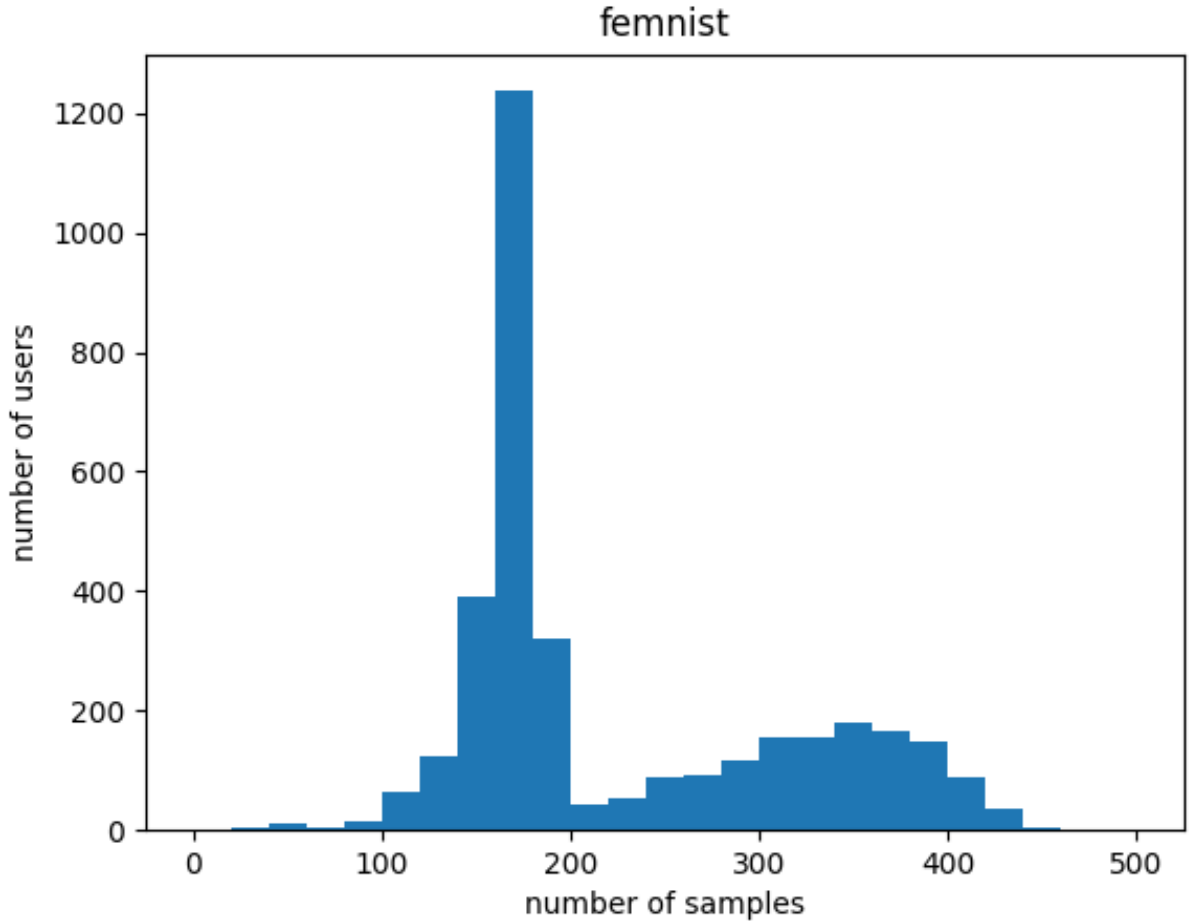
Momin Ahmad Khan Yasra Chandio Fatima Muhammad Anwar

University of Massachusetts Amherst

University of Massachusetts Amherst

NEURAL INFORMATION PROCESSING SYSTEMS

Data Heterogeneity in FL



University of Massachusetts Amherst

NEURAL INFORMATION PROCESSING SYSTEMS

Case Study

NeurIPS 2022

Preservation of the Global Knowledge by Not-True Distillation in Federated Learning

Gihun Lee*, Minchan Jeong*, Yongjin Shin, Sangmin Bae, Se-Young Yun
KAIST
{opcrisis, mcjeong, yj.shin, bsmn0223, yunseyoung}@kaist.ac.kr

Abstract

In federated learning, a strong global model is collaboratively learned by aggregating clients' locally trained models. Although this precludes the need to access clients' data directly, the global model's convergence often suffers from data heterogeneity. This study starts from an analogy to continual learning and suggests that *forgetting* could be the bottleneck of federated learning. We observe that the global model forgets the knowledge from previous rounds, and the local training induces forgetting the knowledge outside of the local distribution. Based on our findings, we hypothesize that tackling down forgetting will relieve the data heterogeneity problem. To this end, we propose a novel and effective algorithm, *Federated Not-True Distillation* (FedNTD), which preserves the global perspective on locally available data only for the *not-true* classes. In the experiments, FedNTD shows state-of-the-art performance on various setups without compromising data privacy or incurring additional communication costs[1].

CVPR 2021

Model-Contrastive Federated Learning

Qinbin Li
National University of Singapore
qinbin@comp.nus.edu.sg

Bingsheng He
National University of Singapore
hebs@comp.nus.edu.sg

Dawn Song
UC Berkeley
dawnsong@berkeley.edu

Abstract

Federated learning enables multiple parties to collaboratively train a machine learning model without communicating their local data. A key challenge in federated learning is to handle the heterogeneity of local data distribution across parties. Although many studies have been proposed to address this challenge, we find that they fail to achieve high performance in image datasets with deep learning models. In this paper, we propose MOON: model-contrastive federated learning. MOON is a simple and effective federated learning framework. The key idea of MOON is to utilize the similarity between model representations to correct the local training of individual parties, i.e., conducting contrastive learning in model-level. Our extensive experiments show that MOON significantly outperforms the other state-of-the-art federated learning algorithms on various image classification tasks.

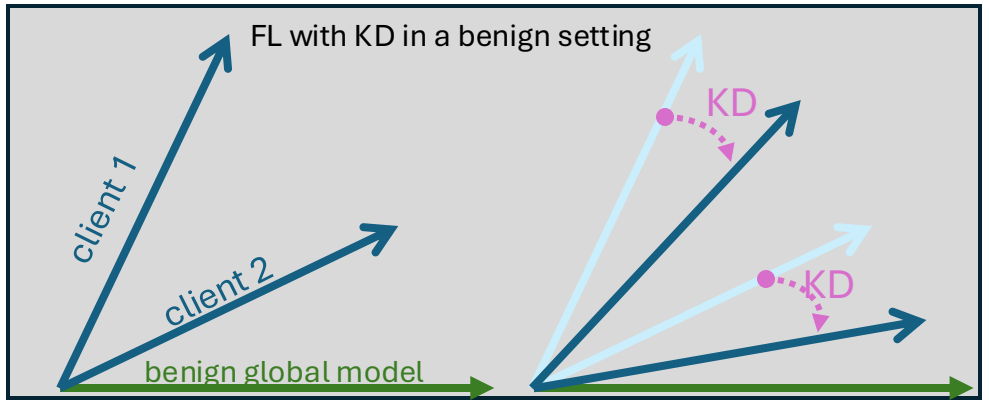
A key challenge in federated learning is the heterogeneity of data distribution on different parties [20]. The data can be non-identically distributed among the parties in many real-world applications, which can degrade the performance of federated learning [22, 29, 24]. When each party updates its local model, its local objective may be far from the global objective. Thus, the averaged global model is away from the global optima. There have been some studies trying to address the non-IID issue in the local training phase [28, 22]. FedProx [28] directly limits the local updates by ℓ_2-norm distance, while SCAFFOLD [22] corrects the local updates via variance reduction [19]. However, as we show in the experiments (see Section 4), these approaches fail to achieve good performance on image datasets with deep learning models, which can be as bad as FedAvg.

In this work, we address the non-IID issue from a novel perspective based on an intuitive observation: *the global*

University of Massachusetts Amherst

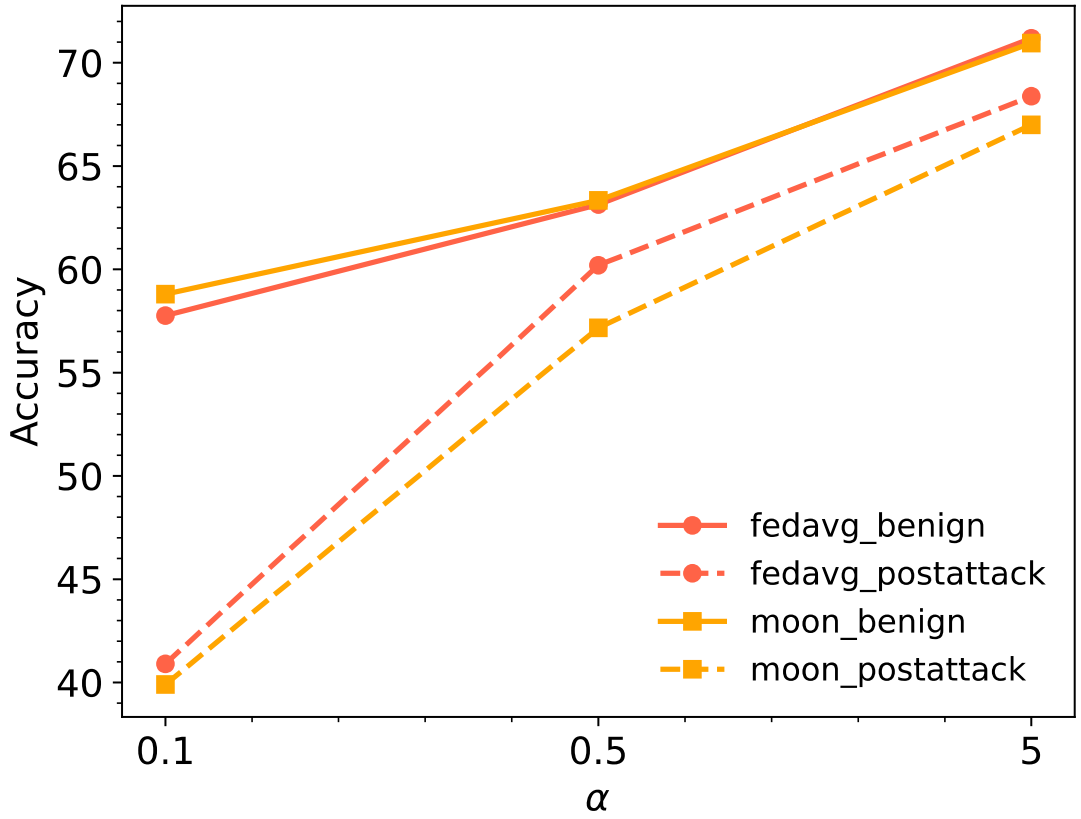
NEURAL INFORMATION PROCESSING SYSTEMS

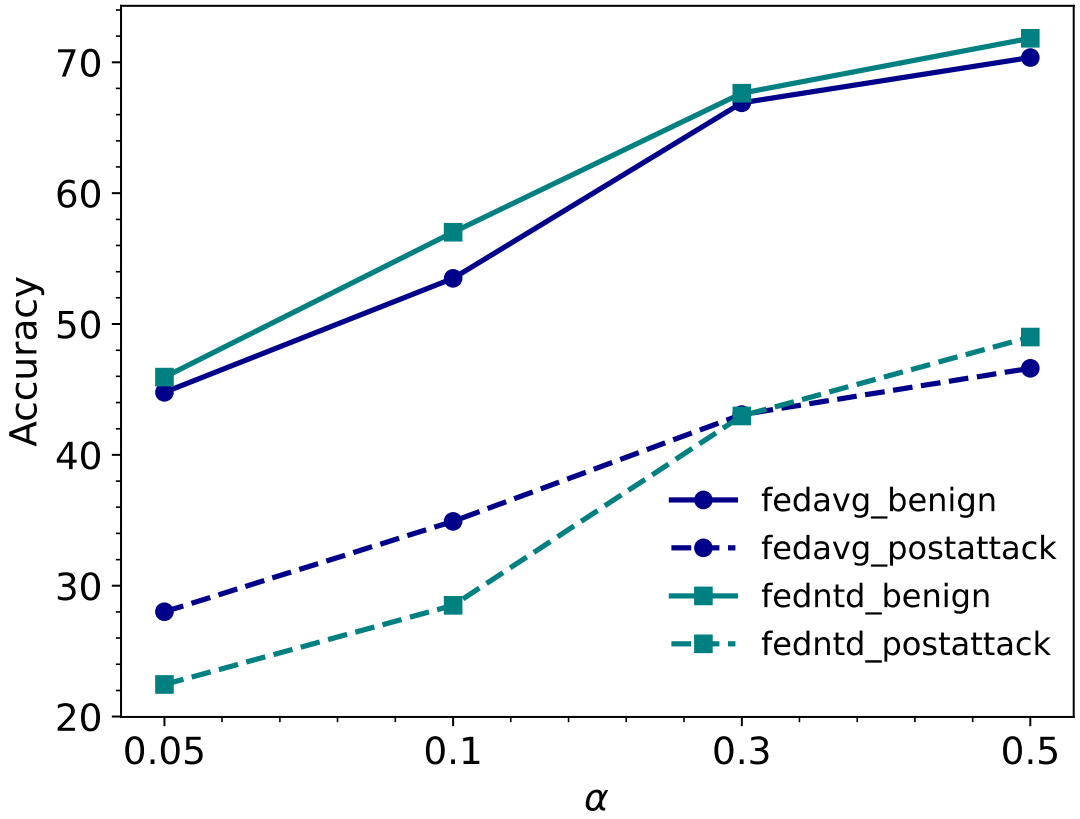
Knowledge Distillation in FL



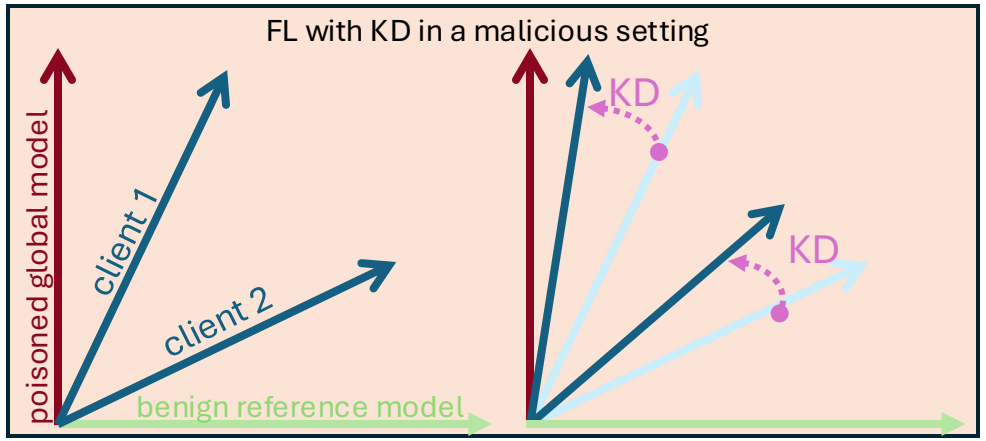
$$\mathcal{L} = (1-\beta)\cdot\mathcal{L}_{CE}(y_s^i, y^i) + \beta\cdot\mathcal{L}_{KL}(softmax(y_s^i/\tau)||softmax(y_t^i/\tau))$$

Impact of Data Heterogeneity



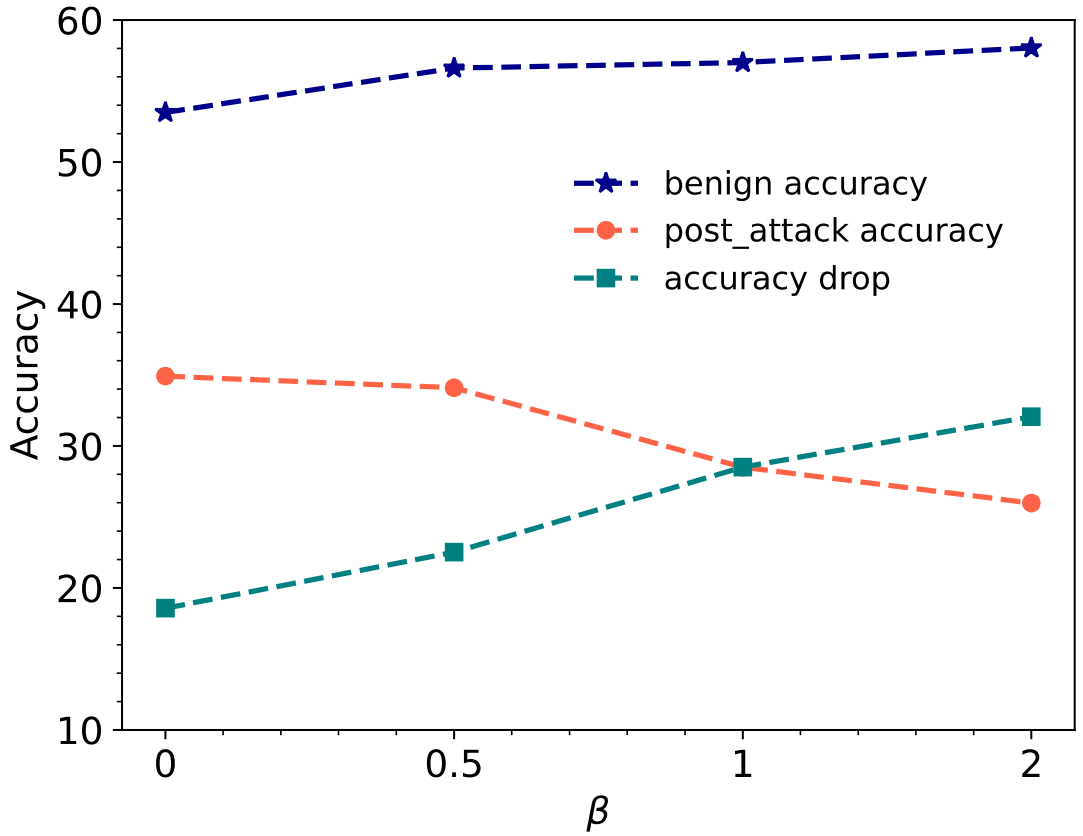


Attack Amplification

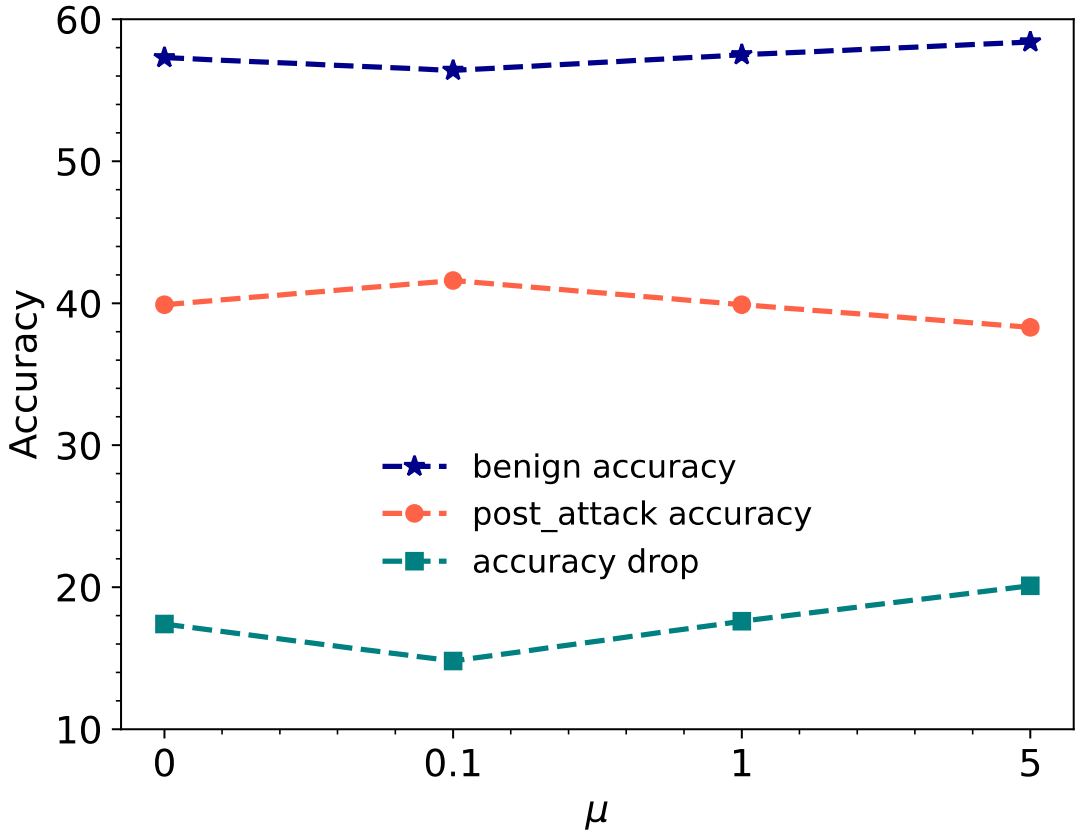


$$\mathcal{L} = (1-\beta)\cdot\mathcal{L}_{CE}(y_s^i, y^i) + \beta\cdot\mathcal{L}_{KL}(softmax(y_s^i/\tau)||softmax(y_t^i/\tau))$$

Attack Amplification

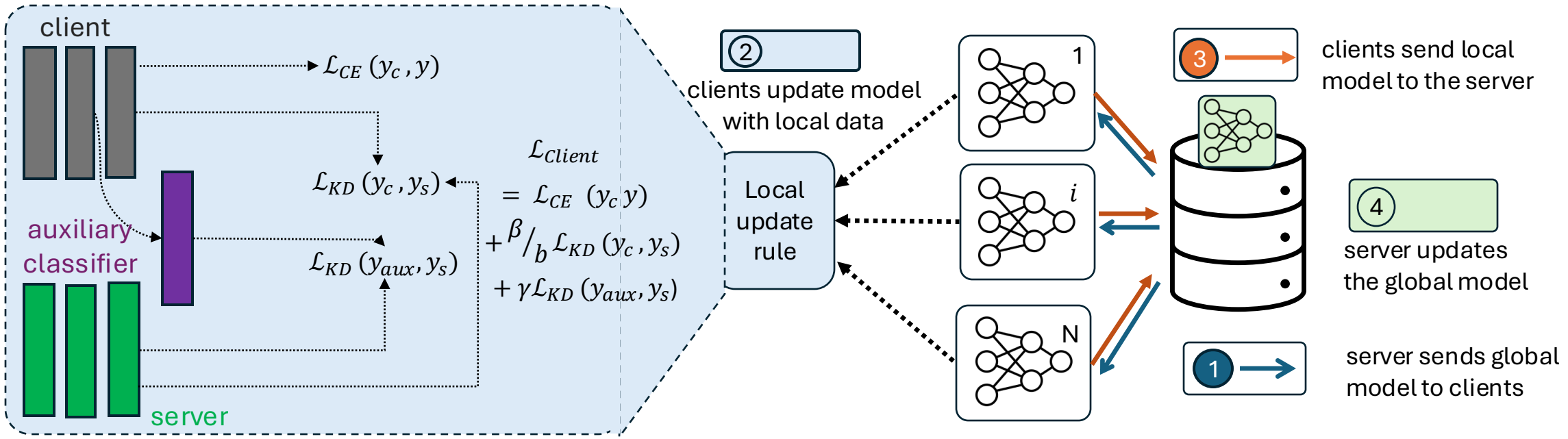


FedNTD



MOON

Our Solution: HYDRA-FL



Quantitative Analysis

Table 1: Test accuracy for three techniques on three datasets. In the no-attack setting, (↑ ↓) shows comparison to FedAvg. In the attack setting, we use bold if our technique outperforms FedNTD.

Dataset	**MNIST**		**CIFAR10**						**CIFAR100**	
			$\alpha = 0.05$		$\alpha = 0.1$		$\alpha = 0.5$			
Techniques	*no attack*	*attack*	*no attack*	*attack*	*no attack*	*attack*	*no attack*	*attack*	*no attack*	*attack*
Fedavg	92.12	74.48	44.69	31.27	54.67	35.67	70.57	48.27	26.17	12.92
FedNTD	93.03↑	58.09	46.94↑	21.72	56.95↑	32.61	71.79↑	52.51	29.1↑	13.92
HYDRA-FL(Ours)	92.69↑	**76.67**	46.92↑	**25.15**	57.12↑	**34.25**	71.22↑	**52.57**	28.9↑	**14.33**

Table 2: Test accuracy for three techniques on three datasets. In the no-attack setting, (↑ ↓) shows comparison to FedAvg. In the attack setting, we use bold if our technique outperforms MOON.

Dataset	**MNIST**		**CIFAR10**						**CIFAR100**	
			$\alpha = 0.1$		$\alpha = 0.5$		$\alpha = 5$			
Methods	*no attack*	*attack*	*no attack*	*attack*	*no attack*	*attack*	*no attack*	*attack*	*no attack*	*attack*
Fedavg	88.02	77.55	57.76	40.9	63.14	60.2	71.19	68.38	28.36	24.21
MOON	91.13↑	72.32	58.8↑	39.9	63.34↑	57.17	70.95↓	67	29.34↑	23.81
HYDRA-FL(Ours)	92.04↑	**76.65**	60.1↑	**43.6**	63.32↑	**59.93**	70.55↓	**68.4**	29.48↑	**25.18**

University of Massachusetts Amherst

NEURAL INFORMATION PROCESSING SYSTEMS

Qualitative Analysis