Federated Learning Under Non-IID Data

Progress report

Elena Voss
Department of Computer Science

August 27, 2026

Outline

Background

Method

Results

Conclusion

UNIVERSITÄT DES SAARLANDES

Why Non-IID Data Is Hard

- Federated averaging assumes roughly similar data across clients
- Real deployments show heavy label and feature skew per client
- Naive FedAvg diverges or converges slowly under strong skew

UNIVERSITÄT DES SAARLANDES

Related Approaches

Regularization-based

- FedProx
- SCAFFOLD

Clustering-based

- Client clustering
- Personalized layers

UNIVERSITÄT DES SAARLANDES

Our Approach

- Adaptive client weighting based on local gradient variance
- Periodic re-clustering as client distributions drift over time
- No extra communication round beyond standard FedAvg

UNIVERSITÄT DES SAARLANDES

Weighting Rule

Client i's contribution weight at round t:

$$w_i^{(t)} = \frac{1/\sigma_i^{(t)}}{\sum_j 1/\sigma_j^{(t)}}$$

where $\sigma_i^{(t)}$ is the local gradient variance estimated from the last three local epochs.

UNIVERSITÄT DES SAARLANDES

Benchmark Setup

Dataset	Clients	Skew (Dirichlet α)
CIFAR-10	100	0.1
FEMNIST	200	0.3

UNIVERSITÄT DES SAARLANDES

Accuracy Comparison

Method	CIFAR-10	FEMNIST
FedAvg	61.2%	78.4%
FedProx	65.8%	80.1%
Ours	**71.3%**	**83.6%**

Table: Top-1 accuracy after 100 communication rounds.

UNIVERSITÄT DES SAARLANDES

Summary

- Variance-based weighting outperforms uniform and loss-based baselines
- Gains grow with skew severity
- No additional communication cost over standard FedAvg

UNIVERSITÄT DES SAARLANDES

Thank You

Questions?

elena.voss@example.edu

UNIVERSITÄT DES SAARLANDES