CA-AFP: Cluster-Aware Adaptive Federated Pruning for Communication-Efficient and Personalized Learning

Federated Learning Course Project

Presented by

Om Govind Jha(22227) **Harsh Shukla(22140)**

Background

- **HAR on Edge Devices**
  - **Continuous Monitoring:** Patients and athletes require continuous, on-device tracking for **efficient, real-time results**.
  - **Deployment Constraints:** Constant processing on wearables requires minimizing battery usage and memory footprint.
- **The Heterogeneity Hurdle (Non-IID)**
  - User data varies drastically (physiology, gait, sensor placement).
- **Impact:** Global models struggle to generalize
- **Our Objective**
  - To simultaneously deliver **High Accuracy**, **High Sparsity**, and **Equitable Performance (Fairness)** for all clients.

Research Objective and Existing Gaps

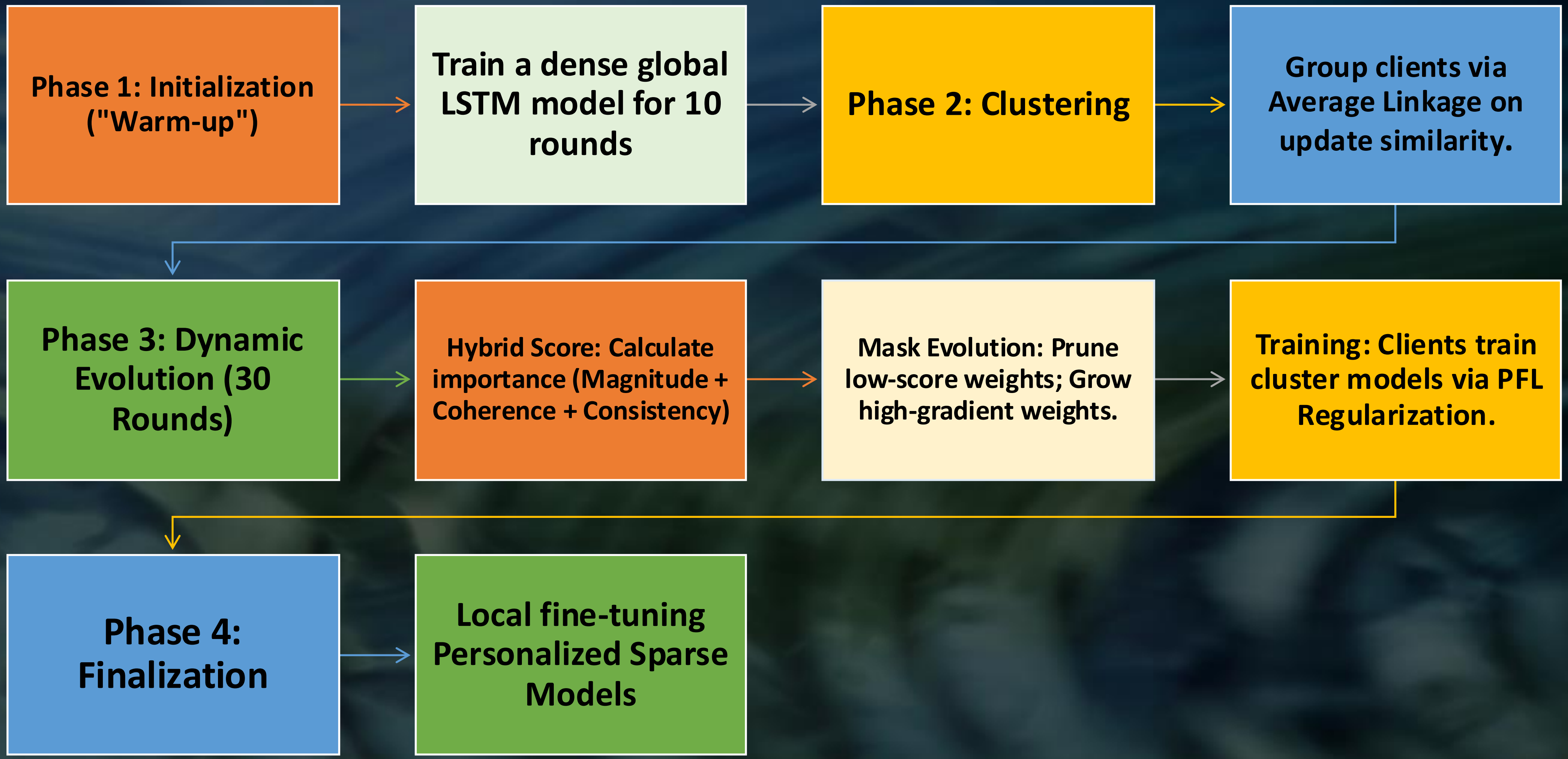
Research Objectives

- **Efficiency:** Achieve **communication reduction** via model sparsity.
- **Robustness:** Maintain high accuracy despite severe **Non-IID data heterogeneity**.
- **Fairness:** Equitable performance for **all** clients

Existing Frameworks

- **Static Constraints:** Conventional pruning is often a "one-way street" (static); once critical weights are removed, they cannot be recovered.
- **Metric Insensitivity:** Standard magnitude pruning relies on **global averages**, often discarding weights that are locally critical for specific minority clients.
- **Structural Instability:** Dynamic methods often lack the necessary clustering stability to converge on highly heterogeneous sensor data.
- **The Unexplored Intersection:** Few existing works optimize for Efficiency, Robustness, and Fairness simultaneously.

CA-AFP Methodology Workflow



Key Properties of Methodology

Consensus-Driven Hybrid Score(Parameter level)

Magnitude: Weight Strength . "Is this connection currently strong?"

Coherence : Value Agreement . "Do all clients have similar weight values?" (Low Variance is good).

Consistency : Direction Agreement. "Do client gradients align?" (High Sign Agreement is good).

Preserves weights that are **universally important** to the cluster.

Dynamic Pruning Mechanism

A continuous **Prune-and-Grow** cycle during training.

- Removes weak connections based on the Hybrid Score.
- Reactivates dead connections that accumulate **high gradients**.

Allows the model to correct early mistakes and adapt to new data.

Structural Stability via Average Linkage

By averaging distances, clusters are defined by their **centroid** rather than extreme outliers.

Prevents the " Cluster collapse" caused by conflicting outlier clients.

PFL Training Objective

Standard Cross-Entropy loss ensures the model learns to classify the specific client's activities correctly.

Cluster Alignment (Regularization)

preventing the personal model from drifting too far from the shared cluster consensus

$$L_{total} = \underbrace{L_{CE}(y, \hat{y})}_{\text{Local Accuracy}} + \underbrace{\frac{\lambda}{2}||W_{personal} - W_{cluster}||^2}_{\text{Cluster Alignment}}$$

Experimental Framework and Dataset

Datasets & Non-IID Partitioning

- **Dataset 1: WISDM** (Smartphone Accelerometer).
  - *Specs:* 36 Users, 3-axis raw accelerometer data (20 Hz).
  - *Preprocessing:* Sliding windows (T=200 , 50% overlap).
- **Secondary Dataset 2: UCI-HAR**
- *Specs:* 30 Users, **9-channel input** (Accel + Gyro), waist-mounted.
- **Non-IID Split :** Clients are partitioned by activity type to force heterogeneity:
  - **Cluster 0:** Dynamic Motion (Walking, Jogging).
  - **Cluster 1:** Vertical Motion (Upstairs, Downstairs).
  - **Cluster 2:** Stationary (Sitting, Standing).
- **Data Skew:** Sample counts follow a **Log-Normal** distribution; Label ratios follow a **Dirichlet** distribution($\alpha = 0.5$).

Model Architecture (Uniform Across All Methods)

- **Type:** Deep LSTM Network.
- **Structure:**
  - 2 LSTM Layers (64 units) + Dropout (0.3).
  - 1 Dense Layer (32 units, ReLU) + Dropout (0.2).
  - Softmax Output (6 Classes).
- **Optimizer used:** Adam

Federated Protocol

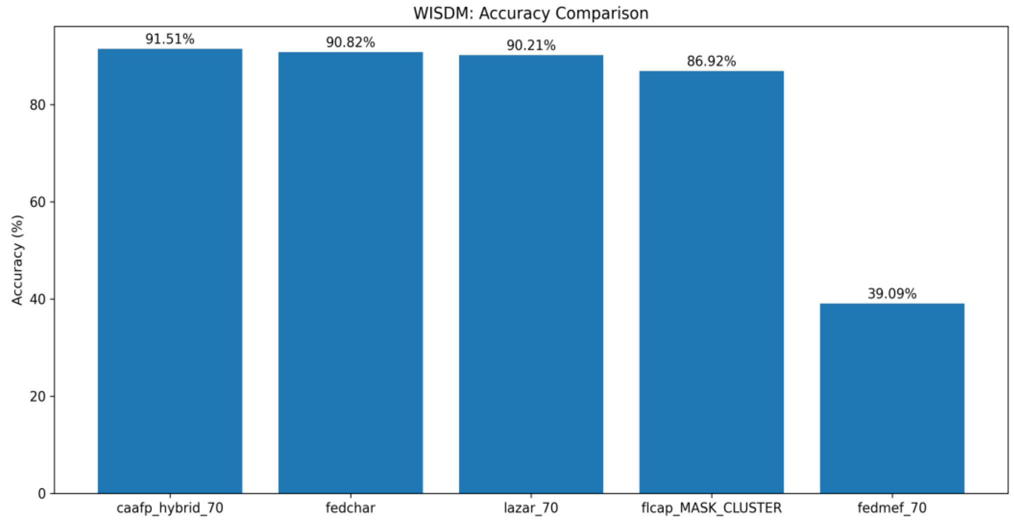
- **Population:** N=30 Clients, Selection fraction 0.33 (10 clients/round).
- **Training Budget: 40 Rounds** Total (10 Warm-up + 30 Dynamic).
- **Local Compute:** 3 Epochs per round.
- **Sparsity Target:** Fixed at **70%** for all sparse methods.

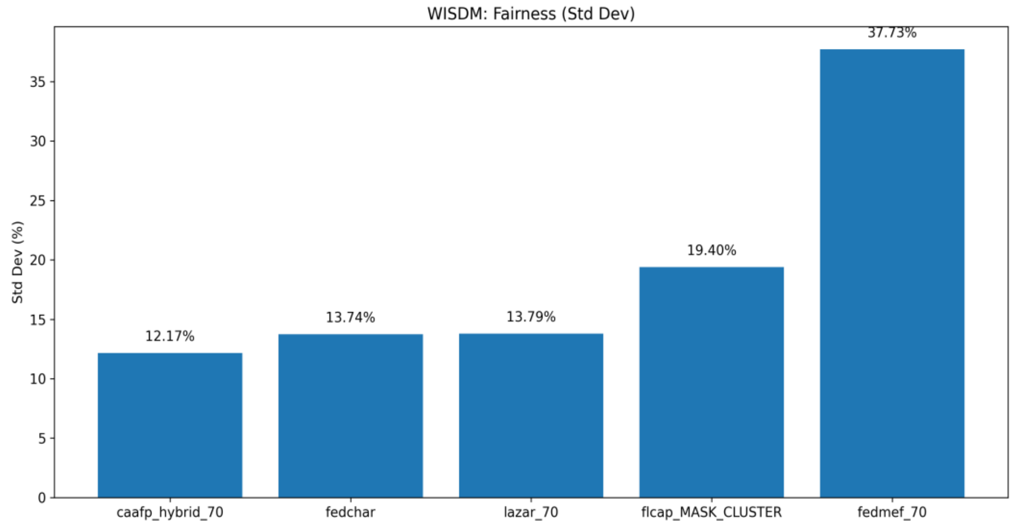
Results- Comparison Table

Method	Dataset	Avg Accuracy(Higher is Better)	Fairness (Std Dev)(Lower is Better)
CA-AFP (Ours)	**WISDM**	**91.51%**	**12.17%**
	UCI-HAR	**94.63%**	**7.47%**
FedCHAR	WISDM	90.82%	13.74%
	UCI-HAR	94.16%	8.66%
FLCAP	WISDM	86.92%	19.40%
	UCI-HAR	90.56%	13.75%
LAZAR	WISDM	90.21%	13.79%
	UCI-HAR	90.64%	12.02%
FedMef	WISDM	39.09%	37.73%
	UCI-HAR	34.74%	35.55%

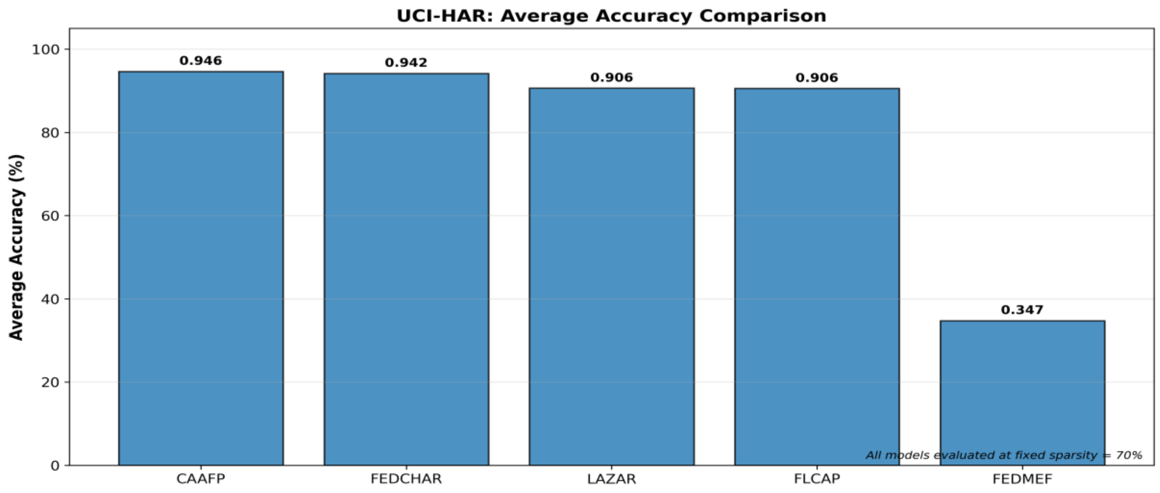
Accuracy and Fairness

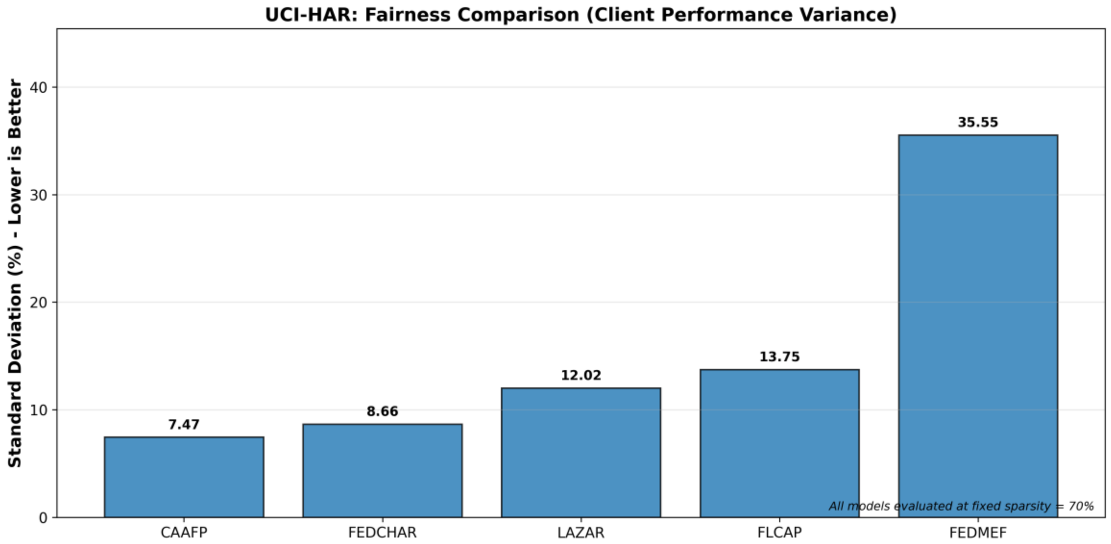
WISDOM DATSET





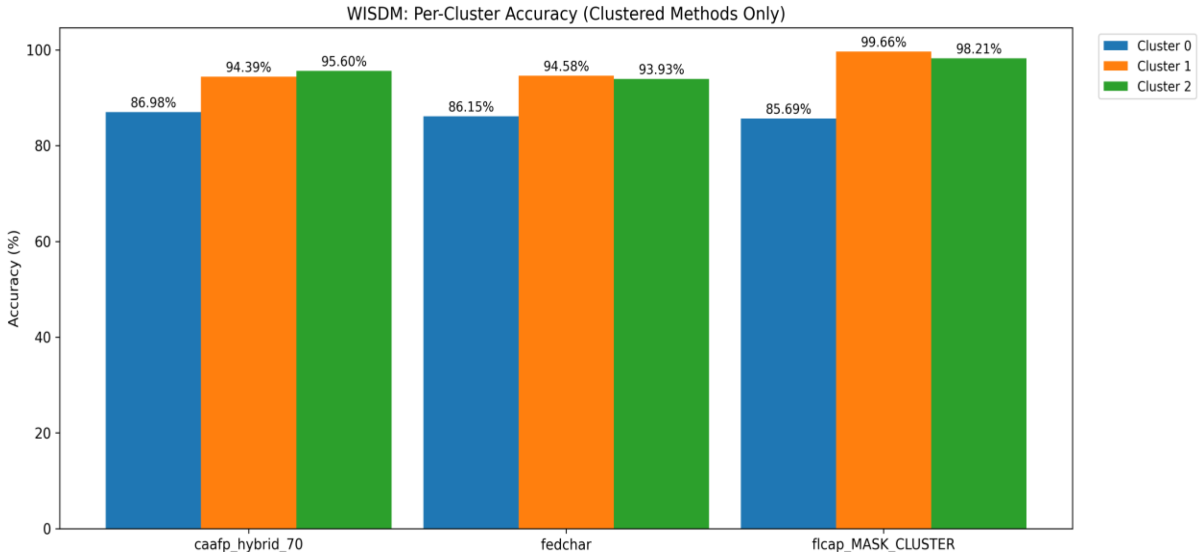
UCI-HAR DATASET

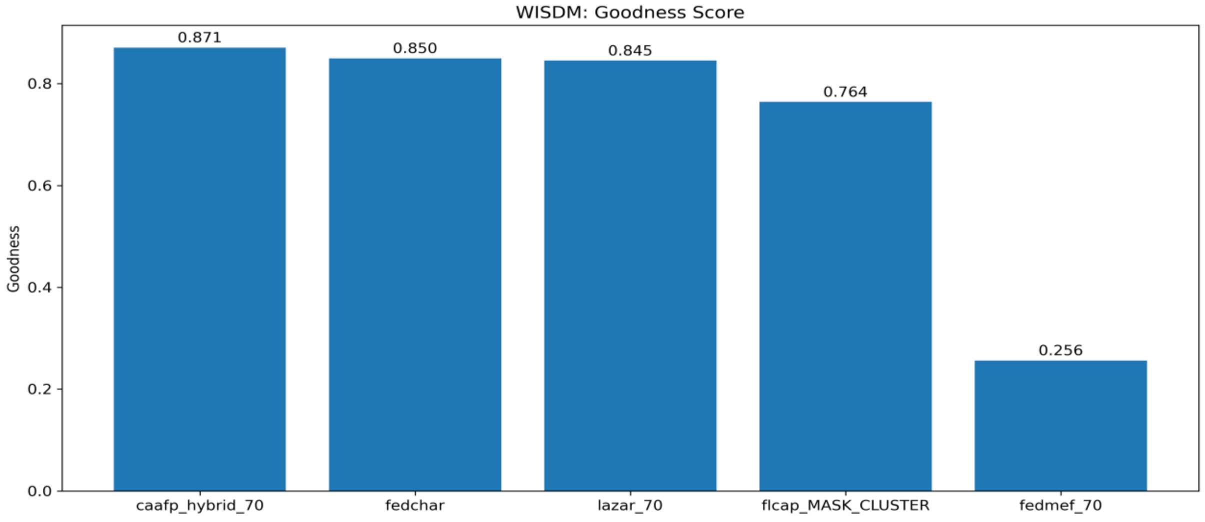




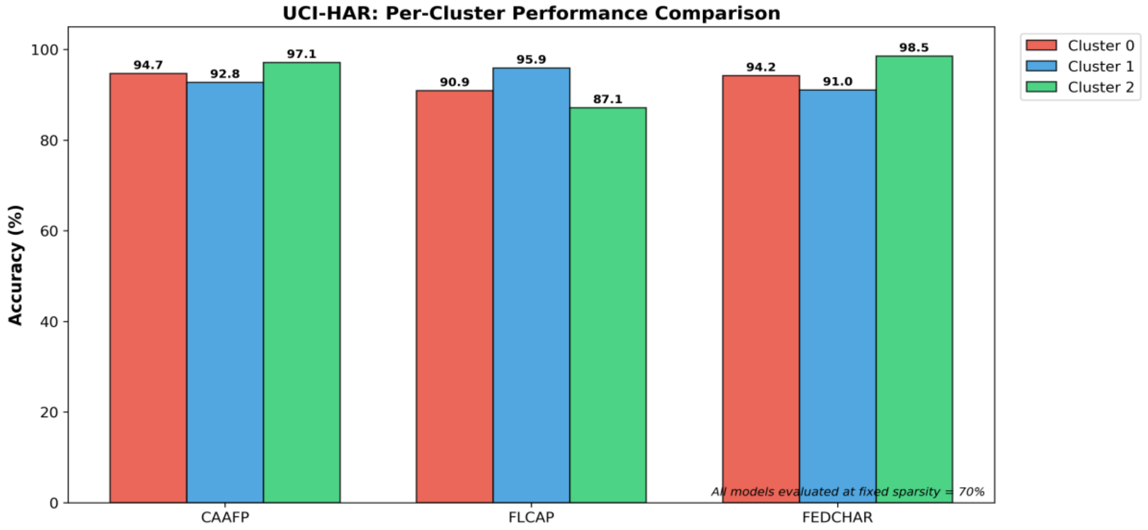
Goodness Score and Cluster Performance

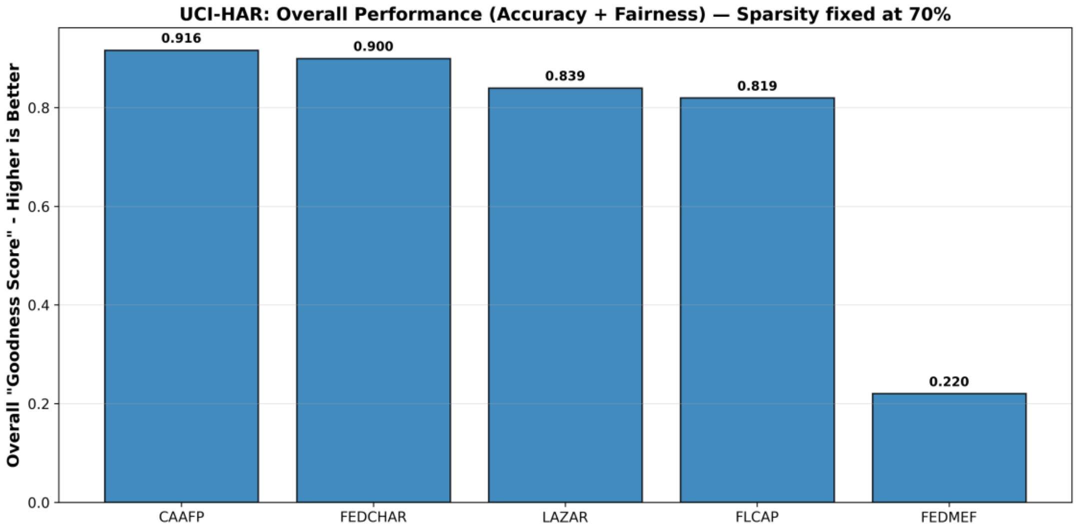
WISDOM DATSET





UCI-HAR DATASET





DISCUSSION

•**Current Limitations:** While communication-efficient, our method introduces slight local computational overhead (gradient calculations) and requires a brief dense "warm-up" phase to stabilize clustering.

• **Future Deployment:** We can validate energy efficiency on physical edge hardware (e.g., Raspberry Pi) and extend the framework to be more robust as user behaviors evolve over time.

• **Privacy Enhancements:** We can integrate **Differential Privacy** and **Secure Aggregation** concepts to further protect gradient updates without compromising the clustering quality.

THANK YOU

Questions?