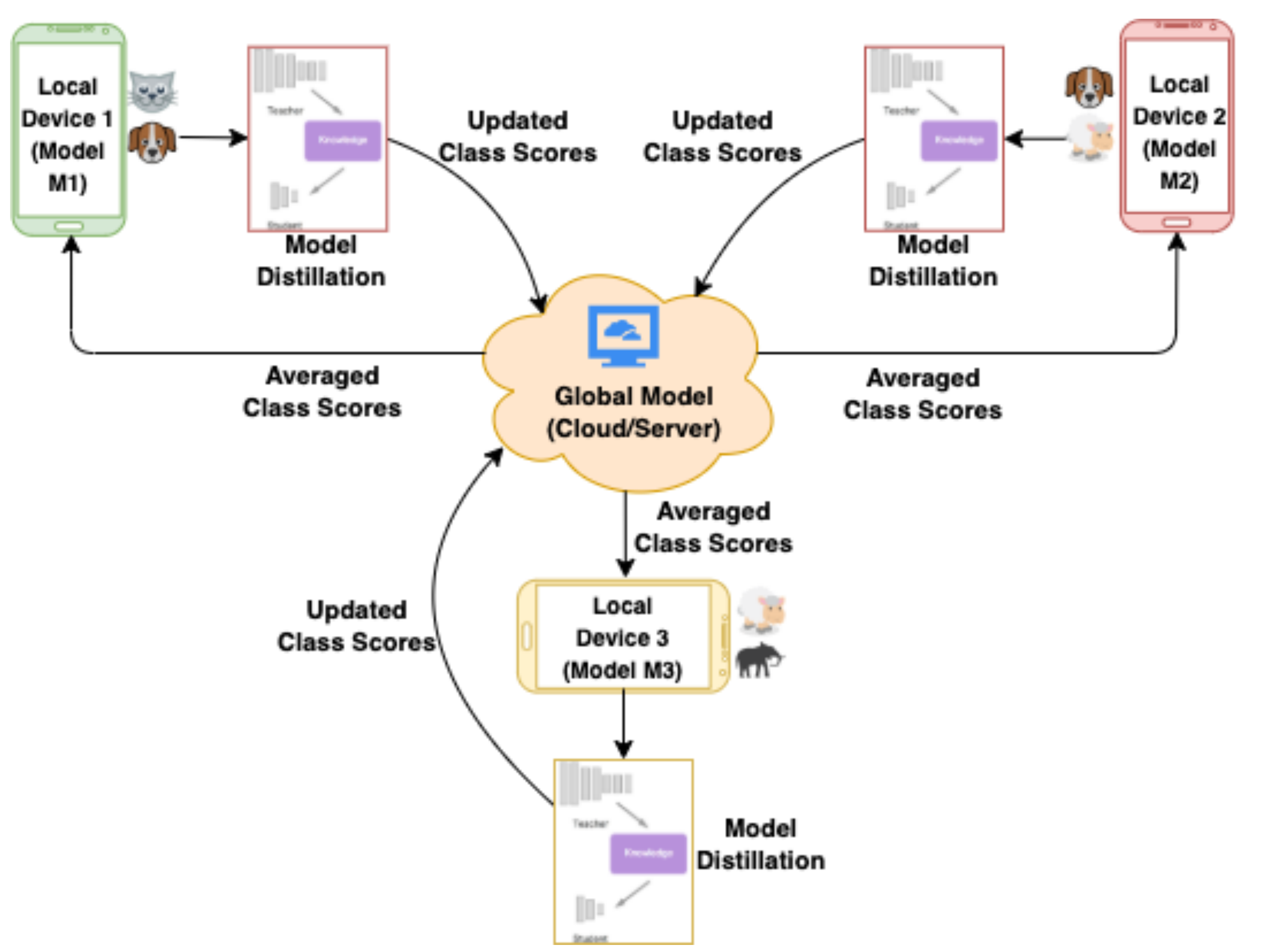
Federated Learning with Heterogeneous Labels and Models for Mobile Activity Monitoring

Gautham Krishna Gudur, Global AI Accelerator, Ericsson
Satheesh Kumar Perepu, Ericsson Research

Motivation

- Various mobile health applications require modeling of user behavior through **Human Activity Recognition (HAR)**.
- On-device Federated Learning – characterization from multiple user devices for effective personalized activity monitoring.
- Address *statistical heterogeneities* in HAR – label (activity), model heterogeneities, non-IID data.
- Federated label-based aggregation, which leverages overlapping information gain across activities using a *Model Distillation* update.

Overall Block Diagram



Dataset and Preprocessing

Heterogeneity Human Activity Recognition (HHAR) Dataset

- Accelerometer data of four different mobile phones and six daily activities
- Two second non-overlapping windows
- Decimation (downsample) all windows to least sampling frequency (50 Hz)
- Discrete Wavelet Transform (DWT) to extract temporal and frequency information, with Approximate coefficients only

Algorithm 1 Our Proposed Framework

Input: Public Dataset $\mathcal{D}_0\{x_0, y_0\}$, Private Datasets $\mathcal{D}_m^i$, Total users M, Total iterations I, LabelSet l_m for each user
Output: Trained Model scores f_G^I
Initialize $f_G^0 = \mathbf{0}$ (Global Model Scores)
for $i = 1$ **to** I **do**
  for $m = 1$ **to** M **do**
    Build: Model $\mathcal{D}_m^i$ and predict $f_{\mathcal{D}_m^i}(x_0)$
    Local Update (Model Distillation):
    Build a distilled model only on respective local model labels with the new data $\mathcal{D}_m^i$.
  end for
  Global Update: Update label wise

$$f_G^{i+1} = \sum_{m=1}^{M} \beta_m f_{\mathcal{D}_m^i}(x_0), \text{ where}$$

$$\beta = \begin{cases} 1 & \text{If labels are unique} \\ \mathrm{acc}(f_{\mathcal{D}_m^{i+1}}(x_0)) & \text{if labels are not unique} \end{cases}$$

end for

Experiments – Distribution of Models, Activities

	User 1	User 2	User 3	Global User
Architecture	2-Layer CNN (16, 32) Softmax Activation	3-Layer CNN (16, 16, 32) ReLU Activation	3-Layer ANN (16, 16, 32) ReLU Activation	–
Activities	{Sit, Walk}	{Walk, Stand}	{Stand, StairsUp}	{Sit, Walk, Stand, StairsUp}
Activity Windows per iteration	{2000, 2000} = 4000	{2000, 2000} = 4000	{2000, 2000} = 4000	{2000, 2000, 2000, 2000} = 8000

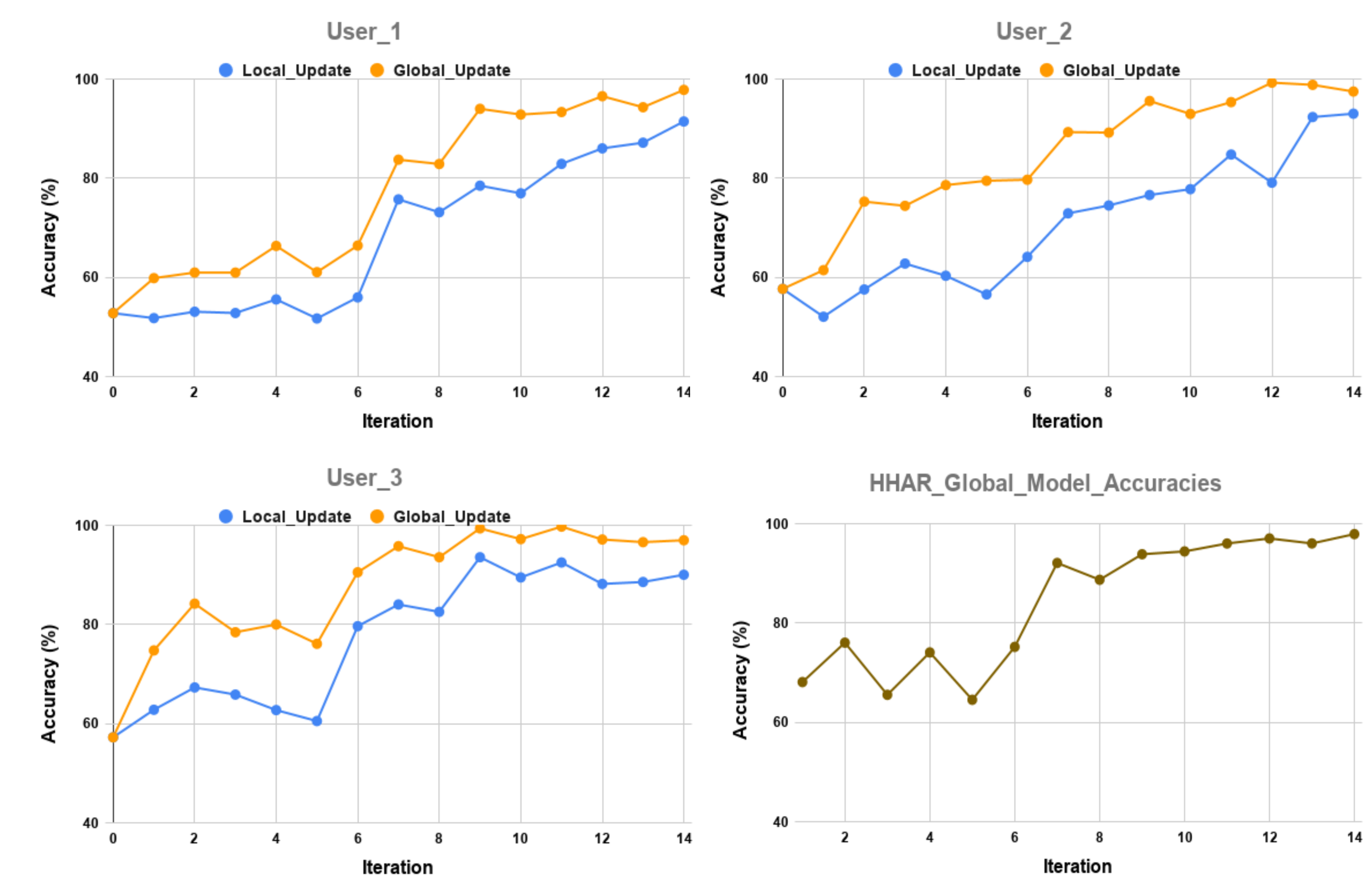
Heterogeneous Model Architectures, Activities and Activity Windows per Iterations across all users

Student Distillation Model: Simple 2-Layer ANN model (8, 16)

Iteration	New Model Architecture
User_1 Iteration_10	3-Layer ANN (16, 16, 32) ReLU Activation
User_1 Iteration_14	1-Layer CNN (16) Softmax Activation
User_2 Iteration_6	3-Layer CNN (16, 16, 32) Softmax activation
User_3 Iteration_5	4-Layer CNN (8, 16, 16, 32) Softmax activation

Model Heterogeneities across Federated Learning Iterations

Results



Iterations vs Local Update and Global Update Accuracies across all three users

	Local Update	Global Update	Acc. Increase
User 1	68.38	77.61	9.23
User 2	70.82	84.4	13.58
User 3	77.68	87.9	10.22
Average	72.293	83.303	11.01

On-Device Performance

Process	Computation Time
Training time per epoch in an FL iteration (i)	$\sim$1.7 sec
Inference time	$\sim$15 ms
Discrete Wavelet Transform	$\sim$0.45 ms
Decimation	$\sim$4.6 ms

On-Device Performance Metrics

References

[1] H Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas, (2017), "Communication-efficient learning of deep networks from decentralized data" In: 20th International Conference on Artificial Intelligence and Statistics (AISTATS).

[2] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean, (2015), "Distilling the knowledge in a neural network," In: NIPS Deep Learning and Representation Learning Workshop.