

---

## A EXPERIMENT DETAILS

### A.1 DATASET

**Opportunity** This is the human activity recognition dataset that consists of the on-body sensor records from 4 participants doing kitchen activities. For each participant, they recorded five different runs. In this dataset, we utilize the 18 kitchen activities as 18 classes and consider two modalities, accelerometers (Acce) and gyroscope (Gyro) for our experiment. We use runs 4 and 5 from participants 2 and 3 as the testing data, runs 4 and 5 from participants 1 and 4 as proxy data, and the remaining data for training data, respectively.

**mHealth** This dataset consists of the on-body sensor records from 10 participants doing 13 daily living and exercise activities. In this dataset, we consider three modalities, accelerometer (Acce), gyroscope (Gyro), and magnetometer (Mage), in our experiments. We randomly use data from one participant’s data as testing data, one participant’s data as training data, and the remaining data as training data.

**UR Fall Detection** This dataset contains 70 video clips of human activities with three classes. We consider three modalities RGB camera (RGB), depth camera (Depth), and sensory data of each video frame from accelerometers (Acce) in the experiments. As mentioned above, we follow the setup from MM-FedAvg work to generate RGB data. We randomly sample 1/10 of data as testing data, 1/10 of data as proxy data, and the remaining data as training data.

Table 1: Detailed information of the multimodal autoencoder architecture used in our experiment

Components	Parameter & Shape
Encoder A	(lstm): LSTM(3, 32, batch_first=True) (lstm2): LSTM(32, 2, batch_first=True)
Decoder A	(lstm): LSTM(2, 32, batch_first=True) (lstm2): LSTM(32, 3, batch_first=True)
Encoder B	(lstm): LSTM(512, 32, batch_first=True) (lstm2): LSTM(32, 2, batch_first=True)
Decoder B	(lstm): LSTM(2, 32, batch_first=True) (lstm2): LSTM(32, 512, batch_first=True)

### A.2 MODEL ARCHITECTURE

In table 1, we summarize the autoencoder architecture that we used in this paper.

## B ADDITIONAL RESULTS

### B.1 GLOBAL PERFORMANCE

In this section, we provide some supplementary results for the mHealth and Opp datasets. Particularly, we show the curve of global performance of two datasets in Fig. 1 and Fig. 2. As can be seen, for the mHealth dataset, FedMEKT achieves a stable convergence with better performance, while MM-FedAvg illustrates the degradation in the performance in the final rounds. Regarding Opp dataset, our proposed method obtains the comparable performance to the baselines.

### B.2 LOCAL PERFORMANCE

In this section, we provide the supplementary results in the local performance for all datasets. We depict the curve of personalized performance of all three datasets in Fig. 3, Fig. 4, and Fig. 5. Our method shows a more stable performance in most combinations of modalities. We still achieve comparable performance without degradation for the Acce-Mage combination in the mHealth dataset even though our method cannot outperform MM-FedAvg.

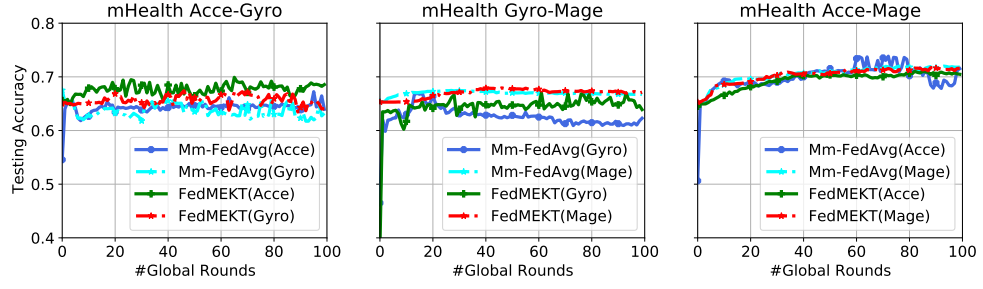


Figure 1: Global Performance of mHealth dataset.

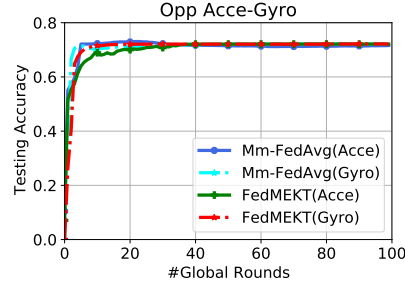


Figure 2: Global Performance of Opp dataset.

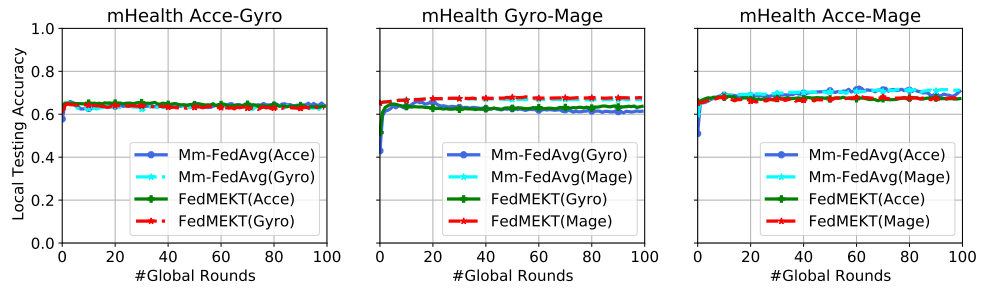


Figure 3: Local Performance of mHealth dataset.

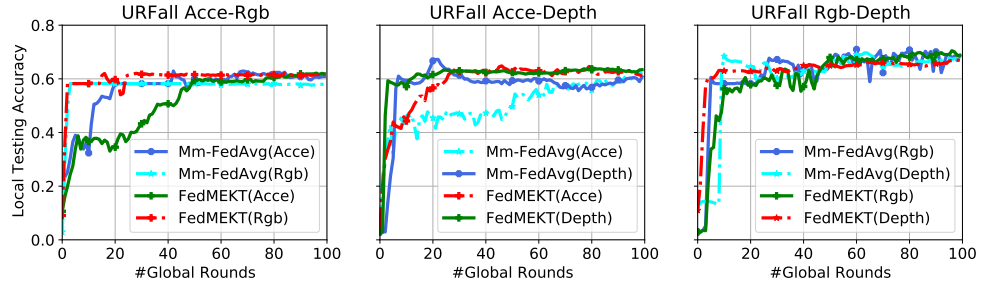


Figure 4: Local Performance of UR Fall Detection dataset.

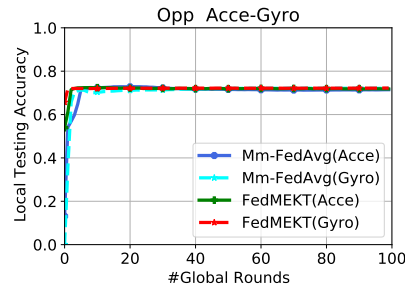


Figure 5: Local Performance of Opp dataset.