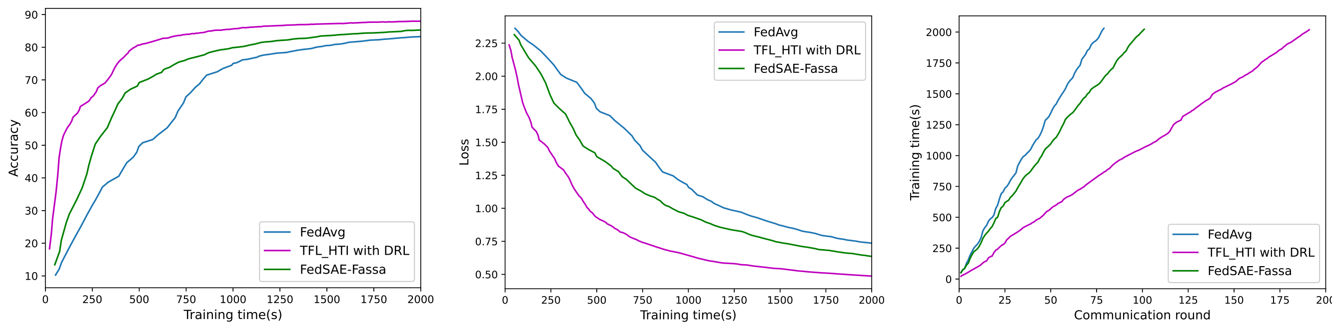
# Supplementary file to TFL_HTI

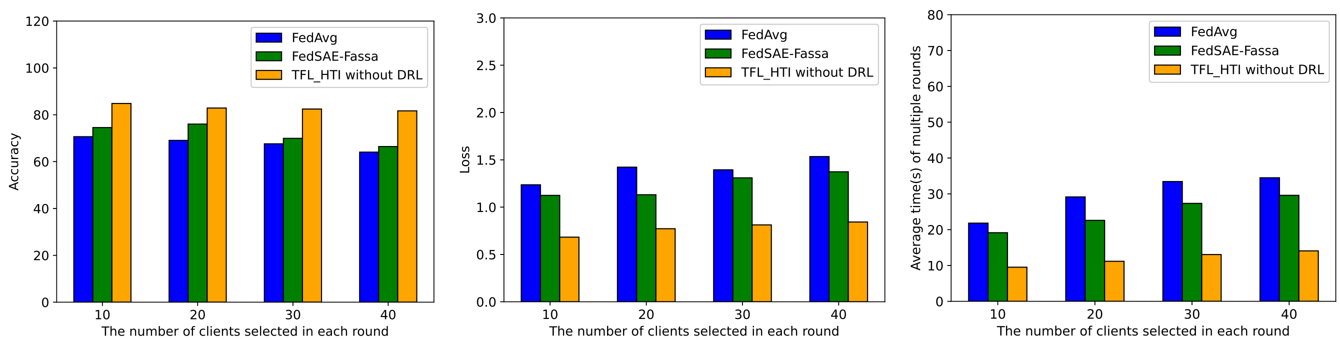
We here present simulation results on a commonly used dataset, *MNIST*. MNIST dataset contains square $28 \times 28 = 784$ pixel gray-scale images of $70,000$ handwritten digits ($60,000$ for training and $10,000$ for testing).




(a) Test accuracy on MNIST (b) Loss on MNIST (c) Average time on MNIST

Fig. 1: Performance on MNIST dataset.

As shown in Fig. 1, we set the training time budget as $2000$s. It is shown that compared with two baseline algorithms, our proposed scheme significantly improves the training accuracy, reduces the training loss and the required training time, which thus confirms its efficiency.




(a) Test accuracy on MNIST (b) Loss on MNIST (c) Average time on MNIST

Fig. 2: Performance impacted by $C^k$ on MNIST dataset.

We also study the performance of our proposed scheme impacted by the number of clients participating in $k$-th round, e.g.$C^k$.

As shown in Fig. 2, we range the number of clients chosen in a communication round from 10 to 40. It shows that the test accuracy and training loss of three algorithms change lightly under MNIST dataset, when increasing the number of clients. This is possible, as the overall training intensity is fixed and clients will be allocated with smaller local iterations when $C^k$ increases. Nonetheless, our proposed algorithm still performs the best. It can also be seen that the average training time of baseline algorithms increases significantly when the number of clients becomes larger. On the contrary, the average training time of our local iteration allocation scheme keeps stable with the increasing number of clients participating in a round, which shows the efficiency and robustness of our strategy.