# iSample: Intelligent Client Sampling in Federated

**HamidReza Imani, Jeff Anderson**
**Advisor: Tarek El-Ghazawi**

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

GW SEAS

## Motivation

The pervasiveness of AI in society has made machine learning (ML) an invaluable tool for mobile and internet-of-things (IoT) devices. However:

- The data available to any one device is limited to train an accurate model.
- Training data movement among devices increases the bandwidth requirements and causes privacy issues.

Federated Learning (FL) allows client devices to share ML models instead of data to learn from one another. However, heterogeneity in device resources still impose limitations on training performance.

## Introduction

In order to improve performance while maintaining good levels of accuracy, we introduce iSample. iSample, an intelligent sampling technique, selects clients by jointly considering known network performance and model quality parameters, allowing the minimization of training time.

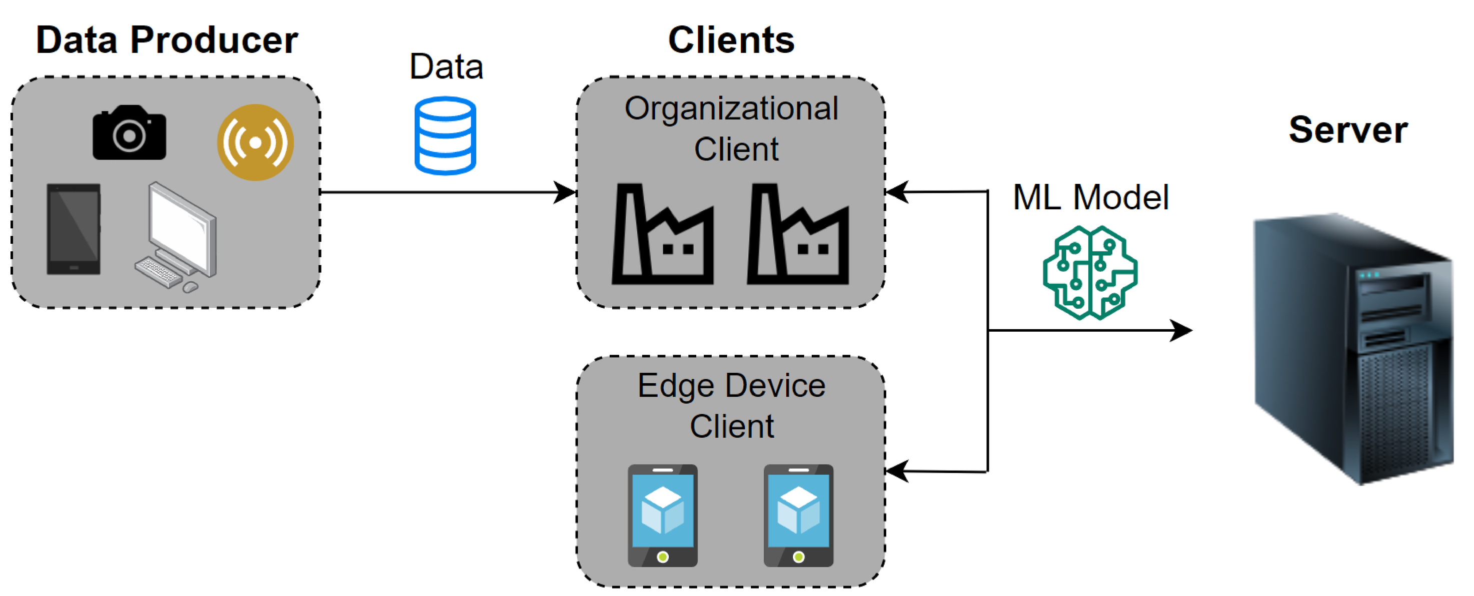


Figure 1. FL aggregates models trained on disparate data sets to derive a unified model used for client update.

## Prior Works

Standard FL [2]: Providing the global model only to a random sub-sample of clients

FedCS [3]: Estimating client resources in a heterogeneous environment and providing the global model only to the ones that can meet a certain deadline

Selective Aggregation [4]: Only including the clients with high quality training data (images) in aggregation

Selecting a subset of clients with significant weight updates [5]: Sampling the clients based on $\|\Delta W\|$.

$$\|\Delta W\| = \sqrt{\sum_{i=1}^{k} \Delta w_i^2}$$

## iSample

- A surveying phase where clients can report their performance parameters
- A grading and a subsequent ranking system to sample the efficient clients (a, b, c and d are tunable coefficients)

$$grade = a.accuracy + b.throughput + c.\|\Delta W\| - d.latency$$

- Increasing the inclusivity by providing the global model to the majority of clients

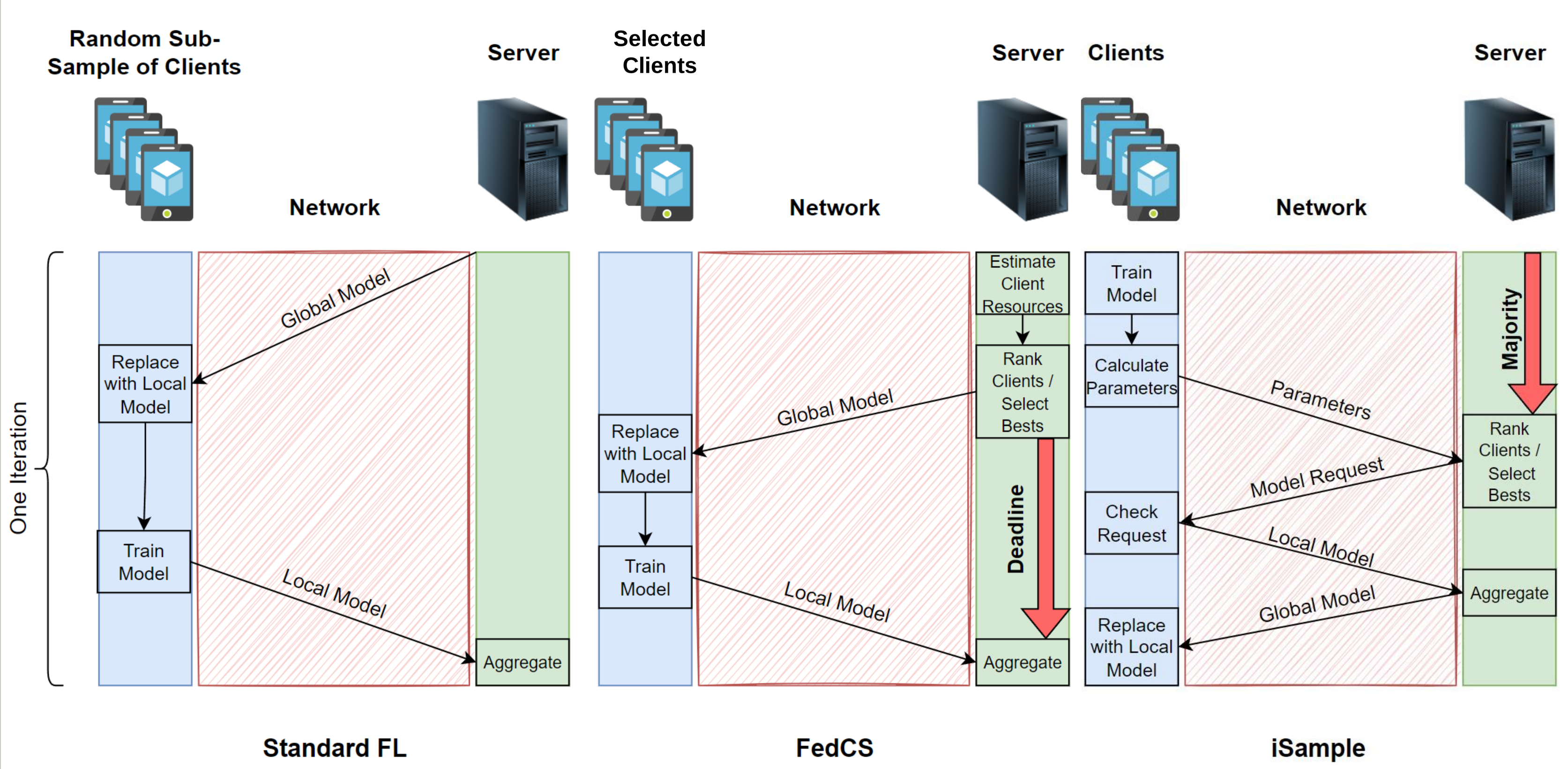


Figure 2. Comparison of iSample protocol with Standard FL and FedCS

- To narrow the search space for the optimal configuration for coefficients c and d are set to 0.1 and 0.01 respectively

## Evaluation and Results

- Server and clients are deployed using AWS EC-2 instances in the cloud
- 80 clients in total residing in four different regions: Oregon, Virginia, London and Frankfurt
- Limited network throughput for each client to emulate heterogeneity
- ML task: image classification using CIFAR-10
- Quarter of the clients are sampled in each iteration

- Tested the protocols on two different ML model:
  - VGG (1060130 parameters → 4.04 MB )
  - CNN (122570 parameters → 0.46 MB)

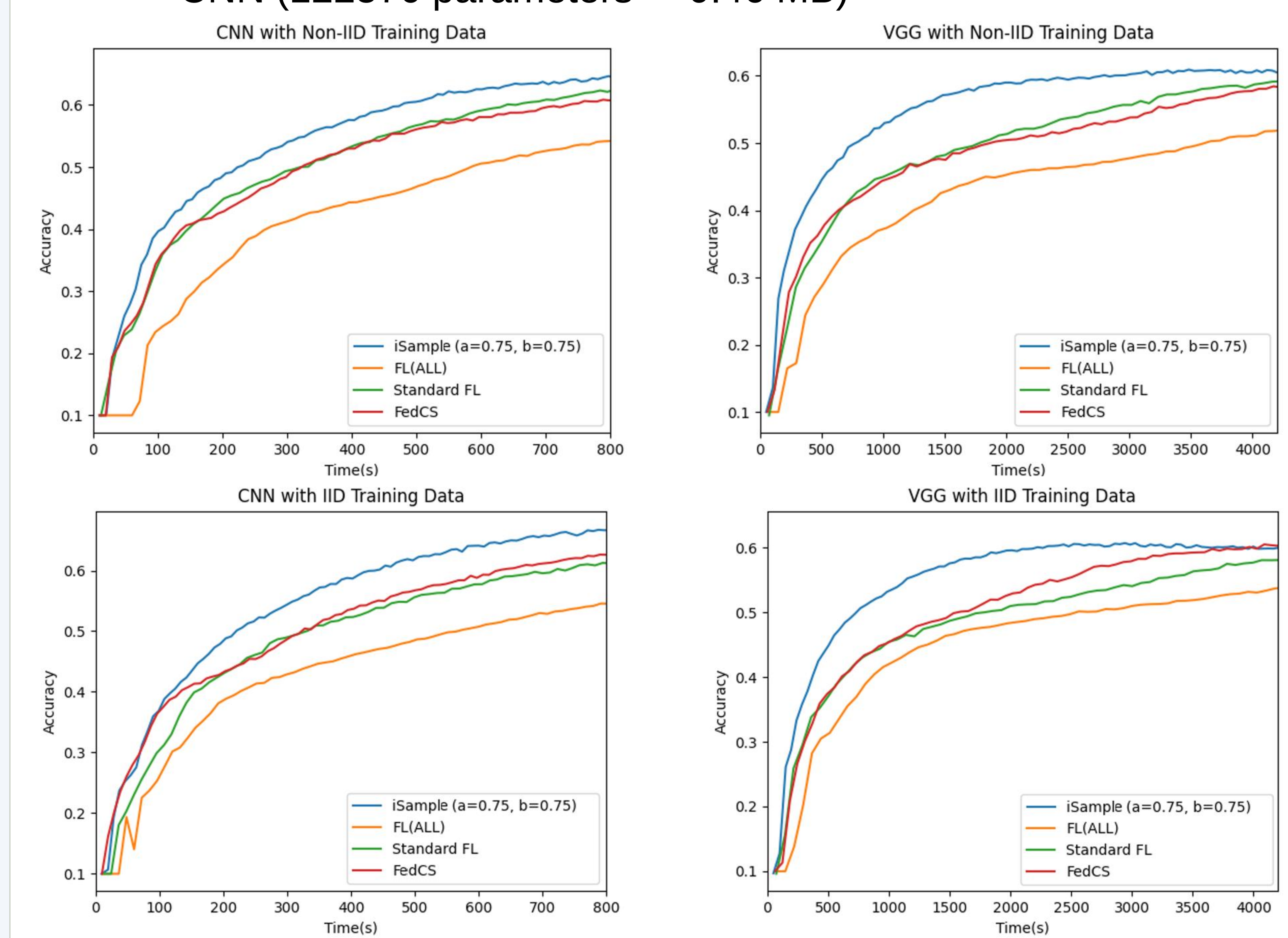


Figure 3. Comparison of one configuration of iSample with FedCS and Standard FL using different ML models for IID and non-IID training data

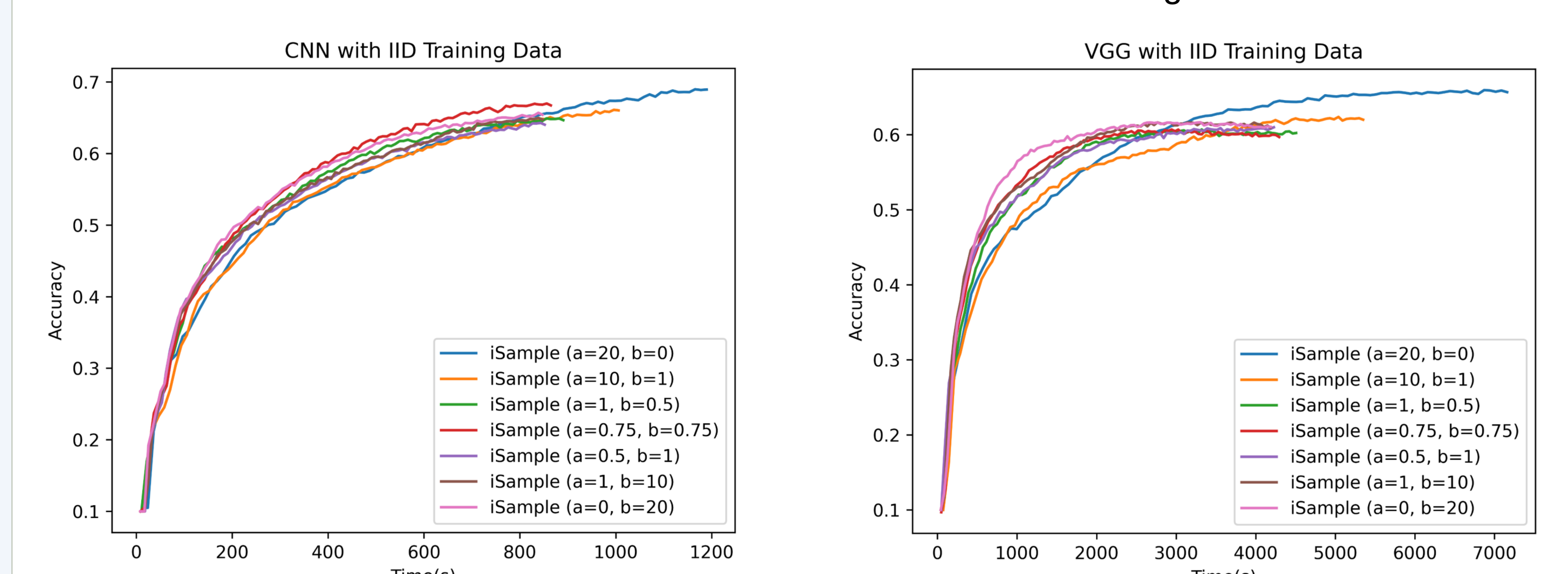


Figure 4. Performance comparison of different coefficient configurations (100 iterations)

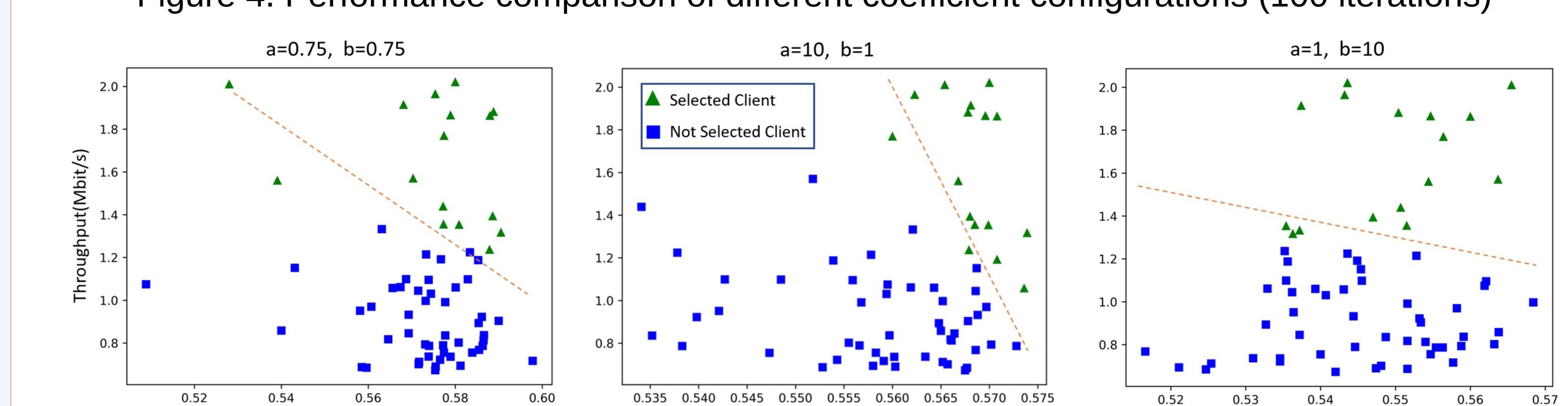


Figure 5. Selected clients for CNN, training with non-IID data at epoch 50 using iSample with different configurations

## References

[1] H. Imani, J. Anderson and T. El-Ghazawi, “iSample: Intelligent Client Sampling in Federated”. 2022 IEEE International Conference on Fog and Edge Computing (ICFEC)
[2] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-efficient learning of deep networks from decentralized data,” Artificial intelligence and statistics.
[3] T. Nishio and R. Yonetani, “Client selection for federated learning with heterogeneous resources in mobile edge,” in ICC 2019-2019 IEEE International Conference on Communications (ICC).
[4] D. Ye, R. Yu, M. Pan, and Z. Han, “Federated learning in vehicular edge computing: A selective model aggregation approach,” IEEE Access, vol. 8, pp. 23 920–23 935, 2020.
[5] M. Ribero and H. Vikalo, “Communication-efficient federated learning via optimal client sampling,” arXiv preprint arXiv:2007.15197, 2020.