COMPUTER SCIENCE AND DATA ANALYTICS

Course: Guided Research I

Report 2

Title: Federated Machine Learning Implementation on Image Classification

Student: Ali Asgarov

Instructors: Prof. Dr. Stephen Kaisler, Assoc.Prof Jamaladdin Hasanov

Problem Description:

This project aims to explore federated learning for image classification while preserving privacy. The goal is to develop a federated learning algorithm tested on image classification tasks and evaluate its performance compared to centralized training approaches.

In FL, each client trains its model decentrally. In other words, the model training process is carried out separately for each client. Only learned model parameters are sent to a trusted center to combine and feed the aggregated main model. Then the trusted center sent back the aggregated main model back to these clients, and this process is circulated.

In this context, I am preparing an implementation with IID (independent and identically distributed) data to show how the parameters of tens of different models that are running on different nodes can be combined with the Federated Learning method and whether this model will give a reasonable result.

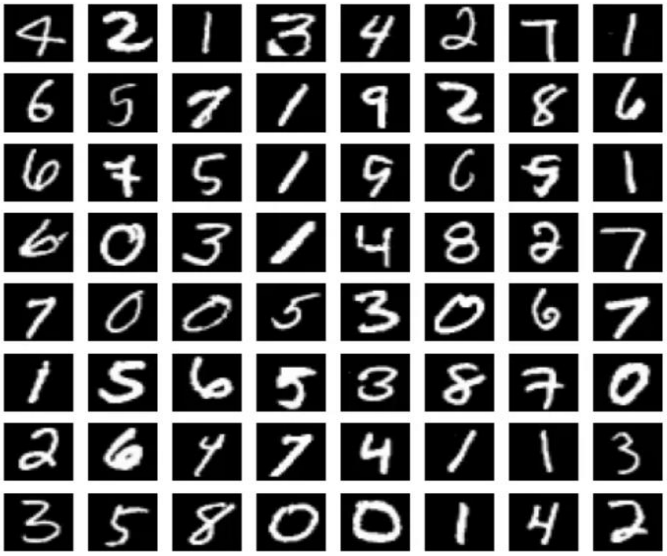
Strategy Definition:

Qualitative analysis is important in my project as I experiment with the implementation of the novel federated learning concept. Given the multiple frameworks available, such as PyTorch and TensorFlow API, qualitative analysis will allow me to assess their efficiency and suitability for my project after data preprocessing is done. It helps me understand the practical aspects, such as ease of use, performance, flexibility, and scalability of each framework. Additionally, qualitative analysis guides me in identifying challenges, limitations, and areas for improvement specific to each framework. By combining qualitative analysis with the experimental implementation of federated learning, I will gain valuable insights to select the most efficient framework for implementation of federated learning concept on my image classification tasks.

Quantitative evaluation is also integral to my project. I am analyzing image data, including accuracy, precision, recall, which will measure performance and efficiency of federated learning for image classification. By comparing it to centralized training approaches, I can objectively assess the impact of federated learning.

In conclusion, by incorporating both qualitative and quantitative analyses, my project aims to gain a comprehensive understanding of the implementation of federated learning for image classification. Through qualitative analysis, I can assess the practical aspects and challenges associated with different federated learning implementation methods, while quantitative evaluation allows for objective performance assessment and comparison with centralized training approaches.

Dataset Selection: This implementation will be carried out on the MNIST Data set. The MNIST data set contains 28 * 28 pixel grayscale images of numbers from 0 to 9.



Handwritten Digits from the MNIST dataset (Image by Author*)

URL http://deeplearning.net/data/mnist/"

Filename mnist.pkl.gz

Dataset Preparation:

The MNIST data set does not contain each label equally. Therefore, to fulfill the IID requirement, the dataset will be grouped, shuffled, and then distributed so that each node contains an equal number of each label. Also will be scaling the image values.

Model Architecture:

Here I will be designing a suitable deep learning model architecture for image classification and determining the hyperparameters and optimization algorithm to be used. A simple 2-layer model will be created for the classification process.

Federated Learning Setup:

In this step I will be doing framework selection for the implementation of the Federated Learning on the image classification. As this is novel approach there are stills points needs to be improved behind the Federated Learning implementation frameworks. I will be evaluating PyTorch and TensorFlow API for implementation and will focus the most efficient one.

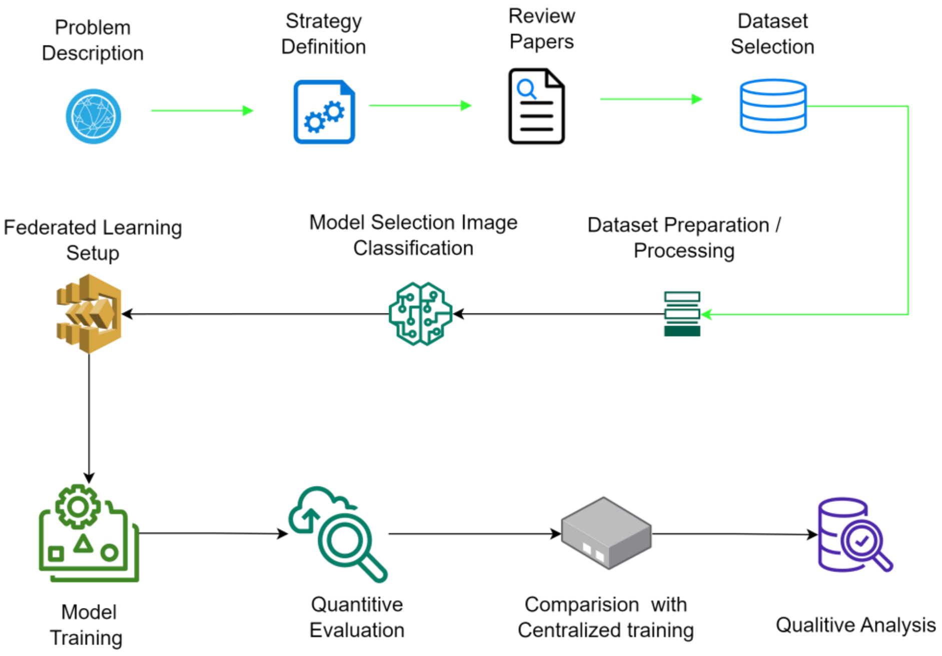
Initial Model Training:

Here I will implement federated learning algorithms using the selected framework, train models on client nodes and aggregate parameters. I will distribute the initial model to the client nodes. Each client trains its model locally on its respective data using the IID approach. Will monitor and record the training progress and performance metrics for each client model.

Comparison with Centralized Training:

Here I will evaluate and compare performance against centralized training approaches based on the precision, recall, accuracy and other classification metrics.

Here is the Flow Diagram:


Problem Description
Strategy Definition
Review Papers
Dataset Selection
Federated Learning Setup
Model Selection Image Classification
Dataset Preparation / Processing
Model Training
Quantitive Evaluation
Comparision with Centralized training
Qualitive Analysis