The University of Texas at Austin
Chandra Department of Electrical and Computer Engineering
Cockrell School of Engineering

Spring 2025

ECE 361E: Machine Learning and Data Analytics for Edge AI
Project Description (Milestones 1, 2, and 3)
Released: Mar 11, 2025

Work in groups of two students. Only one submission per group is required.

Big picture

In this project, you will experiment with federated learning (FL) techniques and optimize an FL system with real data. You will also participate in an FL competition that will award bonus points to the best projects.

Basic setup

- Application: Image classification on MNIST Dataset
- Model: Conv5Small
- Training Framework: PyTorch
- Training platform: Two (one RaspberryPi 3B+ and one Odroid MC1) devices per team

In milestone 1 (**M1**), the primary objective is to become familiar with the FL framework. You will run multiple FL experiments with different configurations and evaluate their performance in terms of convergence time and energy consumption. In milestone 2 (**M2**), you will explore more advanced FL approaches (e.g., FedMAX, FedProx) to improve upon **M1**. Finally, in milestone 3 (**M3**) you will optimize the framework based on your own design choices, as you will be competing against the other teams to see which FL configuration performs best. Both milestones **M1** and **M2** have a *mandatory component* with multiple tasks similar to the homeworks (e.g., **M1.1, M1.2,..., M2.1, M2.2**,...), and an *exploration component* (**M1.Exploration** and **M2.Exploration)** where you need to try different configuration and optimization choices to improve your design. This exploration part is the basis for the M3 competition.

Note: Throughout this project you are restricted to use up to 30 communication rounds per experiment and need to start training from scratch (i.e., no pre-trained models are allowed).

M1 [35 points] FL on Edge Devices - DUE: Monday, Apr 7, 12 noon (CST)

In **M1**, you need to run the required experiments and measure the global test accuracy, total wall clock time, convergence time[1] (in number of communication rounds), average time per communication round, average energy consumption per communication round, and the average amount of communicated data per communication round for each device with the Conv5Small model.

M1.1. Run ***config_1.bash*** as given to you to generate the configurations for **experiments 1 and 2**. These experiments will run both IID (experiment 1) and non-IID (experiment 2) scenarios using FedAvg with Conv5Small, and training for one local epoch per communication round. Run ***generate_figs.py*** with ***--time*** and ***--accuracy*** arguments to plot the global test accuracy [%] vs. number of communication rounds, global test loss vs. communication rounds, and complete ***Table 1*** with the obtained results.

[1] **Note:** The convergence time is the number of communication rounds needed for the model to reach a certain accuracy threshold (in this case, 90%).

Table 1

Aggregation method	**FedAvg**	
Data heterogeneity	**IID**	**Non-IID**
Convergence time [#rounds]		
Avg. time per communication round [s]		
Total wall clock time [s]		
Global test accuracy [%]		

M1.2. Run ***generate_figs.py*** with ***--power***, ***--energy*** and ***--communication*** arguments to plot the average power consumption [W] per device, average energy consumption [J] per device, and average amount of communicated data [MB] per communication round per device. Complete ***Table 2***.

Note: Total average energy consumption per communication round [J] = RPI average energy consumption per round [J] + MC1 average energy consumption per round [J]

Table 2

Aggregation method	**FedAvg**	
Data heterogeneity	**IID**	**Non-IID**
RPI avg. energy consumption per round [J]		
MC1 avg. energy consumption per round [J]		
Total avg. energy consumption per communication round [J]		
Avg. amount of communicated data per communication round [MB]		

M1.3. Using ***Table 1***, ***Table 2*** and the plots from **M1.1** and **M1.2** respectively, compare and contrast the impact of IID and the non-IID scenarios on FL system performance. What do you observe?

M1.Exploration

For this milestone, you need to explore the effects of changing the learning rate, model architecture, and the number of local epochs for the non-IID scenario only.

Create a **config_2.bash** file by copying the **config_1.bash** and define ***at least three distinct experiments*** where you ***modify at least two of the three configuration parameters*** (i.e., you cannot just try three different learning rates for your three experiments).

For each experiment, run ***generate_figs.py*** with ***--time***, ***--accuracy***, ***--power***, ***--energy*** and ***--communication*** arguments to obtain relevant plots and report the metrics in ***Table 1*** and ***Table 2***. **In your M1 presentation include only the top 3 most significant experiments (i.e., best evaluated metrics) you ran in your exploration.**

What to explore: learning rates, model architectures, and number of local epochs.

Of note: Any model architecture you try cannot be larger than 1M parameters (strictly enforced).

Guidelines for presentation and video recording[1]:

- The **M1** project presentation consists of **up to 5 slides** excluding the cover page **(this limit is strictly enforced).** Your presentation must provide the following information:
    - **Cover page:** Project title and Group member(s)
    - **Approach for exploration (1 slide):** Explain the main idea/steps behind your exploration approach. This can include all the exploration you did (not only the top 3 most significant experiments in terms of results).
    - **Experimental results (2-3 slides):** Present the metrics evaluated from ***Table 1*** and ***Table 2*** and discuss your main insights from the top 3 most significant results. If you have more interesting experiments, then present only the insights you learned from them.
    - **Conclusion and plans to improve/explore further (1 slide)**
- **You must** pre-record a *backup* **video presentation** in one of the following formats: AVI, MOV, mp4, MPEG, and WMV. We suggest using Zoom for recording videos, but you can also use this software. Each video can be a maximum of **three minutes** long **(this limit is strictly enforced)**. Record the demo in full screen mode and show the presenter(s) in a small window on the screen.

Of note: All project presentations will be live, in-class. Only in the rare event a team cannot make it we will play the backup video in class.

Submission:

- PDF document with results and discussion from **M1.1, M1.2** and **M1.3**
- PDF version of your slides (PDF file only, **no PowerPoint**)
- Your framework code + *readme.txt* file explaining all your changes to the framework
- Video presentation

All files should be in a single zip file named ***Team<Team#>.zip***

M2 [35 points] FL Optimization - DUE: Monday, Apr 14, 12 noon (CST)

In **M2**, you will explore various optimization techniques to improve the efficiency of federated learning.

M2.1. Create a **config_3.bash** file by copying the **config_1.bash** and create 4 new experiment configurations as follows. Using one local epoch, run FedMAX with **beta=10** for both IID and non-IID scenarios and FedProx with **mu=1** for both IID and non-IID scenarios (i.e., 4 experiments in total). For each experiment, run ***generate_figs.py*** with ***--time***, ***--accuracy***, ***--power***, ***--energy*** and ***--communication*** arguments to draw the global test accuracy [%] vs. #communication rounds and the global test loss vs. #communication rounds plots, and obtain the metrics needed to complete ***Table 3***.

M2.2. How do FedMAX and FedProx compare against FedAvg? Which approach is better? Explain.

M2.3. FedProx allows for partial computation from each client. In the current implementation, do we account for hardware heterogeneity using FedProx? If the answer is yes, explain how that works. If the answer is no, explain what would be needed to make it work.

[1] The detailed schedule for presentations (Tuesday vs Thursday) will be available on Canvas in a timely manner.

Table 3

Aggregation method	**FedMAX**		**FedProx**	
Data heterogeneity	**IID**	**Non-IID**	**IID**	**Non-IID**
Convergence time [#rounds]				
Avg. time per communication round [s]				
Total wall clock time [s]				
Global test accuracy [%]				
RPI avg. energy consumption per round [J]				
MC1 avg. energy consumption per round [J]				
Total avg. energy consumption per communication round [J]				
Avg. amount of communicated data per communication round [MB]				

M2.Exploration

For this milestone, you need to explore the effects of changing the hyperparameters beta (for FedMAX), mu (for FedProx), the learning rate for each device and the number of local epochs for each device (each device can run a different number of local epochs and different learning rate), for the non-IID scenario only.

Note: The number of local epochs can be different for each device, e.g., two local epochs for the RPi and one local epoch for the MC1. Similar for the learning rate, e.g., 0.1 for RPi and 0.01 for MC1.

Create a **config_4.bash** file by copying the **config_1.bash** and define ***at least two distinct experiments*** where you ***modify at least two of the four configuration parameters*** (i.e., you cannot just try two different values of beta for FedMAX as your two experiments).

For each experiment, run ***generate_figs.py*** with ***--time***, ***--accuracy***, ***--power***, ***--energy*** and ***--communication*** arguments to obtain relevant plots and report the metrics in ***Table 1*** and ***Table 2***. **In your M2 presentation include only the top 2 most significant experiments (i.e., best evaluated metrics) you ran in your exploration.**

What to explore: beta (FedMAX), mu (FedProx), different number of local epochs for each device.

Submission (Follow the same guidelines as in **M1**): Record a backup video where you present your experimental results (you must report the metrics from **Table 3** for each experiment) and observations; then upload the video to Canvas. Include the **PDF** version of your slides for **M2** (**no PowerPoint**) and your framework code + readme.txt file explaining all your changes to the framework in a single zip file named ***Team<Team#>.zip***

M3 [30 points] House of the ~~Dragons~~ Federations
DUE: Monday, Apr 21, 12 noon (CST)

In Milestone 3 (**M3**) you will apply everything you learned in this class to compete in the *"House of Federations"* competition. Each team must choose to optimize the FL system for either **Team Black** or **Team Green** and devise an optimized strategy to implement within the federated learning framework. Use the findings reported in **M1** and **M2 to guide your choices**. Feel free to experiment with any combination of the configuration parameters you explored in **M1** and **M2**.
All experiments will have 3 runs and the results will be your average over the 3 runs, each run having a different seed. Record your results in ***Table 4***.

Table 4

Aggregation method	**<fill_in>**
Data heterogeneity	**Non-IID**
Convergence time [#rounds]	
Avg. time per communication round [s]	
Total wall clock time [s]	
Total wall clock time to reach 90.00% [s]	
Global test accuracy [%]	
RPI avg. energy consumption per round [J]	
MC1 avg. energy consumption per round [J]	
Total avg. energy consumption per communication round [J]	
Avg. amount of communicated data per communication round [MB]	

In the project competition, you can only belong to one team:

- **Team Black:** Optimize for the *fastest wall clock time* [s] to reach the global test accuracy threshold of 90.00% in the non-IID scenario.
- **Team Green:** Optimize for the *lowest total average energy consumption per communication round* [J], while reaching a global test accuracy of 90.00% in the non-IID scenario, within 30 communication rounds.

Note: **For both teams**, **once the global test accuracy threshold of 90.00% is achieved, all subsequent communication rounds must maintain or exceed this threshold of global test accuracy**.

For example, for **Team Black**, if your results are in the table below, then the total wall clock time to reach 90.00% will be considered from communication round 1 until communication round 29 (i.e., not from communication round 1 until communication round 27, since the global test accuracy drops under the 90.00% threshold in communication round 28).

Communication round	27	28	29	30
Global test accuracy [%]	90	89	90	91

We will decide the winners of this competition based on the performance of your *final (i.e., most optimized) design.* To make this competition fair, we will evaluate each and every project and thoroughly assess its quality.

The best projects, one selected for **Team Black** and one for **Team Green**, will receive 4 bonus points (added directly to the final score in this class). Each winning team can split the 4 bonus points (only integers, e.g., 2+2, 1+3, 0+4) however they want between the members of the winning team.

For **M3**, you are required to do the following:

- Fill the data from the template slide specifically designed for M3 (**M3_Presentation-v1.pptx**). You need to mention: **Team Black** or **Team Green**, your most notable result (using ***Table 4***), and the most important lesson(s) learned during this project.
- Record a video **up to 90 seconds (1 minute 30 seconds)** where you present your slide and upload the video to Canvas. This presentation should be like a pitch for your project highlighting your best (final) results, as well as lessons learned along the way.

Submission: You must include **the PowerPoint slide we provided updated with your information** for **M3** and your code in a single zip file named ***Team<team_number>.zip***. Make sure to include the framework code with a readme that describes what you changed so we can reproduce your results.

Good luck!