In this project, we leveraged various dataset distillation techniques, focusing more on Attention Matching and Distribution Matching, to create these compact synthetic datasets while maintaining crucial characteristics of source data. Dataset distillation plays an important role when there are resource constraints, be it memory or computational, in order to generate synthetic yet representative small datasets. These synthetic data can be used for training deep learning models with only a minimal loss in the accuracy of the model, which is useful in applications like privacy-preserving machine learning and edge computing.

The experimental investigation used two key datasets: MNIST is the standard dataset used for handwritten digit identification, and MHIST is a medical imaging dataset for histopathological images.

Attention Matching was used to align the attention maps of synthetic and real data with the goal of preserving the critical features in classification tasks. Similarly, training models from synthetic sets and directly comparing their performance with models trained on the original sets was used as another technique. The main metrics used are: classification accuracy, computational efficiency measured in FLOPs, and cross-architecture generalization.

With Attention Matching, in Task 1, we achieved a modest test set classification accuracy of about 50-65%, depending on the dataset, with class features clear in the synthetic data but with some noise. However, Task 2's Distribution Matching technique, which optimizes dataset condensation directly using MMD maximization, significantly outperformed Attention Matching. Models trained on synthetic datasets generated through Distribution Matching reached as high as 97.3% accuracy on MNIST and reached close to real-data performance. Although that may be true, Distribution Matching is more accurate, converges better, generalizes well, and generates synthetic images that are visually appealing; usually, it takes longer when considering training iterations. The project illustrates how dataset distillation efficiently trains models under resource constraints, while Distribution Matching proves to be a formidable approach when high accuracy is in view.