CSC 591 - High Performance Machine Learning
Unit Project 3

Team Member Names: Harshil Shah, Md Jakaria
Team Member Unity IDs: hshah6, mjakari

1. What HPO tuners you have applied to which DNN model.

Answer. To the base model using MNIST dataset, I have tuned using the following tuners:
- TPE Tuner
- Evolution Tuner
- Metis Tuner

2. What configurations did you try on each HPO tuner, and what machine(s) you have used, including the CPU and GPU models of the machine(s).

Answer. I have run all of the experiments on my local pc, with the following hardware:
- CPU: AMD Ryzen 7 5800H with Radeon Graphics 3.20 GHz
- GPU: NVIDIA GeForce RTX 3060 Laptop GPU

Note: I have run each experiment for 10 trials, and with 10 epochs of training for each trial.

I have tried the following configurations with the 3 tuning methods:

- **TPE Tuner:**
    - *Configuration 1*:
        - {
            'optimize_mode': 'maximize',
            'tpe_args': {
                'constant_liar_type': 'mean',
            }
          }
    - *Configuration 2*:
        - {
            'optimize_mode': 'maximize',
            'tpe_args': {
                'constant_liar_type': 'best',
            }
          }

- **Evolution Tuner:**
    - *Configuration 1:*
        - {
            'optimize_mode': 'maximize',
            'population_size': 10,
          }
    - *Configuration 2:*

- ■ {
     'optimize_mode': 'maximize',
     'population_size': 6,
  }

- ● **Metis Tuner:**
  - ○ *Configuration 1:*
    - ■ {
         'optimize_mode': 'maximize',
         'exploration_probability': 0.2,
      }
  - ○ *Configuration 2:*
    - ■ {
         'optimize_mode': 'maximize',
         'exploration_probability': 0.8,
      }

3. What results you have obtained, including the hyperparameters the tuner finds at the end and the impact (e.g., the accuracy and speed of the resulting model), and also your explanations of the reasons for the differences in the results from the different methods and configurations.

Answer. Here are the experiment results for each configuration of each of the 3 tuners used.

- ● **TPE Tuner**
  - ○ *Hyperparameters modified:*
    - ■ Learning rate
    - ■ Momentum
  - ○ *Configuration 1*:
    - ■ {
         'optimize_mode': 'maximize',
         'tpe_args': {
            'constant_liar_type': 'mean',
         }
      }
  - ○ Max Accuracy: 99.14%
  - ○ Inference Speed on testing set: 4.76s
  - ○ Tuning speed (total time taken to run 10 trials): 59m42s
  - ○ Hyperparameters with max accuracy:
    - ■ Learning rate: 0.0479
    - ■ Momentum: 0.7033
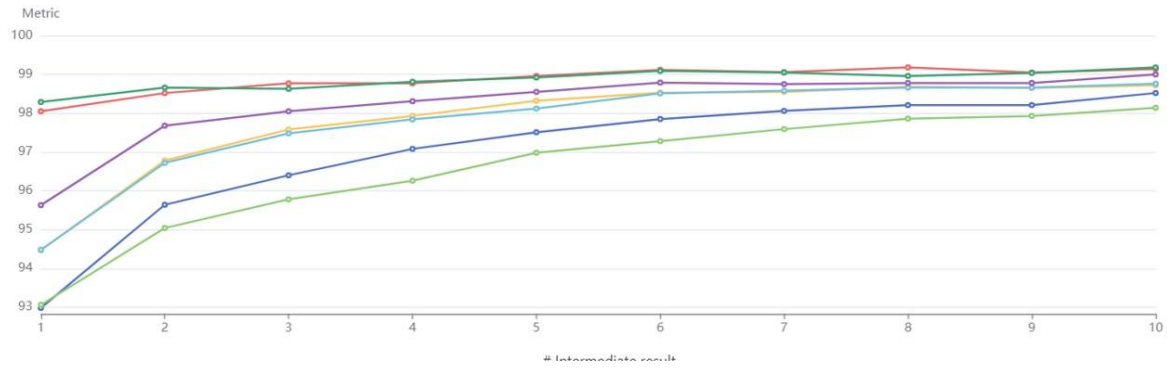
- **TPE Tuner**
  - *Hyperparameters modified:*
    - Learning rate
    - Momentum
  - *Configuration 2*:
    - {
        'optimize_mode': 'maximize',
        'tpe_args': {
            'constant_liar_type': best,
        }
      }
  - Max Accuracy: 99.15%
  - Inference Speed on testing set: 5.28s
  - Tuning speed (total time taken to run 10 trials): 1h4m45s
  - Hyperparameters with max accuracy:
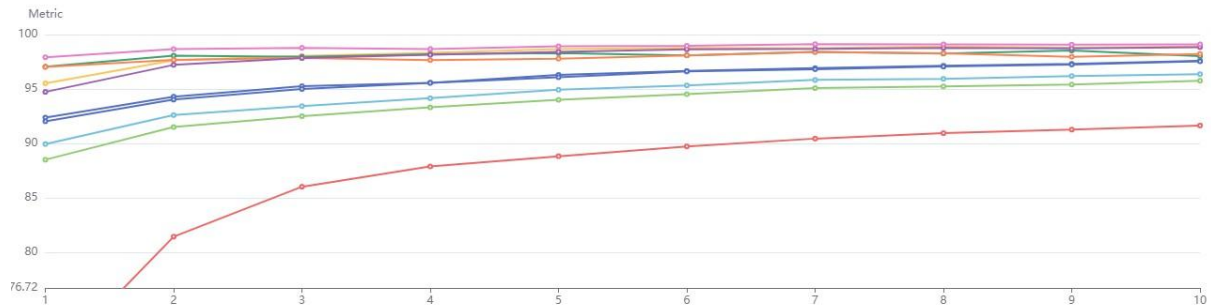   - Learning rate: 0.0499
    - Momentum: 0.6682
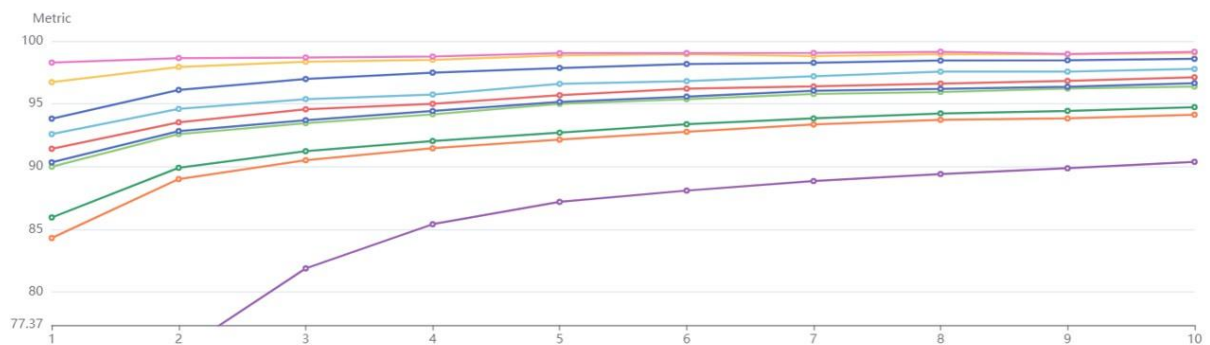
- **Evolution Tuner:**
  - *Hyperparameters modified:*
    - Learning rate
    - Momentum
  - *Configuration 1:*
    - {
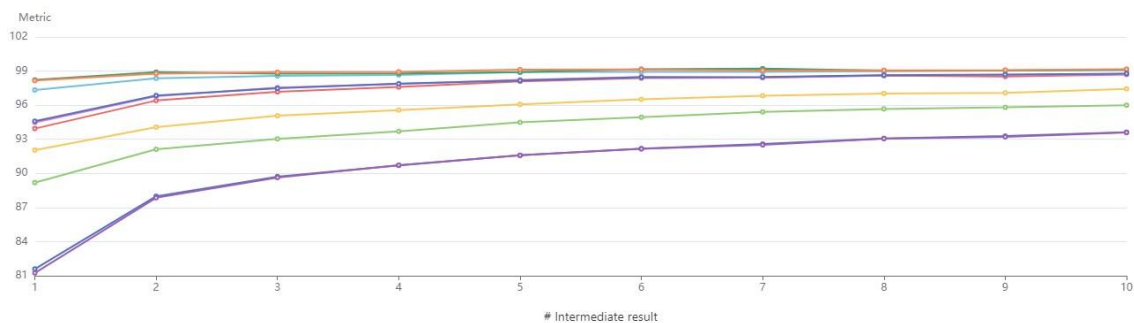   'optimize_mode': 'maximize',

<pre>   'population_size': 10,
                }
</pre>
- ○ Max Accuracy: 99.10%
- ○ Inference Speed on testing set: 5.36s
- ○ Tuning speed (total time taken to run 10 trials): 53m48s
- ○ Hyperparameters with max accuracy:
    - ■ Learning rate: 0.0090
    - ■ Momentum: 0.8829

- ● **Evolution Tuner:**
    - ○ *Hyperparameters modified:*
        - ■ Learning rate
        - ■ Momentum
    - ○ *Configuration 2:*
        - ■ {
            <pre>   'optimize_mode': 'maximize',
                'population_size': 6,
            }</pre>
    - ○ Max Accuracy: 99.17%
    - ○ Inference Speed on testing set: 5.27s
    - ○ Tuning speed (total time taken to run 10 trials): 45m49s
    - ○ Hyperparameters with max accuracy:
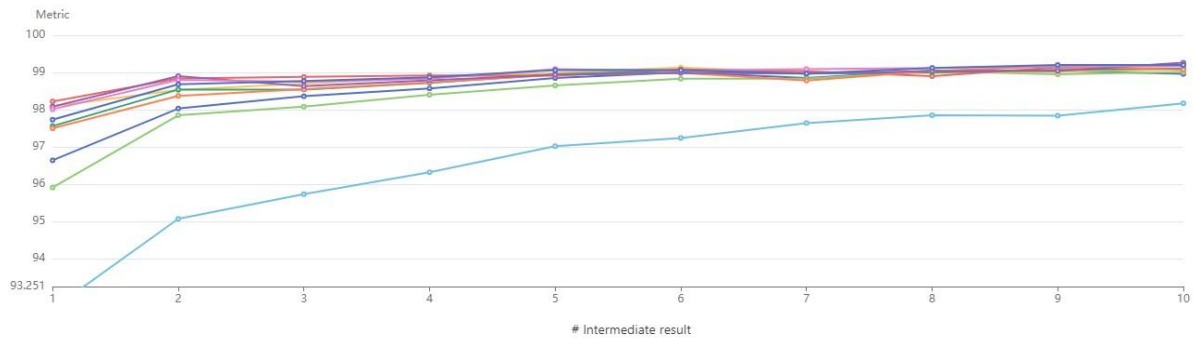   - ■ Learning rate: 0.0846
        - ■ Momentum: 0.4394

- ● **Metis Tuner:**
    - ○ *Hyperparameters modified:*
        - ■ Learning rate
   - ■ Momentum
    - ○ *Configuration 1:*

- {
    - 'optimize_mode': 'maximize',
    - **'exploration_probability': 0.2,**
    - }
  - ○ Max Accuracy: 99.26%
  - ○ Inference Speed on testing set: 5.28s
  - ○ Tuning speed (total time taken to run 10 trials): 44m1s
  - ○ Hyperparameters with max accuracy:
    - ■ Learning rate: 0.0607
    - ■ Momentum: 0.5825

- ● **Metis Tuner:**
  - ○ *Hyperparameters modified:*
    - ■ Learning rate
    - ■ Momentum
  - ○ *Configuration 2:*
    - ■ {
      - 'optimize_mode': 'maximize',
      - **'exploration_probability': 0.8,**
      - }
  - ○ Max Accuracy: 99.13%
  - ○ Inference Speed on testing set: 5.17s
  - ○ Tuning speed (total time taken to run 10 trials): 43m34s
  - ○ Hyperparameters with max accuracy:
    - ■ Learning rate: 0.0641
    - ■ Momentum: 0.0887
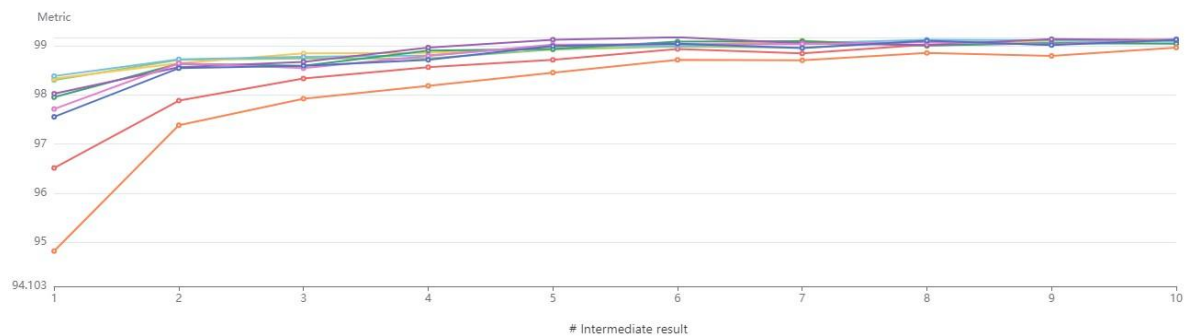
Reasons for the differences in results:
- For the most part, the time taken to run on the test set during inference is approximately the same for most of the tuners. The reason for this is that the speedups are during training time, and not during the inference time.
- There are, however, differences in the tuning time, that is, the time to run all trials on all of the models.
- We see that most models reach 99+% test accuracy by the end of tuning. This is because, I have kept the search space relatively small, and run 10 epochs on a dataset such as MNIST, which is a lot of good training.
- In the case of the differences due to varying configurations, we see little to no difference in the results. The reason for this, again, might be the fact that the hyperparameters tuned are learning rate and momentum, both in a relatively smaller search space, so all of the tuners are able to reach close to optimal values for those hyperparameters.

4. What lessons you have learned through the project.

A. I have learned the following about hyperparameter tuning through this project:
- I've learned what a hyperparameteer means. I understand that it represents a variable involved in the training process of a neural network.
- I understand that with a good understanding of the training process, we can choose better values for these variables.
- This is what the process of hyperparameter optimization means. In simple terms, it attempts to find the best values for hyperparameters to maximize the accuracy, or minimize the loss. In this case, we maximize the accuracy on the testing set.
- This could be more than just one variable, and we need to consider the interactions of these hyperparameters and find good values for their combinations, rather than considering each individual one in a vacuum.
- In this project, we have chosen TPE, Evolution, and Metis tuners from NNI.
- TPE, or Tree-structured Parzen Estimator is a lightweight tuner that only finds best values for individual hyperparameters (no extra benefit when using combinations), and models the conditional probability of the hyperparameter against the evaluation result.
- The Evolution tuner starts with a population generated from a random seed, and with each generation or next set of hyperparameters, it applies some mutations or changes to aim for better results.
- The Metis tuner offers multiple benefits. First, it lets you determine the probability to exploit/explore.
- Secondly, it gives you not only the intermediate results, but also possible recommendations for subsequent trials, and indications of possibly noisy training data, which can be put to use to make the experiments even more productive.
- From the tuners here, all of them got 99+% accuracies, but the Metis tuner with a 20% exploration probability achieved the top accuracy of 99.26%.

5. A link to your GitHub repository:

A. https://github.com/Harshil-Shah99/CSC591-Unit-Project-3
****