Faculty of Computers and Artificial Intelligence

Project Summary:

With access to the dataset provided in the next section,

- Download and analyze the dataset.
- Start working through the steps outlined in the requirements section.

Dataset:

Handwritten Alphabets (hyperlink)

The dataset contains greyscale images for 26 (A-Z) handwritten alphabets of size 28*28 pixels. Your goal is to classify them.

Requirements:

- **Data exploration and preparation:**

  - Identify the number of unique classes and show their distribution.
  - Normalize each image.
  - Reshape the flattened vectors to reconstruct and display the corresponding images while testing the models.

- **Experiments and results:**

  - Split the data into training and testing datasets
  - *First experiment (You can use scikit-learn):*
    - Train 2 SVM models with linear and nonlinear kernels.
    - Test the models and provide the confusion matrix and the average f-1 scores for the testing dataset.
  - Split the training dataset into training and validation datasets.
  - *Second experiment (Build from scratch):*
    - Implement logistic regression for one-versus-all multi-class classification.
    - Train the model and plot the error and accuracy curves for the training and validation data.
    - Test the model and provide the confusion matrix and the average f-1 scores for the testing dataset.

- o *Third experiment (You can use TensorFlow):*
    - ▪ Design 2 Neural Networks (with different number of hidden layers, neurons, activations, etc.)
    - ▪ Train each one of these models and plot the error and accuracy curves for the training data and validation datasets.
    - ▪ Save the best model in a separated file, then reload it.
    - ▪ Test the best model and provide the confusion matrix and the average f-1 scores for the testing data.
    - ▪ Test the best model with images representing the alphabetical letters for the names of each member of your team.
- o Compare the results of the models and suggest the best model.

Deliverables:

You are required to submit ONE zip file containing the following:
- Your code (.py) file. If you have a (.ipynb) file, you have to save/download it as (.py) before submitting.
- A report (.pdf) containing the team members' names and IDs, the dataset you chose and the code with screenshots of the output of each part. If you have a (.ipynb) file, you can just convert it to pdf.
- A text file containing the team members' names, IDs, and groups, and a link to a video uploaded on your drive.
  In the video, each team member should speak stating his/her ID at the start, and take one of the main parts in the code (exploration, preparation, model(s), results) and explain it. The overall duration of the video should not exceed 10 minutes.

The zip file must follow this naming convention: **Group_ID1_ID2_ID3_ID4_ID5_ID6**

Instructions:

1. The minimum number of students in a team is 3 and the maximum is 6.
2. No late submission is allowed.
3. Team members must be from the same lab (or have the same TA).
4. All team members must understand all parts of the code.
5. Stick to uploading ONLY the required files following the naming convention.
6. Cheating students will take **ZERO** and no excuses will be accepted.
7. Grades are given based on the performance of the team and satisfying the requirements.

Grading Criteria:

Data Preparation	2
SVM	2
Logistic Regression	3
Neural Networks	3
Results in a well-written report	1
Total = 11 points	

If the video (link) is not included in your deliverables, you will lose 5 points.