

## WIX3001 Soft Computing

### Assignment 2: Python Programming

1. Group Assignment: 2 – 5 members per group.
2. For this assignment, each group has to collect their own dataset similar to MNIST. Use this as a guideline: <https://www.kaggle.com/code/mikalaichaly/diy-mnist-dataset>
3. Instead of labeling the characters as digits or alphabets, the characters are labeled with the name of the person writing them. The objective of this assignment is to **identify the user from their handwriting**.  
Example:  
First set of characters: 0 1 2 3 4 5 6 7 8 9 A B C D ... -> Labeled as Member1  
Second set of characters: 0 1 2 3 4 5 6 7 8 9 A B C D ... -> Labeled as Member2  
...
4. Each member in the group must contribute 50+ handwritten characters, either digit, alphabet, or both. Ideally each member should write each character multiple times.
5. When completed, the dataset should have X class labels, where X is the number of members in your group. The dataset should look similar to the MNIST dataset.
6. **Use soft computing methods to design a method to identify the user from their handwriting**. Examples of methods:
  - Fuzzy logic.
  - Neural networks.
  - Genetic algorithms to optimize (fuzzy logic / neural network) parameters.
  - Fuzzy extreme learning machine.
  - Others...
7. The methods should be programmed using Python, and using Jupyter Notebooks to show your steps.
8. Write a report. It should contain the following:
  - Details of your group members.
  - Details of your dataset. How is it collected? What preprocessing methods are used? Show some examples of handwritten characters from each member of your group.
  - Details of the soft computing methods used. Example:
    - Fuzzy logic: how many membership functions are used for input and output? How many fuzzy rules are in the rule set? Which membership function is used and why? How is the fuzzification and defuzzification performed?
    - Neural networks: which NN model is used? What parameters are used (i.e. hidden layers, number of nodes per layer)?
  - What results are obtained? What metric did you use (i.e. Accuracy, Sensitivity) and what score did you get?
  - If you are using fuzzy logic, extract the fuzzy ruleset from your best model and show it using IF-THEN format.
  - If you are using neural networks, how do you extract explainable information from your model? You can use methods like [SHAP](#) or [LIME](#) to generate explainable information.
  - Or if you are using decision tree methods, then you can plot your model using some of the methods listed here: <https://mljar.com/blog/visualize-decision-tree/>
  - If you are using genetic algorithms, show the graph of the fitness score vs generations, similar to what is done in Assignment 1.

- Analyze the results. Which digit or alphabet has the best and worst accuracy? Which member's handwriting has the best and worst accuracy? Show some examples of correct predictions and wrong predictions.
9. Zip your report and the Jupyter Notebook file and upload to UM Spectrum by **23<sup>rd</sup> June 2023**. Only the group leader needs to submit the files.

Marking scheme (10%):

- Explained the dataset in detail: 1%
- Explained the soft computing methodology in detail: 5%
- Explained the results in detail: 2%.
- Generated explainable information: 2%.