Senior Project:

Description: (Recognition of handwritten characters) Develop an application which apply Image processing and machine learning models to recognize characters (English alphabet and digits) from a hand written document and converts it to text output.

* Use EMNIST dataset to train a CNN model.
* Train CNN model to uses 26x26 numpy array as input vector and UTF-8 encoded digit/alphabets as output

**Deliverables**:

A Test harness with following functions

1. Load dataset <file name>
   * Description: load the dataset and divide in Training and Test Set
   * Input: data file name
   * Output: description of the dataset. (plot with character, frequency)
2. Select Model <model name>
   * Description : Load from different pre-trained models saved in file
   * Input: model name
   * Output: Load the model and display the model name, model type and accuracy with Test Set
3. Predict

<n>

* + Description: use the currently selected model to predict the value for nth data in the Test Set
  + Input: integer less than length of Train Set
  + Output: Display test image and character predicted/recognized

<image>

* + Description: Scan the image file and recognize the Character
  + Input: Image filename
  + Output: Display input image and character predicted/recognized

1. Train Model <model type> <name>
   * Description : Train new models using the currently selected dataset
   * Input: Model Type (i.e Linear, **Decision Trees,** CNN) and a name to save the model.
   * Output: create/save a new model and display the Accuracy.
2. Evaluate
   * Description: Compare different model accuracy.
   * Input: N/A
   * Output: List of model to select and accuracy

Complete project report and documentation.

**Grade C:**

* Function 1-4 with 2 model
* Evaluate model for at least > 40% accuracy.
* Data Set: Small (~300 MB)
* Correctly predict single characters from the test Dataset.

**Grade B:**

* Function 1-5 with 5 more models
* Evaluate model for at least > 60% accuracy.
* Data Set: Medium (~1 GB)
* Correctly scan and recognize single character from image file

**Grade A:**

* Function 1-6
* Evaluate model for at least > 80% accuracy.
* Data Set: Large (~ 3-4 GB)
* Correctly scan and recognize multiple character from image files