**HANDWRITTEN DIGIT RECOGNITION**

Project Group - Survi, Prakash, Aditya

**Introduction**

Handwritten recognition is the ability of a computer to receive and interpret intelligible handwritten input from source such as paper documents, photographs, touch-screen and other devices. In this project, we are planning to use three classification methods for the recognition namely - **Support Vector Machine (SVM)**, **K-Nearest Neighbor (KNN)** and a **Neural Network** to recognize the handwritten digits. We shall be working with two datasets - **The MNIST database of handwritten digits** and the **Semeion Handwritten Digit Data Set**

**The Classification Methods**

SVMs are a set of supervised learning methods used for classification, regression and outliers detection. For our use case, we plan to use a **linear kernel based SVM** as they have more flexibility in the choice of penalties and loss functions and can also scale better for larger samples

KNN is the most simple image classification algorithm. We will perform cross validation to find the optimal value for k and use the **Euclidean distance measure** for neighboring data points. And then we will use this k for testing on our trained KNN model.

And finally we will also be training a Neural Network (Convolutional / Recurrent) for comparison against the other two models.

For the each of these methods we will be extracting the **Histogram of Oriented Gradient (HOG) features** from the pixel-based representation of samples in our datasets to train our models with.

**The Interest**

We are interested in **evaluating and comparing our performance** on the chosen datasets using the classification techniques described above. Potential work has been done for classifying MNIST dataset using the three methods. But we also take into consideration the second Semeion dataset to evaluate the robustness of our classification models.

**Why This Problem?**

We know that the latter scope for our defined problem is **recognition of handwritten characters**. By evaluating on the above classification models we can perhaps save time and choose the ideal model based on the characteristics we acquire after testing on them. Thus, this project adds a nudge towards **optical character recognition** as a whole.

**Databases:**

[1] MNIST Handwritten Digits database of New York University : MNIST is one of the most famous and popular used database for handwritten digits recognition which contains 70,000 samples included two parts of 60,000 and 10,000 samples corresponding to training and test data.

[2] Semeion Handwritten Digit Data Set of UCI: 1593 handwritten digits from around 80 persons were scanned, stretched in a rectangular box 16x16 in a gray scale of 256 values.Then each pixel of each image was scaled into a boolean (1/0) value using a fixed threshold.

**References:**

[1] Norhidayu binti Abdul Hamid and Nilam Nur Binti Amir Sjarif, “Handwritten Recognition Using SVM, KNN and Neural Network”

[2] Reza Ebrahimzadeh and Mahdi Jampour ,“Efficient Handwritten Digit Recognition based on Histogram of Oriented Gradients and SVM”

[3] Subhransu Maji and Jitendra Malik,”Fast and Accurate Digit Classification”