

# **A Novel Method for Handwritten Digit Recognition System**

## **Team Members:**

RISHIKA V (Roll no : 19Z341)  
SWETHA M (Roll no : 19Z355)  
CS THARUNRAJ (Roll no : 19I208)  
GOKULRAJAN N (Roll no : 19I218)

## **Faculty Mentor:**

S Sangeetha

## **Problem Statement:**

To develop an application that is capable of identifying and understanding handwritten digits or characters.

## **Introduction:**

Due to the fact that everyone in the world has a distinctive writing style, handwriting identification is one of the most interesting areas of research now being conducted. It is the ability of a computer to recognise and comprehend manually entered numbers or letters automatically. Everything is becoming digitalized to minimize human effort as a result of advancements in science and technology. As a result, many real-time applications require the ability to recognise handwritten digits. For this recognition process, the MNIST data collection, which contains 70000 handwritten digits, is frequently employed. We train these photos using artificial neural networks and create a deep learning model. The user can upload an image of a handwritten digit using a web application that has been constructed. The model analyses this picture and returns the result.

## **Literature Survey:**

### **1. A Novel Method for Persian Handwritten Digit Recognition using Support Vector Machines**

Persian handwritten digits categorization has been confronting challenges because of various handwriting styles, between class likenesses, and intra-class contrasts. In this paper, an original technique for identifying Persian handwritten digits is introduced. In the proposed strategy, a blend of Histogram of Arranged Inclinations (Hoard), 4-side profiles of the digit picture, and a few flat and vertical examples was utilized and the element of the component vector was decreased utilizing Head Part Investigation (PCA). The proposed technique applied to the HODA data set, and Support Vector Machine (SVM) was utilized in the classification step.

HODA is one of the biggest and challenging datasets of Persian handwritten digits. This dataset includes 60000 training data samples and 20000 test ones. The proposed method can be divided into three major parts, namely pre-processing, feature extraction and classification. the dimension proportion of various pictures is different intrinsically. For example, pictures of '1' or '9' are outlined upward, while '0' and '5' pictures are the square shapes. For overcoming this issue, a picture figuring out algorithm is directed. The proportion of the length to the width of the image is determined and approximately zero lines or sections are cushioned to it evenly, on the off chance that that the proportion is less than 0.95 or more prominent than 1.05. In the subsequent stage, the size of all images are changed, to produce exceptional block size and feature length during applying the feature extractor algorithm. The final pre-handling part is noise reduction and Binarization of the pictures. Several features are extricated from each picture and connected together to construct feature vectors. They are Hoard, 4-side profiles of the digit picture and some level and vertical examples. Histogram of situated slopes (Hoard) is a strong feature extracting strategy which is often utilized in character and digit recognition algorithms in various dialects.

Results uncovered that the recognition precision of such a strategy has close to 100% exactness with a sufficient rate due to existing unsatisfactory examples in the data set. In this manner, the proposed

technique could further develop the results thought about to other existing strategies.

## **2. A novel method for Handwritten Digit Recognition with Neural Networks**

Handwritten character recognition is a system generally utilized in the world to perceive postal districts or postal code for mail arranging. There are various strategies that can be utilized to perceive handwritten characters. In this paper two procedures explored are Pattern Recognition and Artificial Neural Network (ANN). Both procedures are characterized and various strategies for every method are additionally talked about. Bayesian Decision theory, Nearest Neighbor rule, and Straight Arrangement or Segregation are sorts of techniques for Pattern Recognition. Shape recognition, Chinese Character and Handwritten Digit recognition utilizes Neural Network to remember them. Neural Network is utilized to train and distinguish written digits.

An image is fed into the network to train. Back-propagation neural network is utilized for training the network. In spite of the fact that it is just a single image, it contains 100 examples of the equivalent number. For each 10 ages, the data is saved into the network. In the wake of training, the network was tried and the accuracy rate came to almost 100%. This is an extremely high accuracy rate. The network was not steady on the grounds that the training results changed regularly. On the off chance that we take numerals — 2" as an example, today we would need to train multiple times to reach almost 100% accuracy, yet tomorrow perhaps we need to train multiple times in order to arrive at close to 100% accuracy.

Utilizing the Neural Network system, back-propagation learning, to perceiving handwritten digits was extremely effective. An image, which contained 100 samples of each number, was trained and tested. The accuracy rate of perceiving the number was close to 100%. This accuracy rate is extremely high. From the training and testing results, it was inferred that the system had more inconvenience distinguishing numerals — 5". This perhaps brought about by the reality that the digit is running together or perhaps it isn't completely associated. The system was not steady. It gave unique preparing and testing results consistently for every numeral.

From the net-document, the system had the option to create an image-record. The image-record created showed the perceived number.

### **3. A Novel Approach to Recognize the off-line Handwritten Numerals using MLP and SVM Classifiers**

In this paper, a novel method for offline handwritten numeral recognition is presented. One of the hardest tasks in pattern recognition has been reading handwritten digits. The significant degree of individual variation in numeral shapes makes it difficult to recognise handwritten numerals. This study uses MLP and SVM classifiers to recognise handwritten numbers offline. What kind of feature extraction approaches are being employed has a significant impact on how well a character recognition system performs.

Optical character recognition, also referred to as OCR, is the mechanical or electronic conversion of images of handwritten, typewritten, or printed text into machine-editable text or computer processable format, like ASCII code (typically acquired by a scanner). Every time a page is scanned, a bit-mapped file of that page is kept. We can read the image when it is shown on the screen. But to the computer, it is just a collection of dots. Any "words" on the image are not recognised by the computer. These words are read by the computer using OCR. It analyses each line in the image to identify which specific character is represented by which dot.

The various individuals' data have been collected for the experiment. Currently the developing dataset is for English numerals. 1200 instances of handwritten numerals from 24 different writers total. Each writer received a blank A4 page and was instructed to write the numbers 0 through 9 five times. The database was built to validate the recognition system and is completely unrestricted. The HP-scan jet 5400c scans the collected papers at 300 dpi, producing typically clean, low noise images. The digitized photos are saved in BMP format as binary images.

**Support vector machine:** SVM is a class of learning algorithms used for regression and classification. The hyper plane or set of planes created by support vector machines is used to divide the data into two classes. SVM classifies unknown data based on a collection of labeled

training data sets and employs the supervised learning methodology, which implies the training data is managed by an external agent (experts). The input data is mapped by SVM into a higher-dimensional space, where a hyperplane with the greatest possible separation is built. In order to create integrated multiple binary SVM classifiers for multiclass classification, we divide the multiclass classification issue into multiple binary class problems.

**Multi layer Perceptron classifier:** MLP's are feed-forward networks of simple processing units (neurons) with at least one "hidden" layer. The MLP classifier is used for the classification. The MLP network consists of three layers namely, input layer, hidden layer and output layer.

## **References:**

[1] A Novel Method for Persian Handwritten Digit Recognition using Support Vector Machines by Mojtaba Mohammadpoor, Abbas Mehdizadeh2 and Hava Alizadeh Noghabi

[2] A novel method for Handwritten Digit Recognition with Neural Networks by Malothu Nagu, N Vijay Shankar and K.Annapurna

[3] A Novel Approach to Recognize the off-line Handwritten Numerals using MLP and SVM Classifiers by Mamta Garg and Deepika Ahuja