A Project Report on

**HANDWRITTEN CHARACTER RECOGNITION**

Submitted in partial fulfillment of the requirements

in

### COMPUTER DEPARTMENT

by

### Shweta chauhan(17102009)

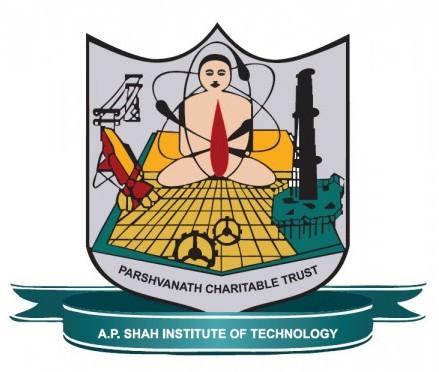
### Nehal Barot(17102060)

### Yash awasthi(15202013)

### Anjali Solanki(18202003)

Under the Guidance of

### Prof. Amol Kalugade



#### Department of Computer Engineering

#### A.P. Shah Institute of Technology G.B.Road,Kasarvadavli, Thane(W), Mumbai-400615

UNIVERSITY OF MUMBAI

#### Academic Year 2020-2021

**Approval Sheet**

This Project Report entitled ***“Handwritten Character Recognition”*** Submitted by ***“Shweta Chauhan”(17102009),“Nehal Barot”(17102060),“Yash Awasthi”(15202013),“Anjali Solanki”(18202003)***is approved for the par tial fulfillment of the requirement in ***Branch Name*** from ***University of Mumbai*** .

(Name)

Guide:Prof.Amol Kalugade

Prof. S.H.Malave

Head, Computer Engineering Department

Place:A.P.Shah Institute of Technology, Thane Date:

### CERTIFICATE

This is to certify that the project entitled ***“Title of project”*** submitted by ***“Shweta Chauhan” (17102009),“Nehal Barot” (17102060),“Yash Awasthi” (15202013),“Anjali Solanki” (18202003)*** for the partial fulfillment of the requirement for award of a degree ***Bachelor of Engineering*** in ***Computer Department***,to the University of Mumbai,is a bonafide work carried out during the academic year 2020-2021.

(Name)

Guide:Prof.Amol Kalugade

Prof. S.H.Malave Dr. Uttam D.Kolekar Head,Computer Engineering Department Principal

External Examiner

Place:A.P.Shah Institute of Technology, Thane Date:

### Declaration

We declare that this written submission represents our ideas in our own words and where others’ ideas or words have been included, We have adequately cited and referenced the orig- inal sources. We also declare that We have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

———————————————

(Signature)

———————————————

(Shweta chauhan,17102009)

(Nehal Barot,17102060)

(Yash Awasthi, 15202013)

(Anjali Solanki, 18202003)

Date:16/12/2020

**1.project conception and initiation**

1.1 Abstract

Handwriting character recognition remains largely unsolved problems due to the presence of many handwritten characters present around the world. There are many existing advance methods which do not lead to a proper solution for handwriting character recognition. In this paper we will describe the approach to get

maximum accuracy (90%) in the field of handwriting character recognition. The handwriting character recognition will be done by using PyTesseract, Convutional Neural Network and Tensor Flow.Automatic identification of handwritten script facilitates many important applications such as automatic transcription of

multilingual documents and search for documents on the Web containing different particular scripts. The increase in usage of handheld devices which accept handwritten input has created a growing demand for algorithms that can efficiently analyze and retrieve handwritten data. Handwritten character recognition is an

area of research where many researchers have presented been done and is still an area under research to achieve more accuracy. This project is aimed at developing software which is helpful to recognize English characters. Our project is restricted to English characters only. It is still possible to further develop to recognize the characters of many languages present in the world. The most common use of neural network is

pattern recognition. Neural network is used for solving problems such as the recognition of patterns, classifying the patterns into groups, data mining etc. Neural network contains a vector which have the pattern

information as well as a target vector. Here the pattern information could be an image and handwritten data. Neural network attempts to determine the handwritten characters or image that the neural network has learned. Neural network is designed to take the input data and classify the data into groups. These groups can be fuzzy

or not clear.

1.2 Objectives

1. To provide an easy user interface to input the source image.

2. User should be able to upload the image.

3. System should be able to pre-process the given input to suppress the

Background.

4. System should detect text regions present in the image.

5. System should retrieve text present in the image and display them to the

User.

1.3 Literature review

A..Neural Networks For Handwritten English Alphabet Recognition The paper makes use of Neural Networks to recognize the English handwritten alphabets. The alphabets are represented as binary values in the form of 0 and 1. These binary images are then used as an input to the feature extraction phase, and the output of this phase is fed as an input in Neural Network system. Similar appearing alphabets were seen to be misclassified in the experiments carried out, as the data set over which the tests were carried out was small. The larger the data set used the more accurate results the system would generate

B. Direction Based Feature Extraction Paper presents a feature extraction method referred as the direction feature to recognize the handwritten characters. This new direction feature extraction method proved to give better accuracy result on the segmented characters when compared to the transition feature extraction method and other methods as well. The recognition rate is above 80% proven experimentally. The characters are segmented automatically from the Cursive Digit and character Recognition (CEDAR) benchmark

C. Diagonal Based Feature Extraction System Using Neural Network The characters in this paper are recognized using a new technique called the diagonal feature extraction technique.There were two approaches used, firstly with 54 features and secondly with 69 features which constituted the Neural Network character recognition system. The diagonal feature extraction method was compared with the traditional horizontal and vertical feature extraction methods, by training the neural network with both the methods. Six recognition neural networks were built in process. The results experimentally revealed that 69 features gave a better accuracy rate than 54 features. The system presented in the paper seems ideal to convert handwritten text documents into structural format.

D. Handwritten English Character Recognition Using Neural Network The paper demonstrates the use of Feed Forward Algorithm along with Back Propagation Algorithm. We paper successfully recognizes characters using a multilayer perceptron with 1 hidden layer. Higher performance can be achieved in Back Propagation once the number of hidden nodes to be used is successfully determined. The recognition of characters is proved to be better and gives an accuracy of 70% and above for English handwritten characters.

E. Digital Image Processing Techniques In Character Recognition The paper explains the various phases of image processing being used in character recognition such as Image Restoration, Image Enhancement, Segmentation, Feature Extraction and Classification with Recognition. These techniques are used along with Neural Network due to its high tolerance to noise. This helps in removing all unwanted signals in images that are distorted over years. Successful character recognition becomes possible for such documents as well and the systems generate perfect results.

F. Character Recognition Using Neural Network In this paper, neural network is used to recognize characters. It improves the recognition rate as the system is developed for isolated English characters – A to Z. The paper makes use of feed forward back propagation and the Neural Network is trained using Back Propagation to classify and recognize characters. The English characters are represented in the binary form as is then fed to the Neural Network for further processing. The paper fails to recognize cursive handwritings.

G. Handwritten Digit Recognition The paper presents a comparison of the feature vectors, the feature extraction strategies are proven to perform better than their baseline counterparts. The gradient feature extraction technique works best for gray

scale images giving the most accuracy rate of characters and also the Normalization-Cooperated Feature Extraction (NCFE) yields a good performance result. The gradient feature extraction technique is applied on the gray scale images and other feature extraction techniques are applied on the binary images. The combination of feature extraction along with normalization has proven to yield higher accuracy rates of

character recognition.

H. Analysis Drawn

After successfully reviewing a number of papers, the following experimental results were analyzed and a comparison of the character recognition accuracy rates is given below along with the different algorithms and techniques used in each corresponding paper.

1.4 Problem Definition

Traditionally, the system exists only for the character recognition but our project extends this further and converts those characters into digital texts and store them in textual form. Moreover characters are only read through OCR (Optical Character Reader). These text are read by OCR are printed on product items, in our project we aim to recognize and convert the handwritten notes to their digital textual from. Traditional systems are not highly efficient but our project gives 85-90 per cent efficiency due to large datasets and precisely trained neural network model for character recognition and correction.

1.5 Scope

1. Converting handwritten notes to digital notes.

2. Converting old scriptures to digital library.

3. House Number Recognition through Google street images.

4. Physical written regional language to other.

5. Quick digitization of printed hand-filled form.

6. Converting postal address to digital texts.

1.6 Technology stack

## 1.PyTesseract: Simple Python Optical Character Recognition

2.Python OpenCV

3.Tensor Flow

4.Neural Networks

1.7 Benefits for environment and society

### **Healthcare and pharmaceuticals**

Patient prescription digitization is a major pain point in healthcare/pharmaceutical industry. For example Roche is handling millions of petabytes of medical PDFs daily. Another area where handwritten text detection has key impact is patient enrollment and form digitization. By adding handwriting recognition to their toolkit of services, hospitals/pharmaceuticals can significantly improve user experience

### **Insurance**

A large insurance industry receives more than 20 million documents a day and a delay in processing the claim can impact the company terribly. The claims document can contain various different handwriting styles and pure manual automation of processing claims is going to completely slow down the pipeline

### **Banking**

People write cheques on a regular basis and cheques still play a major role in most non-cash transactions. In many developing countries, the present cheque processing procedure requires a bank employee to read and manually enter the information present on a cheque and also verify the entries like signature and date. As a large number of cheques have to be processed every day in a bank a handwriting text recognition system can save costs and hours of human work

### **Online Libraries**

Huge amounts of historical knowledge is being digitized by uploading the image scans for access to the entire world. But this effort is not very useful until the text in the images can be identified which can be indexed, queried and browsed. Handwriting recognition plays a key role in bringing alive the medieval and 20th century documents, postcards, research studies etc.

**2.Project Design**

2.1 proposed system

In this section, the proposed recognition system is described. A typical handwriting recognition system consists of pre-processing, segmentation, classification and post processing stages. Image acquisition, the recognition system acquires a scanned image as an input image. The image should have a specific format such as JPEG, BMT etc. This image is acquired through a scanner, digital camera or any other suitable digital input device.

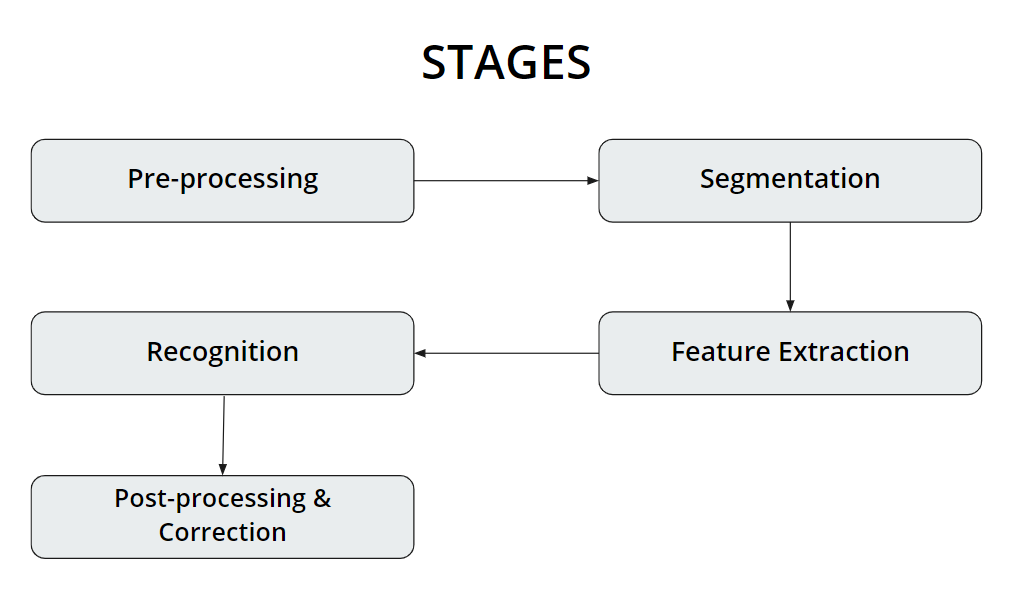
Pre-processing: The pre-processing is a series of operations performed on the scanned input image. It essentially enhances the image rendering it suitable for segmentation. The various tasks performed on the image in pre-processing stage are shown in Fig.2. Binarization: process converts a gray scale image into a binary image using global thresholding technique. Dilation of edges in the binarized image is done using sobel technique, dilation the image and filling the holes present in it are the operations performed in the last two stages to produce the pre-processed image suitable for segmentation.

Segmentation: In the segmentation stage, an image of sequence of characters is decomposed into sub-images of individual character In the proposed system, the pre-processed input image is segmented into isolated characters by assigning a number to each character using a labeling process. This labeling provides information about number of characters in the image. Each individual character is uniformly resized into 30X20 pixels.

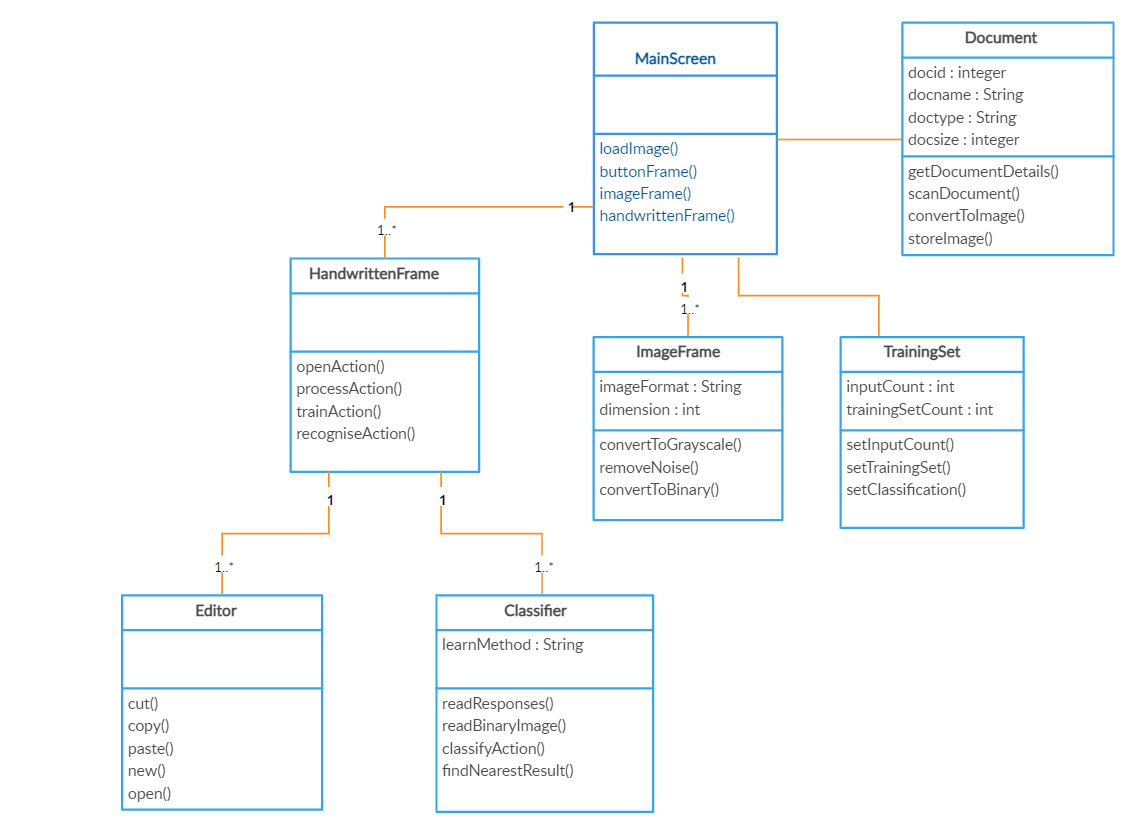
Classification and Recognition: The classification stage is the decision making part of the recognition system . A feed forward back propagation neural network is used in this work for classifying and recognizing the handwritten .characters. The 600 pixel derived from the resized character in the segmentation stage form the input to the classifier. The neural classifier consists of two hidden layers besides an input layer and an output layer as shown in. The hidden layers use log sigmoid activation function and the output layer is a competitive layer as one of the characters is required to be identified at any point in time. The total number of neurons in the output layer is 26 as the proposed system is designed to recognize English alphabets

Post-processing: Post-processing stage is the final stage of the proposed recognition system. It prints the corresponding recognized characters in the structured text form by calculating equivalent ASCII value using recognition index of the test samples.

2.2 Design(flow of modules)



2.3 class diagram



2.4 module-1

module-2.....

.....module-n

**3.Planning for next semester**

Planning to execute the program with optical features and training of dataset .