HANDWRITTEN CHARACTER RECOGNITION

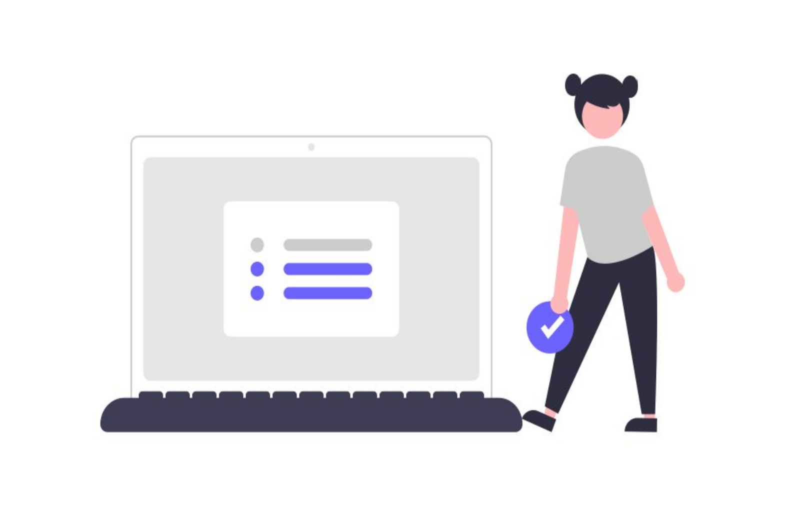
Project Members :

Shweta Chauhan

Nehal Barot

Anjali Solanki

Yash Awasthi



ABSTRACT

Handwriting Character Recognition has been one of the active and challenging research areas in the field of image processing and pattern recognition. It has numerous application which includes reading aid for blind, bank cheques and conversion of any handwritten document into structural text form. The handwriting character recognition will be done by using PyTesseract, Convolutional Neural Network and Tensor Flow.

INTRODUCTION

The problems in handwritten character recognition (HCR) system is the difference of the handwriting styles, which can be completely unlike for different writers. HCR is mainly classified into offline and online character recognition. In offline handwritten character recognition, the handwriting is scanned and understood by computer. In online character recognition, handwriting is recognized while writing through touch pad using stylus pen etc.

Neural network is used for solving problems such as the recognition of patterns, classifying the patterns into groups, data mining etc. Neural network is designed to take the input data and classify the data into groups. These groups can be fuzzy or not clear.

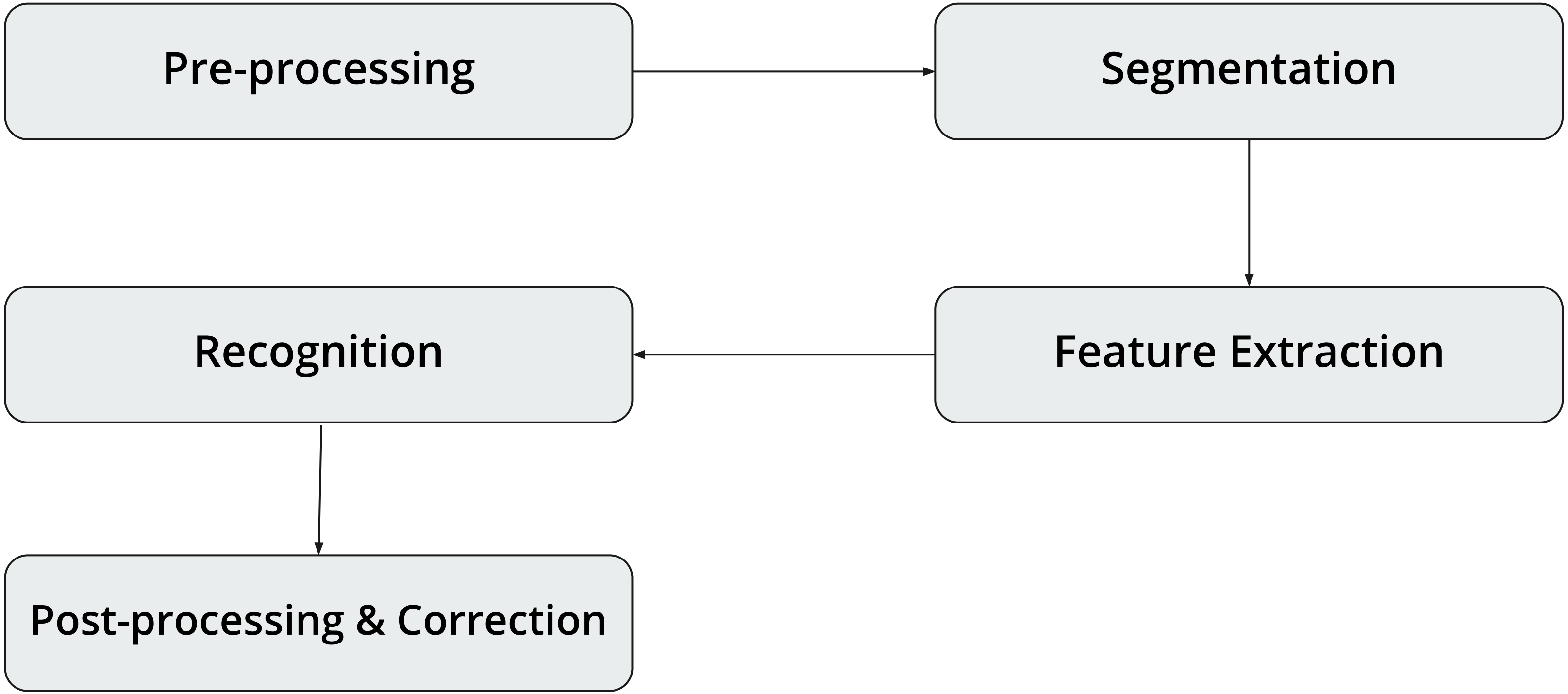
PROBLEM ?

The problems in handwritten character recognition (HCR) system is the difference of the handwriting styles, which can be completely unlike for different writers.

Handwritten physical documents are difficult to store and access in an efficient manner, search through them efficiently and to share them with others.

How can we **Improve the efficiency of manual conversions by automating the process.**

STAGES



Pre-processing

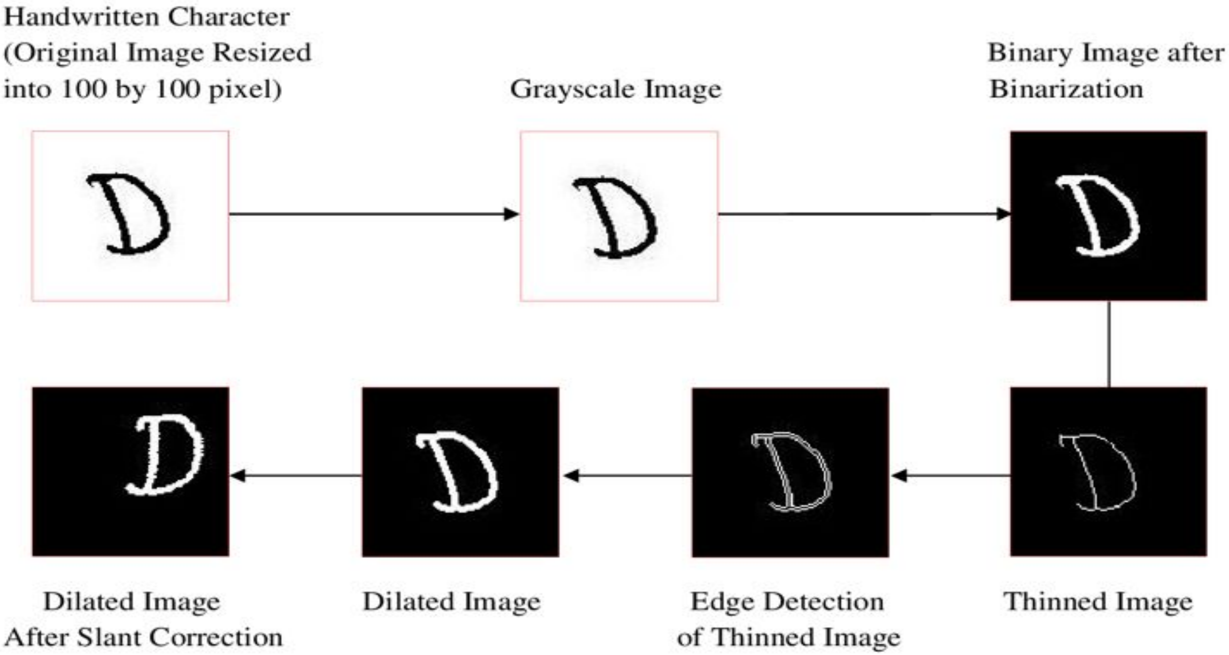
Segmentation

Feature Extraction

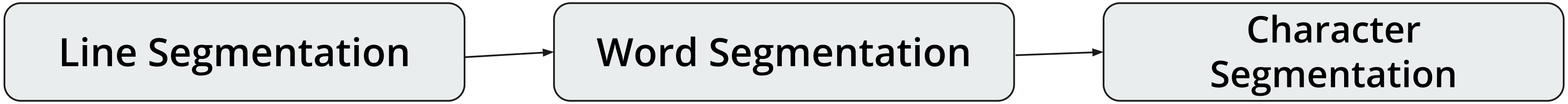
Recognition

Post-processing & Correction

Pre-Processing


Handwritten Character (Original Image Resized into 100 by 100 pixel)
Grayscale Image
Binary Image after Binarization
Thinned Image
Edge Detection of Thinned Image
Dilated Image
Dilated Image After Slant Correction

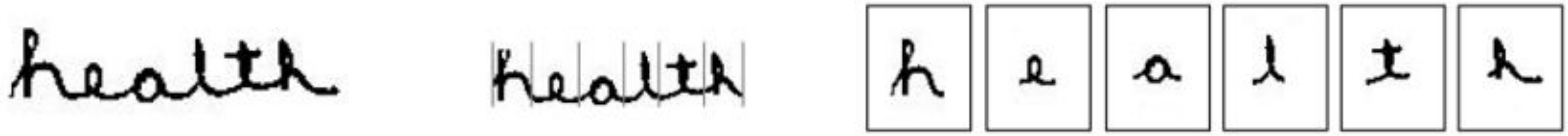
Segmentation



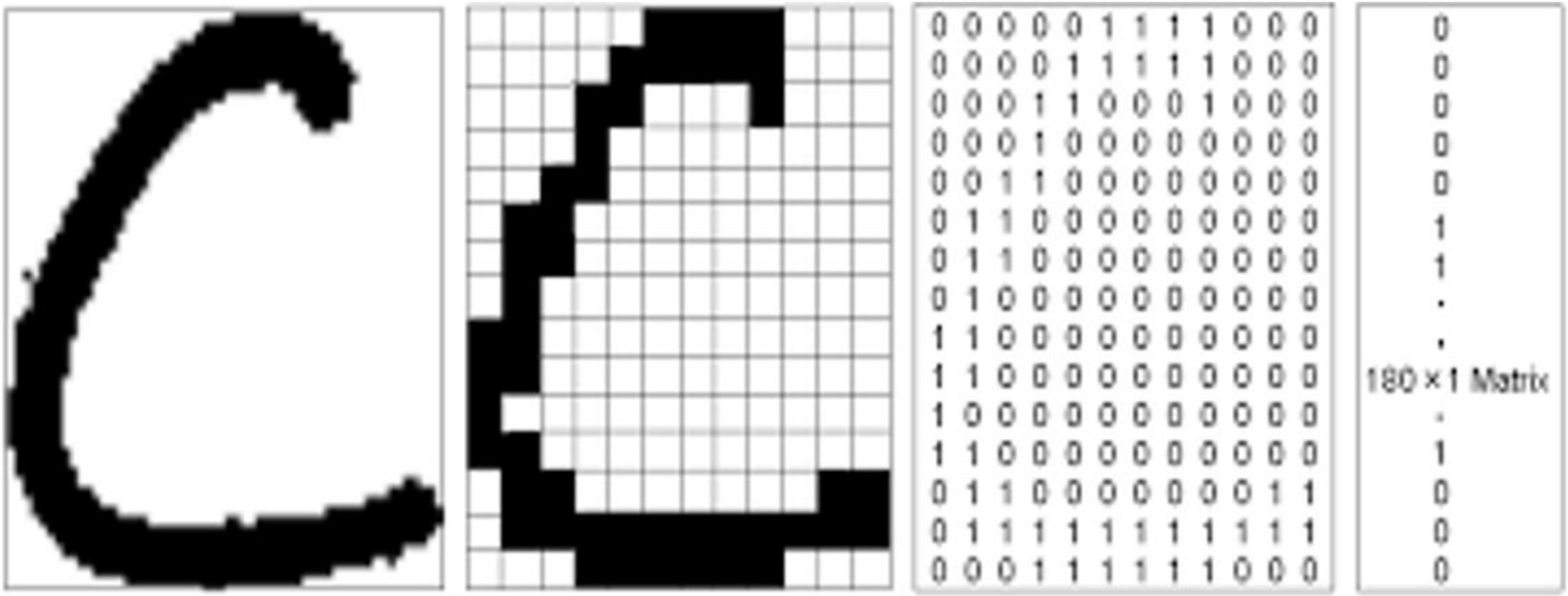
Line Segmentation

Word Segmentation

Character Segmentation



Feature Extraction



(a)

(b)

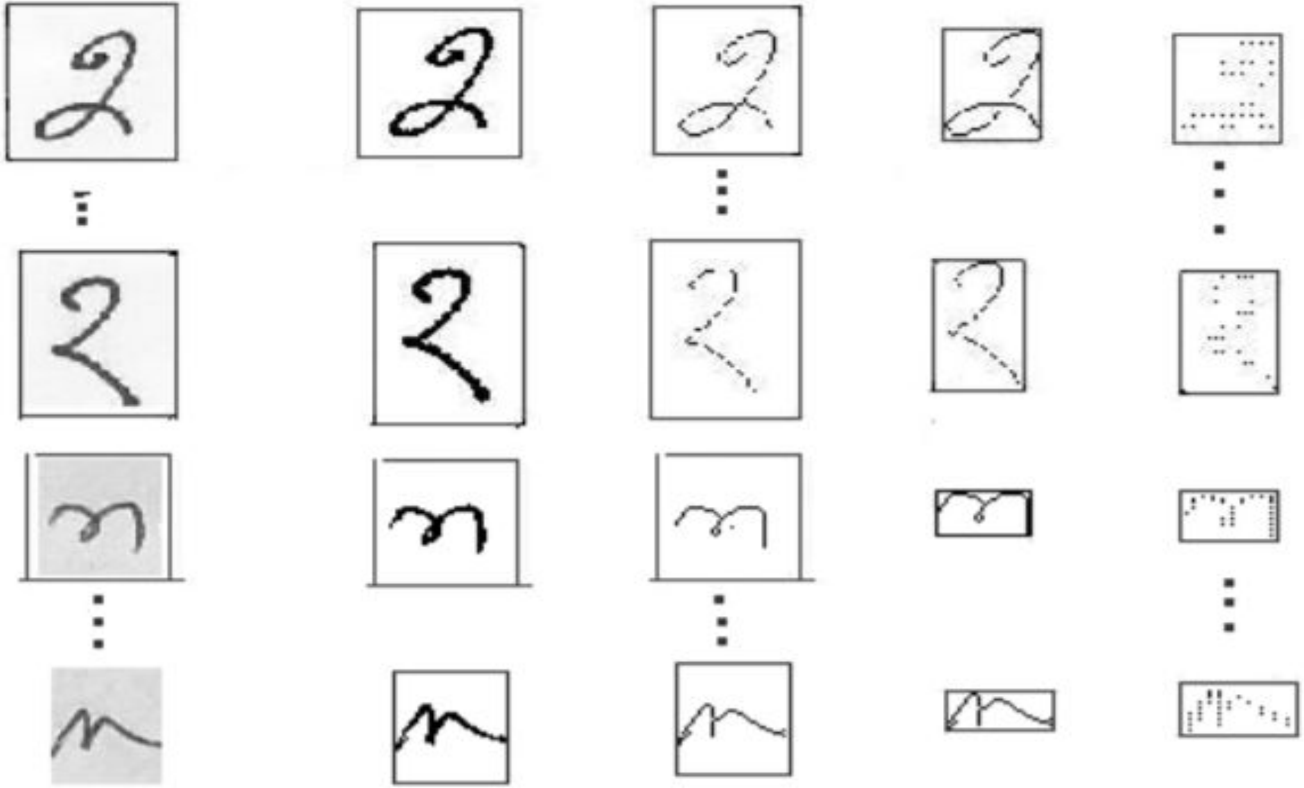
0	0	0	0	0	1	1	1	1	0	0	0
0	0	0	0	1	1	1	1	1	0	0	0
0	0	0	1	1	0	0	0	1	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0
0	0	1	1	0	0	0	0	0	0	0	0
0	1	1	0	0	0	0	0	0	0	0	0
0	1	1	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0
0	1	1	0	0	0	0	0	0	0	1	1
0	1	1	1	1	1	1	1	1	1	1	1
0	0	0	1	1	1	1	1	1	0	0	0

(c)

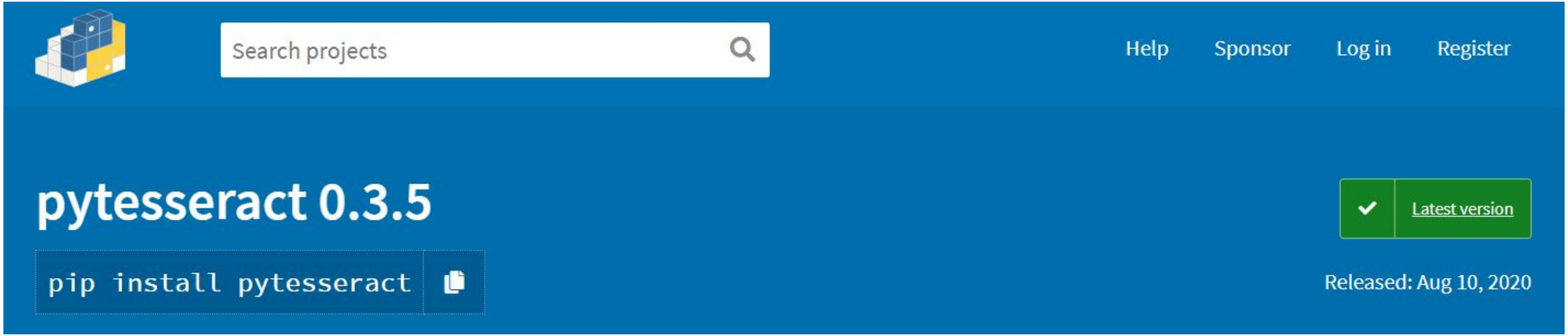
0
0
0
0
0
1
1
.
.
180 × 1 Matrix
.
1
0
0
0

(d)

Recognition



TECHNOLOGIES USED



pytesseract 0.3.5

```
pip install pytesseract
```

Released: Aug 10, 2020

Python OpenCV

High Processing Computer Vision Library

Used for Input Image Processing

TECHNOLOGIES USED

TensorFlow

To train the software for recognition, TensorFlow provides an open source library.

Step 1 : Configuring the Project

Step 2 : Importing the MNIST Dataset

Step 3 : Defining Neural Network architecture

Step 4 : Building the TensorFlow Graph

Step 5 : Training and Testing

CONCLUSION

Many regional languages throughout the world have different writing styles which can be recognized with HCR systems using proper algorithm and strategies.

Scanned image is pre-processed to get a clean image and the characters are isolated into individual characters. Preprocessing work is done in which normalization, filtration is performed using processing steps which produce noise free and clean output. Managing our evolution algorithm with proper training, evaluation of step wise process will lead to successful output of system with better efficiency.

This work will be helpful to the researchers for the work towards other script.

FUTURE SCOPE & USAGE

1. 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 forms
6. Postal address to digital texts.

THANK YOU !