HANDWRITING RECOGNITION USING MACHINE LEARNING
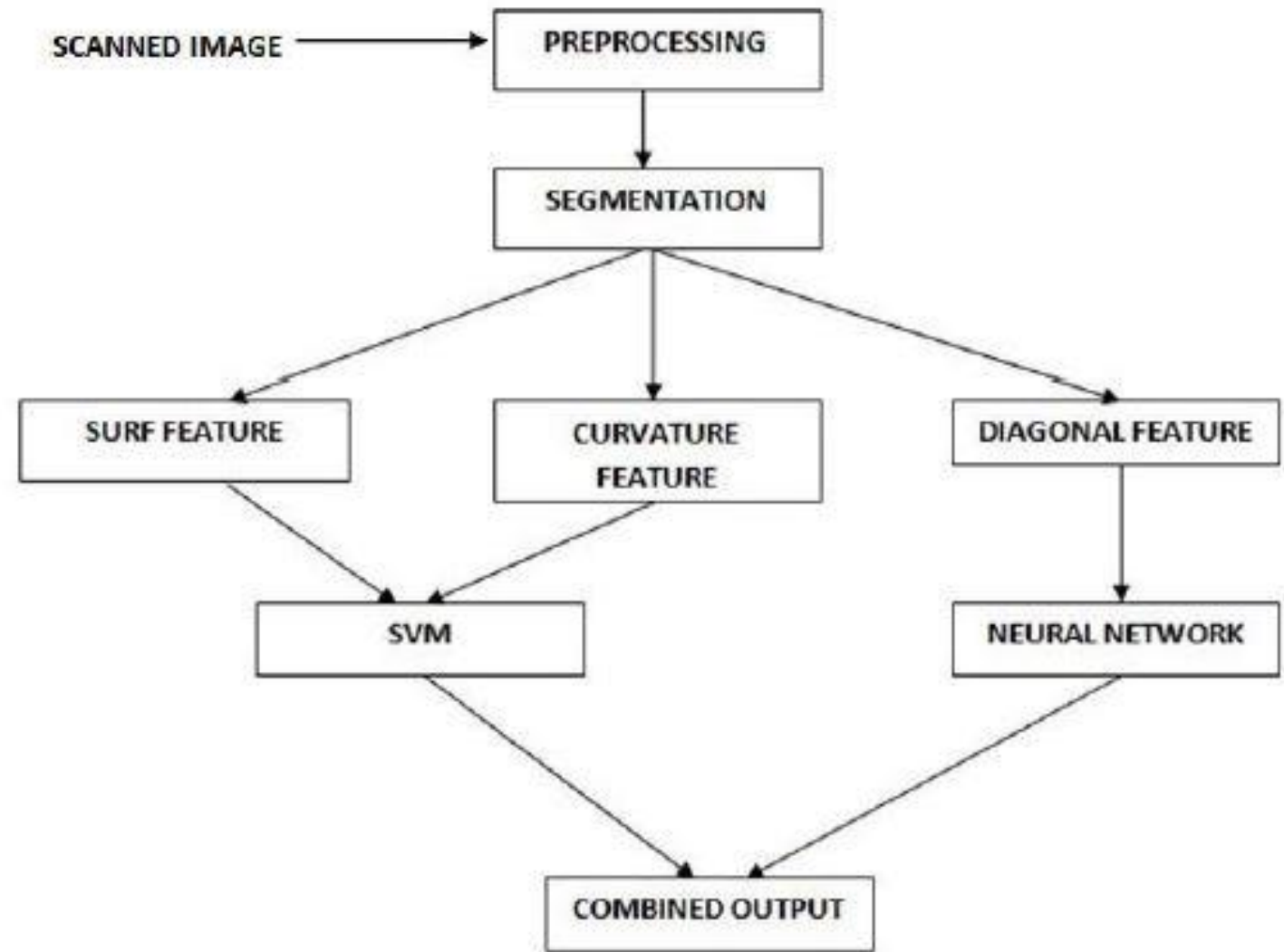
MAIN PROJECT

INTRODUCTION

- The handwriting recognition is a challenging as well as emerging area of pattern recognition. It is a tedious process mainly due to its enormous character set.
- Here we propose a novel method for handwriting recognition by using two dissimilar classifiers. It can also be called as an ensemble method in which multiple classifiers are combined to solve a particular problem and thereby improve the performance of the system.
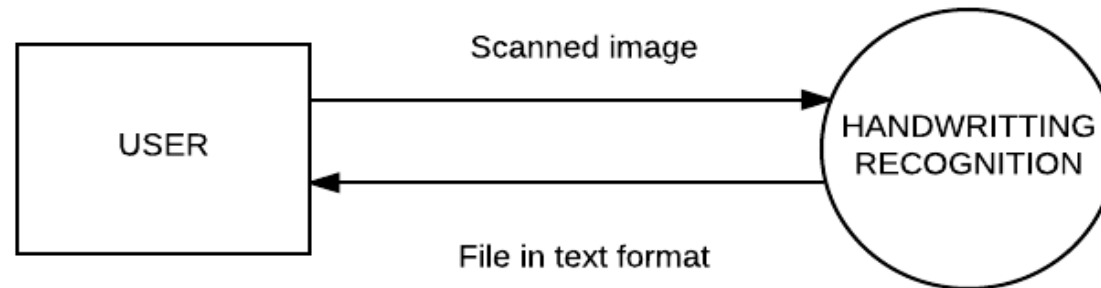- The experiment is conducted in 2 phases.

PROPOSED METHOD

The proposed system uses Image processing as well as the use of two dissimilar classifiers. This is aimed to produce a better and efficient method of solving the digitization problem.

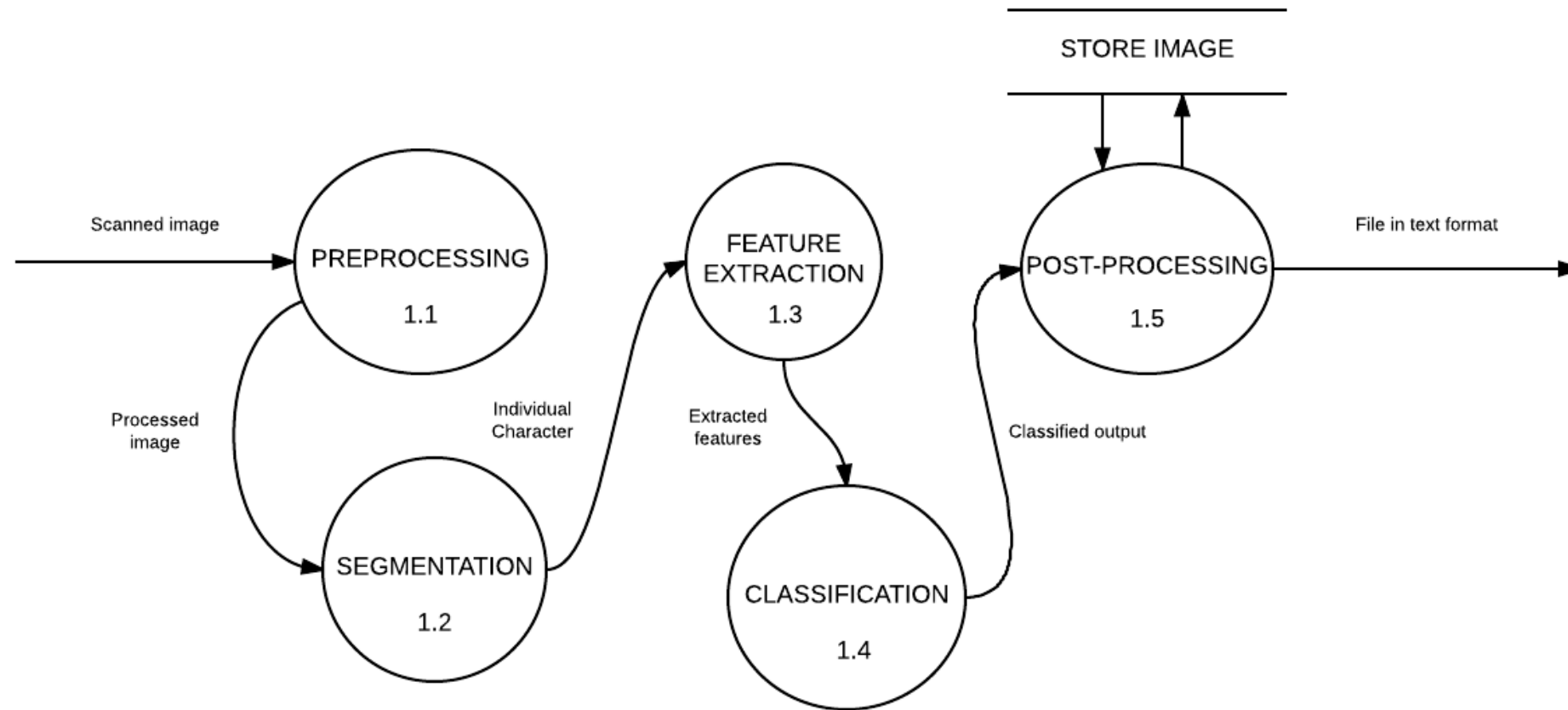The various phases of the project will be explained in the upcoming slides.
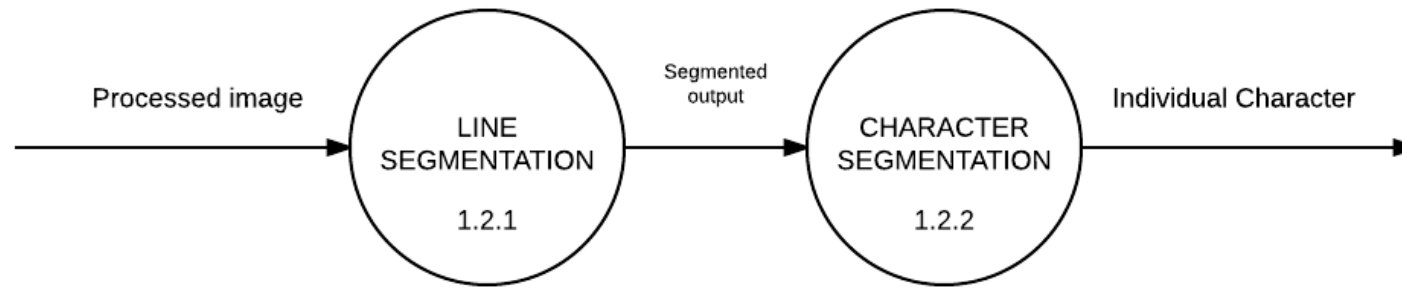
FLOW OF CONTROL

DATA FLOW DIAGRAM
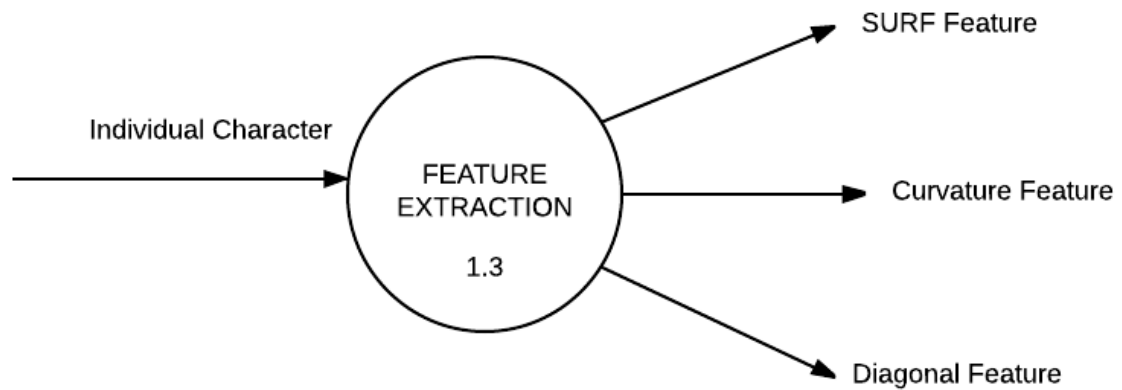
STORE IMAGE

Scanned image → PREPROCESSING 1.1

Processed image

SEGMENTATION 1.2

Individual Character

FEATURE EXTRACTION 1.3

Extracted features

CLASSIFICATION 1.4

Classified output

POST-PROCESSING 1.5

File in text format

LEVEL-1

Processed image → LINE SEGMENTATION 1.2.1 → Segmented output → CHARACTER SEGMENTATION 1.2.2 → Individual Character

LEVEL 1.2

Individual Character → FEATURE EXTRACTION 1.3 → SURF Feature / Curvature Feature / Diagonal Feature

LEVEL 1.3

SURF Feature

Curvature Feature

SUPPORT
VECTOR
MACHINE

1.4.1

Diagonal Feature

NEURAL
NETWORK

1.4.2

RESULT
AGGREGATOR

1.4.3

Classified output

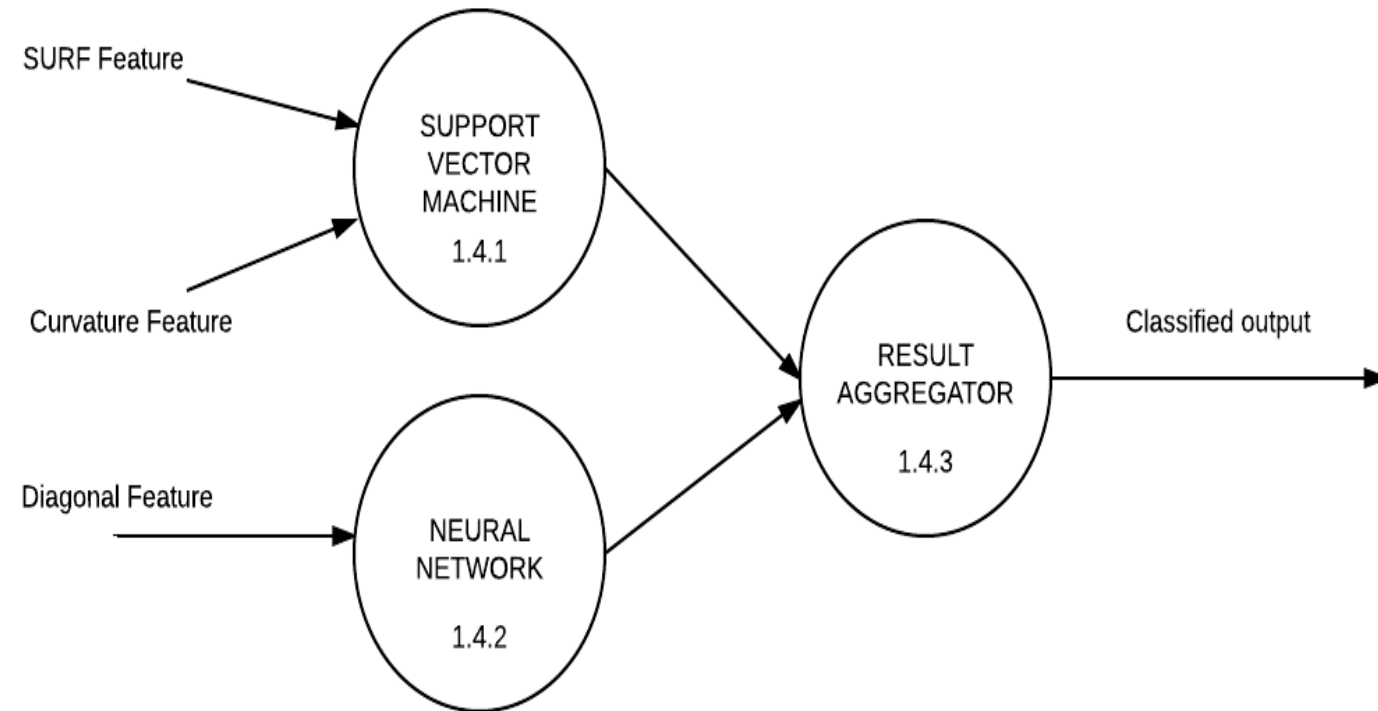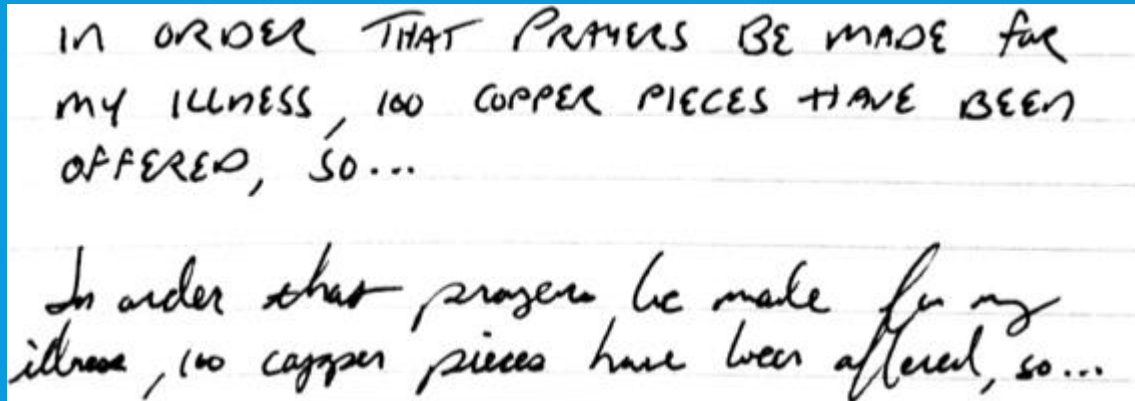**LEVEL 1.4**

MODULE DESCRIPTION

The different modules that we plan to use to implement the project are;

- Image Acquisition
- Pre-processing
- Segmentation
- Feature Extraction
- Classification
- Post-processing

1. IMAGE ACQUISITION

- Handwritten scanned Images can be represented in any format such as JPEG, PNG, etc.
- The Image should have a minimum resolution

These images are given as input to the system for further steps.

2. PRE-PROCESSING

This step removes the distortions in the scanned image. The steps involved are,

1. Converting to grey scale.
2. Converting the grey scale image to binary. This is done by using the OTSU'S Method of global thresholding.
3. Noise reduction using filter.
4. Relevant features are extracted by thinning using HILDITCH Algorithm.

3. SEGMENTATION

The purpose of this step is to isolate individual characters.

1. Line segmentation
   - Using horizontal projection profile method
2. Character segmentation
   - Using vertical projection profile method

   The step produces individual characters from the scanned image

4. FEATURE EXTRACTION

The module extracts salient characteristics of the individual characters. The characteristics are ;

- SURF- It is used to distinguish between points and intersections. It produces 3 parts
    i. Interest point detection
    ii. Local neighborhood description
    iii. Matching
- Curvature – It provides clearer structural sketch of the image
- Diagonal – Input is divided into zones and the zones provide features to the character

5. CLASSIFICATION

In training phase, based on the features extracted unique labels are assigned to each characters. In recognizing phase the characters are mapped to the previously assigned labels stored in the database. Two dissimilar classifiers are used ;
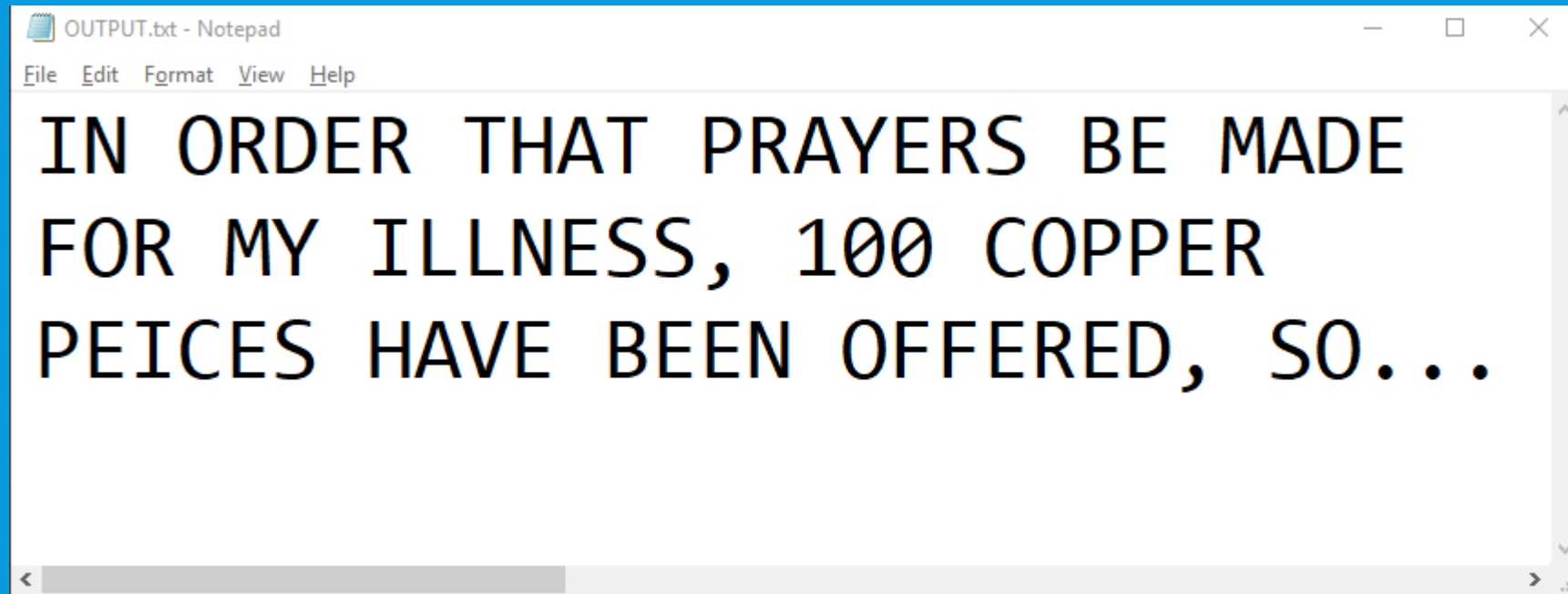
- SVM (Support Vector Machine)
   It is a supervised learning model that performs linear classification. The SVM learns fast but the prediction is slow.
- Neural Network
   The Neural Network learns by observing data. The learning process is slow but NN predicts swiftly.

6. POST-PROCESSING

The characters are mapped to the corresponding Unicode values. These values are used when comparing new data input .
The Post-Processing results are stored in the database in the training phase.
In the recognition phase the comparison takes place.

EXPECTED OUTPUT

SCOPE

- Now is the age of digitalization. Rather than storing the data on papers, it can be safely stored and easily accessed, then digitalized. Therefore Handwriting Recognition is of prime importance. Handwriting Recognition is an emerging as well as challenging area in the fields of pattern recognition and computer vision. Other applications include automatic number plate recognition, CTS scanning, preservation of degraded documents etc.
- The aim of a handwriting recognition system is to convert human readable characters which are present in a photographed or digitized sheet of paper and convert it into a machine editable form.
- An upgradation to this project can result in the system to detect a particular persons handwriting. The future scope can be further studies to improve the current system.

CONCLUSION

In this project, we propose a method for the recognition of Handwritten characters using a combination of dissimilar classifiers. The work will be conducted in 2 phases. In order to identify individual characters, we extract SURF, curvature and diagonal feature and for classifying Neural Network and SVM are used.