

# MAJOR - 1 PROJECT

## SYNOPSIS

For

Handwriting Character Recognition

Submitted By

Specialization	SAP ID	Name
Artificial Intelligence & Machine Learning	500075183	Piyush Malviya
Artificial Intelligence & Machine Learning	500076400	Mudit Dagar
Artificial Intelligence & Machine Learning	500076579	Prateek Sehrawat
Artificial Intelligence & Machine Learning	500075154	Rahul Dhanola



Department of Informatics

School Of Computer Science

UNIVERSITY OF PETROLEUM & ENERGY STUDIES,

DEHRADUN- 248007. Uttarakhand

Dr. Virender Kadyan  
**Project Guide**

Dr. Thipendra P Singh  
**Cluster Head**



**School of Computer Science**  
University of Petroleum & Energy Studies, Dehradun

## **Synopsis Report**

### **Handwriting Character Recognition**

#### **Abstract**

Handwriting Character Recognition is the mechanical or electronic conversion of scanned or photographed images of typewritten or printed text into machine-encoded/computer-readable text. It is widely used as a form of data entry from some sort of original paper data source, whether passport documents, invoices, bank statement, receipts, business card, mail, or any number of printed records. It is a common method of digitizing printed texts so that they can be electronically edited, searched, stored more compactly, displayed on-line, and used in machine processes such as machine translation, text-to-speech, key data extraction and text mining. Handwriting Character Recognition is a field of research in pattern recognition, artificial intelligence and computer vision. Handwriting Character Recognition is the electronic translation of handwritten text into machine translated text. It is widely used to recognize and search text from electronic documents or to publish the text on a website. A large number of research papers and reports have already been published on this topic. A Handwriting Character Recognition has a variety of commercial and practical applications in reading forms, manuscripts and their archival etc. The project tries to create software for recognition of a handwritten text from handwritten text images. It uses computer vision and machine learning. And it experiments with different approaches to the problem.

## Table of Contents

Topic		Page No
Table of Content		2
1	Introduction	3
2	Literature Review	4
3	Problem Statement	5
4	Objectives	5
5	Methodology	5
6.	PERT Chart	10
References		10

# Synopsis

## 1. Introduction

Handwriting Character Recognition is a process that translates images of handwritten scanned text into machine-editable text, or pictures of characters into a standard encoding scheme representing them in ASCII or Unicode. A Handwriting Character Recognition system enable us to feed a book or a magazine article directly into an electronic computer file, and edit the file using a word processor. Though academic research in the field continues, the focus on Handwriting Character Recognition has shifted to implementation of proven techniques. Handwriting Character Recognition (using optical techniques such as mirrors and lenses) and digital character recognition (using scanners and computer algorithms) were originally considered separate fields.

Because very few applications survive that use true optical techniques, the Handwriting Character Recognition term has now been broadened to include digital image processing as well. Early systems required training (the provision of known samples of each character) to read a specific font. "Intelligent" systems with a high degree of recognition accuracy for most fonts are now common. Some systems are even capable of reproducing formatted output that closely approximates the original scanned page including images, columns and other non-textual components. However, this approach is sensitive to the size of the fonts and the font type.

For handwritten input, the task becomes even more formidable. Soft computing has been adopted into the process of character recognition for its ability to create input output mapping with good approximation. The alternative for input/output mapping may be the use of a lookup table that is totally rigid with no room for input variations.

A Handwriting Character Recognition has a variety of commercial and practical applications in reading forms, manuscripts and their archival etc. The project tries to create software for recognition of a handwritten text from photos. It uses computer vision and machine learning. And it experiments with different approaches to the problem.

Such a system facilitates a keyboard less user-computer interaction. Also the text which is either printed or hand-written can be directly transferred to the machine. The challenge of building a Handwriting Character Recognition system that can match the human performance also provides a strong motivation for research in this field.

## **2. Literature Review**

### **2.1 Intelligent Systems for Off-Line Handwritten Character Recognition: A Review**

Handwritten character recognition is always a first research area in the field of pattern recognition and image processing and there is a large need for Optical Character. This paper provides a comparatively review of included works in handwritten character recognition based on soft computing technique. [1]

### **2.2 An Overview of Character Recognition Focused on Off-Line Hand writing**

Character recognition (CR) has been extensively studied in the last half century and progressed to a level sufficient to produce technology driven applications. Nowadays there are growing computational power active the implementation of the present CR methods and produce an increasing demand on many emerging application domains, which require more advanced methodologies.[2]

### **2.3 Image preprocessing for optical character recognition using neural networks**

In this paper forward-feed neural networks are used to processing of text for optical character. Application was developed and its characteristics were set according to results of practical experiments.[3]

### **2.4 Recognition for Handwritten English Letters: A Review**

In this paper we get an overview of research work for recognition of hand written English letters. In Hand written text there is no limitations on the writing style. Hand written letters are difficult to recognize because of different human handwriting style, slant, size and shape of letters.[4]

### **2.5 Improving Offline Handwritten Text Recognition with Hybrid HMM/ANN Models**

In this paper Hybrid Hidden Markov Model (HMM) is used for recognizing offline handwritten texts. In this paper, different techniques are applied to remove slope and slant from handwritten text and to normalize the size of text images with supervised learning methods. The key features of this recognition system were to develop a system having high accuracy in preprocessing and recognition, which are both based on ANNs.[5]

### **3. Problem Statement**

The main problem is the handwriting style of every different people has its own approach to handwriting in different languages .This problem motivated us to build a system that will recognize character (English) given as an input image.

### **4. Objective**

To develop a Handwriting Character Recognition system using the Convolutional Recurrent Neural Networks (CRNN) and its accuracy will be tested using Machine Learning metrics.

Therefore, the following objectives need to be achieved to satisfy the development of the project.

- To study Convolutional Recurrent Neural Network (CRNN) algorithm and develop a system that is able to recognize characters
- To detect, extract and recognize characters using Convolutional Recurrent Neural Network (CRNN).
- To reduce noise from handwritten documents.

### **5. Methodology**

#### **4.1 Proposed Method**

The proposed system when subjected to a scenario of a set of handwritten articles, the characters in the articles are converted to. Handwriting Character Recognition (HCR) is using the stages like preprocessing, segmentation, feature extraction and recognition using neural network. In Preprocessing image document to make use for segmentation. In segmentation the image is segmented into individual character then feature extraction technique is apply on character image.

#### **4.2 System Architecture**

A character recognition system receives an input in the form of image which contains some text information. The output of this system is in electronic format. There are three modules:

(A) pre-processing

(B) text recognition

(C) post-processing.

Each module is further described in detail as bellow:

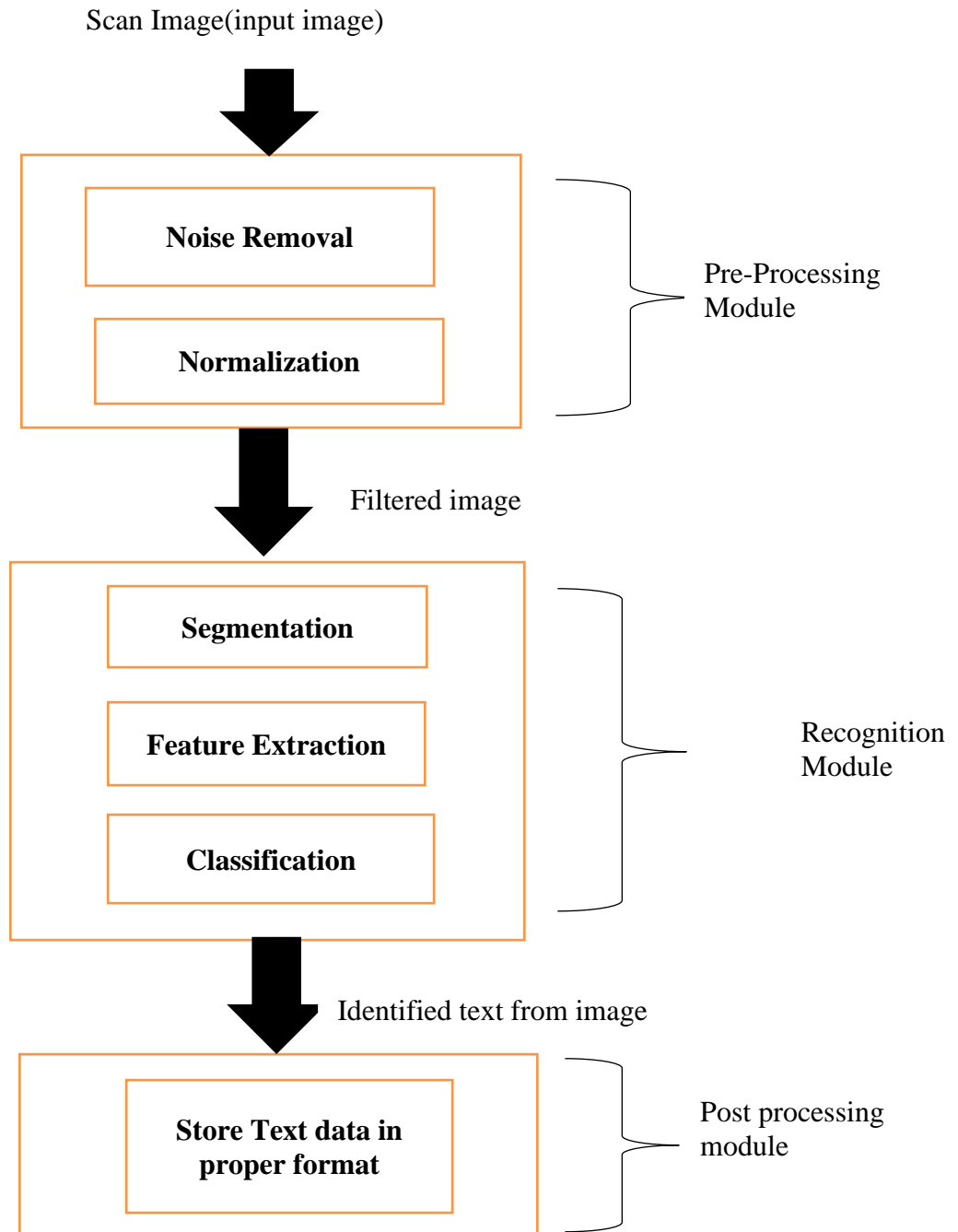


Fig.1. Architecture of character recognition

**(A) Pre-processing Module:** The document is captured by the camera and is converted in the form of a picture. It is the combinations of pixels. At this stage we have the data in the form of image and this image so that's the important information can be retrieved. So to improve quality of the input image, few operation are performed for enhancement of image such as noise removal, normalization, binarization etc.

- **Noise Removal:** Due to this quality of the image will increase and it will affect recognition process for better text recognition in images. And it results in generation of more accurate output at the end of character recognition processing. There are many methods for image noise removal such as mean filter, min-max filter, Gaussian filter etc.
- **Normalization:** The process for which the data need to be organized in the database where range of pixel intensity values changes.
- **Binarization:** A handwritten document is first scanned and is converted into a gray scale image. Gray scale images are converted to binary images by using binarization.

**(B) Recognition Module:** This module can be used for text recognition in output image of pre-processing model and give output data which are in computer understandable form. Hence in this module following techniques are used.

- **Segmentation:** In recognition module, the segmentation is the most important process. Segmentation is done to make the separation between the individual characters of an image. A user can write text in the form of lines. Thus the image is first segmented into line. Then each individual line is segmented into word. Finally each word is segmented into individual character.
- **Feature Extraction:** Feature extraction is the process to separate the most important data from the raw data. There are different classes are made to store the different features of a character. There are many technique used for feature extraction like Principal Component Analysis (PCA), Linear Discriminate Analysis (LDA), Independent Component Analysis (ICA), Chain Code (CC), Gradient Based features, Histogram etc.
- **Classification:** Input to this stage is output of the feature extraction process. The input feature with stored pattern is compared and find out best matching class for input. There are many technique used for classification such as Artificial Neural Network (ANN), Template Matching, Support Vector Matching (SVM) etc.

**(C) Post-processing module:** The output of recognition module is in the form text data which is understand by computer, So there need to store it in to some proper format( i.e. text or MS-Word )for farther use such as editing or searching in that data.



## Block Diagram of The Work

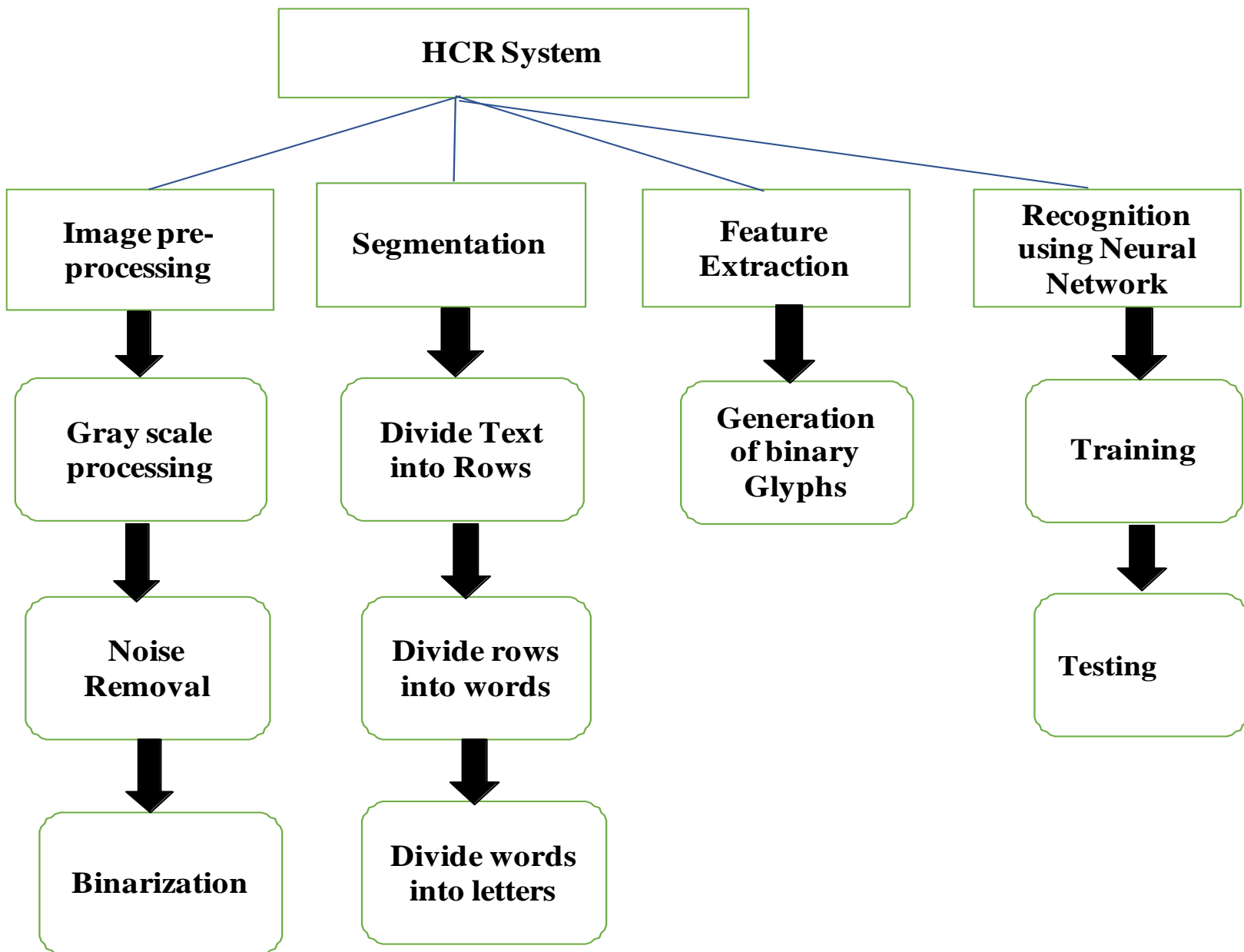


Fig.2. Block diagram of character recognition

## 6. PERT CHART

Our month wise plan to complete the project is as follows –

Date	Expected Work
<b>September 2022</b>	<ul style="list-style-type: none"><li>• Research paper &amp; feasibility understanding of the different methods used in the project.</li><li>• Gather all the requirements like hardware and software tools and modules for the implementation.</li><li>• Also collecting &amp; understanding the required datasets.</li></ul>
<b>October 2022</b>	<ul style="list-style-type: none"><li>• Starting the Implementation part with various pre-processing steps.</li><li>• Arranging the data according to the model requirements.</li><li>• Training &amp; testing of the model.</li></ul>
<b>November 2022</b>	<ul style="list-style-type: none"><li>• Optimizing &amp; tuning the model parameters to increase the model accuracy.</li><li>• Visualization of various approaches for better understanding.</li><li>• Completion of the Project.</li><li>• Preparation of Project Report.</li></ul>

## References

1. Shabana Mehfuz,Gauri katiyar, ‘Intelligent Systems for Off-Line Handwritten Character Recognition: A Review” ,International Journal of Emerging Technology and Advanced Engineering Volume 2, Issue 4, April 2012. Access Date: 09/07/2015.
2. ]Rahul KALA, Harsh VAZIRANI, Anupam SHUKLA and Ritu TIWARI, “An Overview of Character Recognition Focused on Off-Line Handwriting”, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS PART APPLICATIONS AND RE- VIEWS, VOL. 31, NO. 2, MAY 2001. Access Date:09/07/2015
3. ]Miroslav NOHAJ, Rudolf JAKA, “Image preprocessing for optical character recog- nition using neural networks ”Journal of Patter Recognition Research, 2011. Access Date:09/07/2015.
4. ]Nisha Sharma et al, “Recognition for handwritten English letters: A Re- view “International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 7, January 2013. Access Date:09/07/2015
5. Salvador España-Boquera, Maria J. C. B., Jorge G. M. and Francisco Z. M., “Improving Offline Handwritten Text Recognition with Hybrid HMM/ANN Models”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, No. 4, April 2011