Marathwada Shikshan Prasarak Mandal’s

**Deogiri Institute of Engineering and Management Studies,**

**Aurangabad**

**Seminar Report**

**On**

**Handwritten Character Recognition**

Submitted By

**Rijin Reji Thomas (36117)**

**Dr. Babasaheb Ambedkar Technological University**

**Lonere (M.S.)**



Department of Computer Science and Engineering

**Deogiri Institute of Engineering and Management Studies,**

**Aurangabad**

(2019- 2020)

**CERTIFICATE**

This is to certify that the Seminar entitled “**Handwritten Character Recognition**” submitted by **Mr. Rijin Reji Thomas** is a bonafide work completed under my supervision and guidance in partial fulfillment for the award of Bachelor of Technology (Computer Science and Engineering) Degree of Dr. Babasaheb Ambedkar Technological University, Lonere.

Place: Aurangabad

Date:16th October 2019

**Prof. P.Durole Vice Principal & HOD**

**Guide Prof. S.B. Kalyankar**

**Head**

**Dr. Ulhas D. Shiurkar**

**Director,**

**Deogiri Institute of Engineering and Management Studies,**

**Aurangabad**

**Abstract**

The main aim of this project is to design expert system for , “HCR(English) using Neural Network”. that can effectively recognize a particular character of type format using the Artificial Neural Network approach. Neural computing Is comparatively new field, and design components are therefore less well specified than those of other architectures. Neural computers implement data parallelism. Neural computer are operated in way which is completely different from the operation of normal computers. Neural computer are trained (not Programmed) so that given a certain starting state (data input); they either classify the input data into one of the number of classes or cause the original data to evolve in such a way that a certain desirable property is optimized.

Two phase processes are involved in the overall processing of our proposed scheme; the Pre-processing and Neural network based Recognizing tasks. The pre-processing steps handle the manipulations necessary for the preparation of the characters for feeding as input to the neural network system. First, the required character or part of characters needs to be extracted from the pictorial representation. The splitting of alphabets into 25segment grids, scaling the segments so split to a standard size and thinning the resultant character segments to obtain skeletal patterns. The following pre-processing steps may also be required to furnish the recognition process:

I. The alphabet scan be thinned and their skeletons obtained using well-known image processing techniques, before extracting their binary forms.

II. The scanned documents can be “cleaned” and “smoothed” with the help of image processing techniques for better performance.

###### Contents

###### Chapter

###### 

1. **INTRODUCTION** 
   1. Character Recognition

1. **LITERATURE SURVEY**
2. **Brief on System**

General Statement on the problem area

Proposed Recognition System

Pre-Processing Method

Implementation on Graphical User Interface

1. **CONCLUSIONS**

4.1 Conclusion

**REFE.RENCES**

**ACKNOWLEDGEMENT**

1. **Introduction**

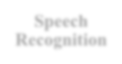
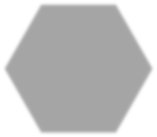
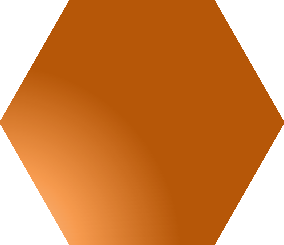
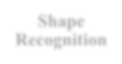
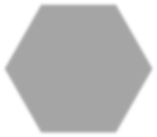
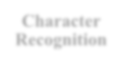
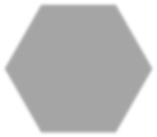
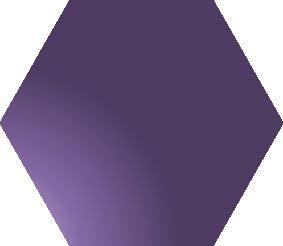
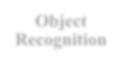
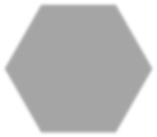
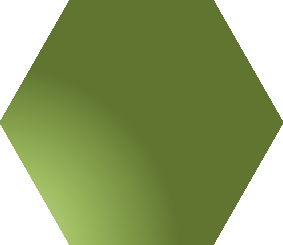
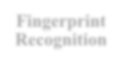
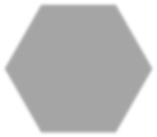
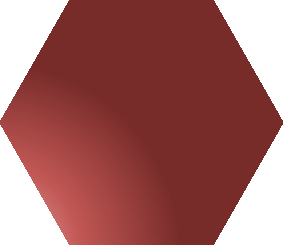
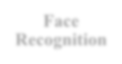
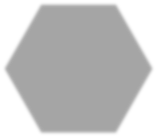
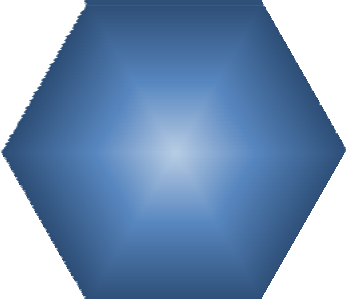
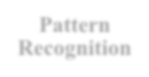
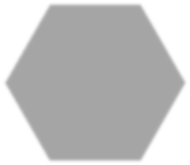
Handwritten character recognition is an area of pattern recognition which defines an ability of a machine to analyze patterns and identify the character. Pattern recognition is the science of making inferences from perceptual data based on either a priori knowledge or on statistical information.

The subject of pattern recognition spans a number of scientific disciplines uniting them in the search for the solution to the common problem of recognizing the members of a class in a set containing elements from many patterns in classes. A pattern class is a category determined by some given common attributes. A pattern is the description of any member of a category representing a pattern class. The basic function of pattern recognition system is to detect and extract common features from the patterns describing the objects that belong to the same pattern classes and to recognize the pattern and classify it as a member of the pattern class under consideration.

Since last few decades and advancement in technology, computers interact more effectively with humans and with the natural world e.g. speech recognition, handwritten recognition, gesture recognition etc. However, Humans are outperforming far better than machines in recognizing patterns. Some of the tasks are generally easy for humans, such as identifying the human voice based on frequency and pitch, recognizing and differentiating aroma of flower and a food, identifying characters etc. These kinds of perceptual problems are difficult for the computer because of voluminous data with composite and hidden information each pattern usually contains.

In a curiosity to understand and uncover secret of how humans can recognize patterns many efforts are being attempted in this area to mimic human behavior. Handwritten character recognition is one such area where “Handwriting” is used to preserve information so as to retrieve it at a later stage and as a facilitating mode of communication using some artificial graphical symbols on a surface. Since many decades paper have been a popular surface to write. Due to the advancement of technology this surface is gradually being taken by a pressure sensitive touch screen. On the basis of surface handwritten character recognition technology is divided into two categories i.e. Online and Offline.

*Figure 1-1 Pattern Recognition Areas*



**Face Recognition**

**Speech Recognition**

**Fingerprint Recognition**

**Pattern Recognition**

**Shape Recognition**

**Object Recognition**

**Character Recognition**

A transition is happening in the field of personal computing to make from the desktop to handheld devices, so it requires change in input paradigm that is suitable for single hand entry than a keyboard. Online handwriting recognition allows for such input modalities. In on-line handwriting recognition, words are generally written on a pressure sensitive surface from which real time information, such as the order of the stroke made by the writer is obtained and preserved as two-dimensional coordinates of successive points.

Computers are able to store large amount of information in a small chip thereby minimizing storage space. However massive data in the world still exist on paper which

includes some important historical documents, books, records etc. Transformation requires to convert textual information available on a paper to store it as electronic format a process known as digitization. Benefits of digitization includes quick searching, minimized storage space, faster retrieval, easy to modify, deducing new information from existing data etc.

There are two general ways for digitization: First is to key in data available on paper manually by human whereas second approach is to use electronic device like scanner. In a former approach where a process of digitization is carried out by humans, it requires a lot of human powers and time for this manually entry of storing information by typing data. Where as in second approach machine is involved in digitization process where a device called scanner will convert a piece of paper into electronic format as image. Comparing these two approaches first approach has a benefit that once data are keyed in further processing operation can be carried out easily as data are in editable format and limitation of having high accuracy, human power and timing. Second approach does digitization faster than first with high accuracy however data stored as images. If a sole purpose is just a preserving then this approach seems better on the other hand images needs to be converted to editable text format using some computer programs for carrying certain operation that involves searching, updating, deleting text, deriving information from text such as charts, graphs etc.

Converting image into editable text requires identifying characters from input image acquired by digitizing paper using scanner and is also known as offline character recognition. Text available on paper may be handwritten or printed. In off-line handwriting recognition no dynamic information is available. Off-line handwriting recognition is the process of finding letters and words that are present in the digital image of a handwritten text. Ability of computer to recognize handwritten character is a challenging area of research as variation in numerous parameter. Character recognition algorithm differs due to the fact that although document is written with single language script following parameter affects a process of recognition as shown in (Figure 1-2). Set of parameters considered here for diversities includes set of alphabets

- a standard set of letters, Direction of writing and nature of script.

Further complexities involved includes the fact that handwriting differs from writer to writer. Every individual human being has an own style of writing. Sometimes even it is difficult for human to identify and read handwritten text of another human. Further handwriting is also affected by situation, environment and also temperament of human while writing a text.



Cursive script, Non cursive script, Character Modifiers, Heading line

Nature of script

Right to Left, Vertical, top to bottom (Chinese, Japanese) [8]

Left to Right, Vertical, top to bottom (Mongolian)[8]

Direction of Writing

Language Script Characteristics

Right to Left, Horizontal (Hebrew, Arabic)[8]

Left to Right, Horizontal (English, Devnagari, Gujarati)[8]

E.g. English (A-Z, a-z) Gujarati (S-7,V-V: )

Set of Alphabets

### Figure 1-2 Language script characteristics and diversities

* 1. **Character Recognition**

Character Recognition in general means identification of printed or handwritten characters or symbols. Character Recognition area is further divided into two subareas

i.e. (1) Handwritten Character Recognition and (2) Machine Printed Character Recognition. Areas and subareas of Character Recognition are shown in (Figure 1-3).

## Machine Printed Character Recognition

If any machine printed document is scanned and is converted to Image, where the intention of preserving that document is solved by digitizing it. But later on to search from that digitized document is not possible as it is saved as image, converting such

document into editable form making it ready to search, index and can be further modified.

Machine Printed Character Recognition tool does really this task of converting characters stored in image to either Unicode or ASCII format making in editable form. Recognition of Machine Printed Characters are little easy as compare to Handwritten Character Recognition as the format of every character throughout document is going to remain same, in another words the shape, style and structure of every character is going to remain same. On the basis of Complexities involved in classifying features and correctly recognizing characters, Machine Printed Character Recognition system can be further classified into two categories i.e. (1) Single Font and (2) Multi font, which are explained in below subsection. Font is really a synonym for typeface.



**Mutiple Script**

**Single Script**

**Document**

**Isolated**

**Character**

**Writer**

**Independent**

**Writer**

**dependent**

**Multiple Font**

**Single Font**

**Offline**

**Online**

**Machine Printed**

**Handwritten**

**Character Recognition**

### Figure 1-3 Character Recognition Sub Areas

* + - 1. **Single Font**

In this category of Machine Printed Character Recognition System, the document is printed using single font and hence this makes it easy to recognize as characteristics of the font is going to remain same.

## Multiple Font

This category of Machine Printed Character Recognition System is little challenging as compare to single font machine printed character recognition as multiple fonts are used within document. Sometimes two fonts may look similar but while carefully examining them using magnifying glass some subtle differences may found among them. Fonts are separated by how they are designed by varying some parameters like weight, slope, width, optical size, font metric comprises of size and space in the font overall including cap height, x-height, ascender height, descended depth, font bounding box etc. If not known in prior that how many fonts are used in document it becomes challenging to correctly identify features and properties for character and correctly classification.

## Handwritten Character Recognition System

In the era of digitization when electronic devices are really becoming smarter it requires simple interface between man and computer. Automatic Handwritten Character Recognition is one such area which has applications in wide variety of areas ranging from form based applications, postal address , Bank cheques, insurance applications, order form etc.

Handwritten Character Recognition is an area in which document /character is written by human being which makes it really dynamic and challenging area to work with it. Handwriting differs from person to person and it is one of the characteristic helpful to identify the person. Sometimes it is very difficult for a person to identify what is written from the handwritten document by other person. This area is more puzzling then the area of multiple font Machine Printed Character Recognition system as writing the same character more diversified ways can be found. Correctly with full accuracy identifying handwritten character recognition is yet a stimulating and fascinating area for researcher. Handwritten Character Recognition System can be divided into (1) online and (2) Offline as explained in subsequent sections.

## Online Handwritten Character Recognition System

In this category of Handwritten Character Recognition System user is allowed to use e- pen or stylus or finger tips which makes it natural and convenient form to input data using various supported handheld devices. In case of online handwritten character

recognition system directional information along with basic characteristics of character can be obtained.

Online handwritten character is convenient form of input for the person even not aware about technological aspects of about operating device, they simply have to use their own style of writing using any of the pointing object.

In online handwriting user's written strokes are captured as they are being formed by sampling the pen's (x,y) coordinates at evenly spaced time intervals, lifting of a pen that is up and down of pointing object makes it dis-ambiguous to classify.

Other than basic properties of character certain other properties such as:

1. Lifting a pointing object is required to write a character
2. Pointing object went from left to right, right to left, top to bottom or bottom to top – directional information.
3. Identifying geometrical properties such as contour detection, line detection, loop detection can be done based on pointing object and coordinates.
4. When user writes on touch sensitive electronic device many real time information such as coordinates, pressure given to write character can be obtained.

In case of online handwritten character recognition system handwritten character written in particular script is classified and stored as Unicode or ASCII format for further processing. Online handwritten character recognition system can be further divided into (1) writer dependent and (2) writer independent character recognition system.

## Writer Dependent

In case of Writer Dependent (WD) online handwritten character recognition system first end user provides some structured input to recognition engine on the basis of which recognition engine is trained and becomes capable of identifying handwriting of that individual user in which user has to invest time for training the system.

## Writer Independent

In case of Writer-independent systems user can start using system without having to worry about underlying recognition engine.

## Offline Handwritten Character Recognition System

In case of Offline Handwritten Character Recognition using a pen writer writes on paper, later that page is scanned to digitize it and stored in form of a image. Considering a scenario where one person sends electronic mail a handwritten scanned digitized copy to another person if the end user who have received that document want to do modification to that document for further processing cannot be done, as image doesn’t really allow text searching, indexing and editing.

Offline Handwritten Character Recognition system really does the task of identifying handwritten character contained in image and converting it into a format which allows recharge, reformat it for further processing.

Off-line handwritten character recognition system doesn’t really require user to have operating knowledge of device as required into online handwritten character recognition system. In a natural way user can write on a piece of paper. While comparing Offline and Online character recognition system, informer the trajectory, dynamic information is not available as with later case.

Off-line handwritten character recognition system can be further divided into two categories on the basis of how much content is stored in it? i.e. (1) Isolated character / symbol containing single character in a image and (2) Document containing group of lines, words and further characters. Both these subsystem are explained in below subsections

## Isolated Character recognition

In this category image contains a single character written in any particular script or symbol which is little easy as compare to document recognition as this system doesn’t really require segmentation of words and sentences as character is in isolation form.

## Document Recognition

Document may contains multiple pages, pages may contain multiple line, further divided into words and characters. Offline handwritten document recognition engine requires them to segment to break apart pages to lines, lines to words and words to character then further processing of feature extraction and classification can be made. Document recognition can be further classified on the basis of language script used within document as (1) Single script & (2) Multi script.

## Single Script

If document is written with any single language script then it is easy to identify features and classify and convert it into Unicode or ASCII equivalent. It is easy as compare to Multi language Script used for document writing as recognition engine has to be programed to identify features of one particular language script.

## Multi Script

If document is written with multiple script or a document is really a mixture of multiple language script then recognition engine has to deal with feature extraction and classification problem with all alphabets used for the document which makes it challenging as compare to single script document recognition.

1. **Literature Survey**

Handwritten character recognition is process of converting the hand written work over page to a attractive digital format. HCR is a intelligent work done throw scanning the images will complete the analysis of character with output. CR require proper handling of complexity of written content, writing environment, materials, etc. HCR techniques are based on extracting various features of handwritten. Isha Vats, Shamandeep Singh[1] In this paper, system was based on recognition of offline handwriting numerals. The main aim of the proposed work in this paper was to efficiently recognize the offline handwritten digits with a higher accuracy. But a difficult problem in this field was the recognition of completely touching handwritten digits and in this paper the proposed system focused on segmentation for isolating the digits so multiple images can be recognize.

Gunjan Singh, Sushma Lehri [2] Handwritten characters was a difficult task because characters are written in various ways, so they could be of different sizes, orientation, thickness and dimension. An offline HCR(Hindi) system using neural network is presented in this paper. Neural networks were good at recognizing handwritten characters as these networks are insensitive to the missing data.A Backpropagation neural network is used for classification. Experimental result of this system shows that results 93%. S S Sayyad, Abhay Jadhav, Manoj Jadhav, Pradip Bele,Smita Miraje, Avinash Pandhare [3] In this paper A neural network approach is proposed for automatic offline character recognition system. In this paper, work has been performed to recognize Devanagari characters using multilayer perceptron.Various patterns of characters were created in the matrix with the use of binary form and stored in the file.This system used the back propagation neural network for efficient recognition and neuron values were transmitted by feed forward method in the neural network.

Shabana Mehfuz, Member IEEE, 2 Gauri katiyar[4] This paper provides review of existing works in HCR based on soft computing technique during the past decade.

Prof. Swapna Borde, Ms. Vinaya Patil, Ms. Ekta Shah, Ms. Priti Rawat [5] This paper presents a fuzzy approach to recognize characters. Fuzzy sets,fuzzy logic were used as bases for representation of fuzzy character and for recognition.Fuzzy-based algorithm which first segments the character and then using fuzzy system gives the characters that match the given input and then using defuzzication system finally recognizes the character. No training is needed by this system for recognition.

Fatos T. Yarman-Vural and Nafiz Arica [6] The rapidly growing computational power enables the implementation of the present Character Recognition methodologies and creates an increasing demand on many emerging application domains, which require more advanced methodologies. The available Character Recognition techniques with their superiorities and weaknesses are reviewed. The Character Recognition is discussed, and directions for future research are suggested. Special attention is given to the off-line HCR since this area requires more research.

1. **Brief on System**

## General Statement of the Problem Area

The main objective of this research is to find a new solution for handwritten text recognition of different fonts and styles by improving the design structure of the traditional Artificial Neural Network (ANN). ANNs have been successfully applied to pattern recognition, association and classification, forecast studies, and control applications, to name a few. The recognition results of such text or handwritten materials are then fed into Optical Character Recognition (OCR) as an electronic translation of images of handwritten, typewritten or printed text into machine-editable text. OCR is a field of research that is fully developed and has been quite useful in pattern recognition, artificial intelligence and machine vision. Consequently, typewritten text recognition that is void of any distortions is now considered largely a solved problem. However the direct use of OCR on handwritten characters remains a very difficult problem to resolve, yielding extremely low reading accuracy. On-line handwritten character recognition involves the automatic conversion of text as it is written on a special digitizer or PDA, where a sensor picks up the pen-tip movements as well as pen-up and pen-down switching. The obtained signal is converted into letter codes, which are usable within computer and text-processing applications. Off-line handwritten document recognition is currently a difficult problem, as different people have different handwriting styles. Scanning, segmentation and classification are the general processes that are being used to recognize handwritten documents. In previous studies, ANNs have proven to be excellent recognizers of printed characters and handwritten digits (0~9), but research into recognition of handwritten words has not always been effective .

## 2. The Proposed Recognition 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. The general schematic diagram of the recognition system is shown in Fig.1.The proposed method which does not include feature extraction stage is shown in Fig1.

Figure 1.Distribution in groups of patients by the stroke type.

## Image acquisition

In 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 preprocessed image suitable for segmentation.

## Segmentation

In the segmentation stage, an image of sequence of characters is decomposed into subimages 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 Page 23762 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 Fig.4.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.

## 3. Pre-Processing Method:

The raw data is subjected to a number of preliminary processing steps to make it usable in the descriptive stages of character analysis. Preprocessing aims to produce data that are easy for the OCR systems to operate accurately. Fig 3.1.1 shows the Pre-processing of handwritten character of image .The main objectives of pre-processing are : • Noise reduction • Binarization • Edge detection • Dilation and fillin • Stroke width normalization • Skew correction • Slant removal • Segmentation of processed image

## Noise Reduction

When the document is scanned, the scanned images might be contaminated by additive noise and these low quality images will affect the next step of document processing. Therefore, a preprocessing step is required to improve the quality of images before sending them to subsequent stages of document processing. Due to the noise there can be the disconnected line segment , large gaps between the lines etc. so it is very essential to remove all of these errors so that’s the information can be retrieved in the best way. There are many kinds of noise in images. One additive noise called “Salt and Pepper Noise”, the black points and white points sprinkled all over an image, typically looks like salt and pepper, which can be found in almost all documents. Noise reduction techniques can be categorized in two major groups as filtering, morphological operations.

## Filtering

It aims to remove noise and diminish spurious points, usually introduced by uneven writing surface and/or poor sampling rate of the data acquisition device. Various spatial and frequency domain filters can be designed for this purpose

## Morphological Operations

Morphological operations are commonly used as a tool in image processing for extracting image components that are useful in the representation and description of region shape. Morphological operations can be successfully used to remove the noise on the document images due to low quality of paper and ink, as well as erratic hand movement.

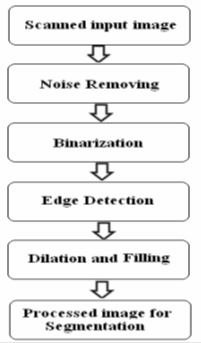


Figure 1.Distribution in groups of patients by the stroke type.

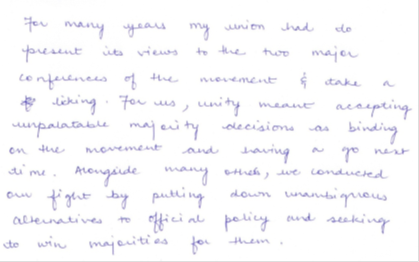


Figure 1.Distribution in groups of patients by the stroke type.

## Binarization

Binarization of gray-scale character images is a crucial step in offline character recognition. Good binarization facilitates segmentation and recognition of characters. Binarization process converts a gray scale image into a binary image. In this it has described new methods for the binarization of noisy gray-scale character images obtained in an industrial setting. Our methods are specially designed to binarize gray-scale character images more effectively by using the fact that characters are usually composed of thin lines of uniform width. Experimental results show that these methods give the best binarization results.

Figure 1.Distribution in groups of patients by the stroke type.

## 

## Edge Detection

Edges characterize object boundaries and are therefore useful for segmentation, registration, and identification of objects. Edge detecting an image significantly reduces the amount of data and filters out useless information, while preserving the important structural properties in an image. There are many ways to perform edge detection. However, the majority of different methods may be grouped into two categories, gradient and Laplacian. The gradient method detects the edges by looking for the maximum and minimum in the first derivative of the image. The Laplacian method searches for zero crossings in the second derivative of the image to find edges.

## Thresholding

In order to reduce storage requirements and to increase processing speed, it is often desirable to represent grey scale or color images as binary images by picking some threshold value for everything above that value is set to 1 and everything below is set to 0. Two categories of thresholding exist: Global and Adaptive. Global thresholding picks one threshold value for the entire document image, often based on an estimation of the background level from the intensity histogram of the image. Adaptive thresholding is a method used for images in which different regions of the image may require different threshold values. A comparison of many common thresholding techniques is given by using an evaluation criterion that is goal-directed in the sense that the accuracies of a character recognition system using different techniques were compared. On those Tested, Niblack’s method produced the best result.

## Dilation and fillin a. Skew Detection

For a document scanning process, there can be the skewness. There are several commonly used methods for detecting skew in a page; some rely on detecting connected components and finding the average angles connecting their centroids. The skewness should be removed because it reduces the accuracy of the document. The skew angle is calculated and with the help skew angle, the skewed lines are made horizontal.

## E.SEGMENTATION

In Character Recognition techniques, the Segmentation is the most important process. Segmentation is done to make the separation between the individual characters of an image. Segmentation of unconstrained handwritten word into different zones (upper middle and lower) and characters is more difficult than that of printed documents. This is mainly because of variability in inter-character distance, skew, slant, size and curved like handwriting. Sometimes components of two consecutive characters may be touched or overlapped and this situation complicates the segmentation task greatly. In Indian languages such touching or overlapping occurs frequently because of modified characters of upper-zone and lowerzone. Segmentation is an important stage, because the extent one can reach in separation of words, lines or characters directly affects the recognition rate of the script.

There are two types of segmentation:

## a) External Segmentation

External segmentation decomposes the page layout into its logical units. External segmentation is the isolation of various writing units, such as paragraphs, sentences or words. It is the most critical part of document analysis. Document Analysis and Recognition (DAR) aims at the automatic extraction of information presented on paper and initially addressed to human comprehension. Segmenting the document image into text and non-text regions is an integral part of the OCR software. Therefore, one who works in the Page 23765 CR field should have a general overview for document analysis techniques. Page segmentation is one important step in layout analysis and is particularly difficult when dealing with complex layouts. Page layout analysis is accomplished in two stages: The first stage is the structural analysis, which is concerned with the segmentation of the image into blocks of document components (paragraph, row, word, etc). The second one is the functional analysis, which uses location, size and various layout rules to label the functional content of document components. Page segmentation is then implemented by finding textured regions in gray-scale or color images. For example, a method for automatically evaluating the quality of document page segmentation algorithms is introduced. They have proposed a bitmap-level automatic scheme to benchmark page segmentation algorithms on mixed text/halftone documents. It provides an accurate qualitative diagnosis of segmentation techniques, from which, a quantitative evaluation is derived.

## b) Internal Segmentation

Internal Segmentation is an operation that seeks to decompose an image of a sequence of characters into sub images of individual symbols. Although, the methods have developed remarkably in the last decade and a variety of techniques have emerged, segmentation of cursive script into letters is still an unsolved problem. Character segmentation strategies are divided into three categories .

## 4. Implementation On Graphical User Interface

A user-friendly front end interface as shown in Fig.7 and Fig.8 has been implemented for the proposed handwritten character recognition system using menu based GUI (Graphical User Interface). The interface system presents the user with two menus - first menu with five processing stages and the second menu to choose the type feature extraction.

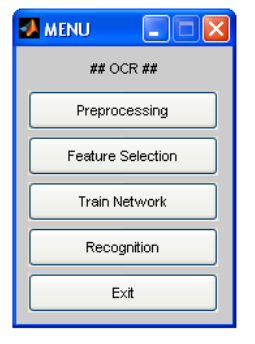


Figure 1.Distribution in groups of patients by the stroke type.

## Based Graphical User Interface

The menu based GUI enables the user to perform pre-processing, select the type of feature extraction, perform the feature extraction using the chosen method and train the network. After the network is trained, the recognition of the test image can be initiated by clicking the recognition bar on the interface. The test image is chosen using the facility provided for selecting the test images. Upon the completion of recognition process, the recognized image appears on the notepad. The same procedure can be used to recognize any number of test images. Finally the exit bar is used to quit from the character recognition system after recognizing all the test images. The GUI frees the user from the difficulties of working from the command line interface.

### 3. Conclusions

A neural network based off line handwritten character recognition system without feature extraction has been introduced in this paper for classifying and recognizing the 26 English alphabets. The pixel values derived from the resized characters of the segmentation stage have been directly used for training the neural network. As a result, the proposed system will be less complex compared to the offline methods using feature extraction techniques. Of the several neural networks architectures used for classifying the characters, the one with two hidden layers each having 100 neurons has been found to yield the highest recognition accuracy of 90.19%. The handwritten recognition system described in this paper will find potential applications in handwritten name recognition, document reading, conversion of any handwritten document into structural text form and postal address recognition.

## References

1. Mori S, Suen CY, Yamamoto K. Historical review of OCR research and development 1992-jul;:1029-1058. [CrossRef](https://doi.org/10.1109/5.156468) | [Google Scholar](https://scholar.google.co.in/scholar?hl=en&q=Historical%20review%20of%20OCR%20research%20and%20development)|
2. Arica N, Yarman-Vural FT. An overview of character recognition focused on off-line handwriting 2001-may;:216-233. [CrossRef](https://doi.org/10.1109/5326.941845) | [Google Scholar](https://scholar.google.co.in/scholar?hl=en&q=An%20overview%20of%20character%20recognition%20focused%20on%20off-line%20handwriting)|
3. Bhattacharya U, Chaudhuri BB. Handwritten Numeral Databases of Indian Scripts and Multistage Recognition of Mixed Numerals 2009-mar;:444-457.
4. Pal U, Sharma N, Wakabayashi T, Kimura F. Handwritten Numeral Recognition of Six Popular Indian Scripts 2007.

**ACKNOWLEDGEMENT**

I would like to place on record my deep sense of gratitude to **Prof. Sanjay Kalyankar**, HOD-Dept. of Computer Science and Engineering, Deogiri Institute of Engineering and Management Studies Aurangabad for his generous guidance, help and useful suggestions.

I express my sincere gratitude to **Prof. S.G.Shah**, Dept. of Computer Science and Engineering, Deogiri Institute of Engineering and Management Studies Aurangabad, for her stimulating guidance, continuous encouragement and supervision throughout the course of the present work.

I am extremely thankful to **Dr. Ulhas Shiurkar**, Director, Deogiri Institute of Engineering and Management Studies Aurangabad, for providing me infrastructural facilities to work in, without which this work would not have been possible.

**Signature of Student**

Rijin Reji Thomas (36117)