

# Survey on Handwritten Character Recognition using Artificial Neural Network

**Cinu George**  
UG scholar

Department of Electronics and Communication Engineering  
Christian College Of Engineering And Technology,  
Oddanchatram, India

**S.Podhumani**  
UG scholar

Department of Electronics and Communication Engineering  
Christian College Of Engineering And Technology,  
Oddanchatram, India

**Reena K Monish**  
UG scholar

Department of Electronics and Communication Engineering  
Christian College Of Engineering And Technology,  
Oddanchatram, India

**T.Pavithra**  
UG scholar

Department of Electronics and Communication Engineering  
Christian College Of Engineering And Technology,  
Oddanchatram, India

## Abstract

Character recognition has defined a lot of attention in the field of pattern recognition due to its various applications. Many researches can be carried out for online characters. An offline handwritten alphabetical characters recognition system using artificial neural network (ANN) is described in this paper. Here handwritten character recognition is done for languages like Tamil, Malayalam and English. The recognition rate is varied for different languages. In this paper the techniques like auto encoder support vector machine is used for extracting the feature of the handwritten alphabets. Handwritten Character Recognition (HCR) is useful in bank-cheque processing, signature verification, handwritten postal adders resolution and many more. In coming days, character recognition system might serve as a key function to create paperless environment by digitizing and processing existing paper documents.

**Keywords:** Handwritten character recognition, artificial neural network, Auto encoder, Restricted Boltzmann machine

## I. INTRODUCTION

Handwritten character recognition is the ability of a computer to receive and interpret intelligible handwritten input from sources such as paper documents, photographs, touch screen and other devices. The handwritten character recognition can be either on-line or off-line. On-line handwriting recognition involves the automatic conversion of text as it is written in special digitizer, where a sensor picks up the pen-tip movements. Off-line handwritten recognition involves automatic conversion of text in an image into letter codes.

Tamil is a Dravidian language predominantly spoken by the Tamil people of India. Tamil scripts have 12 vowels, 18 consonants and one special character. The complete script consists of 31 letters in their independent form, and an additional 216 combination letters representing a total 247 combinations.

Malayalam is a language spoken in India, predominantly in the state of Kerala. It is the family of Dravidian language. Malayalam script has 15 vowel letters, 41 consonants and a few other symbols.

English is a global language. Modern English alphabets are Latin alphabets. It is an official language of almost 60 sovereign states. It is the third most common native language in the world. English script has 26 letters. It has 6 vowels and 21 consonants. It is principal language spoken in Britain, the USA, Canada, Australia, New Zealand and some other countries such as Uganda and Botswana. It is the language of international commerce, business, diplomacy and tourism.

Character Recognition

Mainly the five major stages in the handwritten character recognition are image acquisition, pre-processing, segmentation, feature extraction, classification and recognition.

Image acquisition is the first step in the work flow sequence where the documents are generated. Images for character recognition is acquired by capturing photographs of the handwritten characters using webcam or by scanning the handwritten document. Scanned image is converted to JPEG format. The below figure shows an example of scanned input image of a few tamil characters.

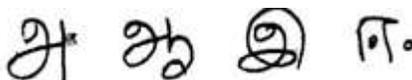


Fig. 1: Scanned Input Image

Preprocessing step commonly involves removing low frequency background noise. In the first phase in our character recognition process is converting the image to binary image by thresholding the given character image. Two intensity values are available in the binary images these values are black and white. We are using zero for black and one for white thus the color of the character is white and background is black.

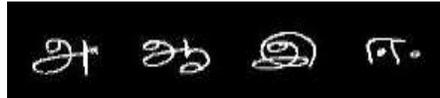


Fig. 2: Preprocessed Image

Segmentation is ... Image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super-pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics.



Fig. 3: Segmented images

After the segmentation process feature extraction is done. Feature extraction is a special form of dimensionality reduction. In machine learning, pattern recognition and in image processing, feature extraction start from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant. Feature extraction is done by various techniques like auto-encoder, support vector machine, Restricted Boltzmann Machine etc.

The wavelet transform have many beneficial features so they provide an appropriate basis for image handling. It has various merits like it has the ability of signal energy compaction and converts it into few transformation coefficients. Classification and Recognition: Support Vector Machine, Self Organizing Map, RCS, Fuzzy Neural Network and Radial Basis Network.

## II. MOTIVATION FOR SURVEY

A great difficulty was faced in the past few years to digitize historic paper based documents like newspapers, books, information material of any field. The purpose behind digitization is to preserve the document, make them easily and fully accessible, searchable and process able in digital form []. Since every human has a different style of writing and bears a different character model in mind, handwriting recognition is a complex problem. So it is necessary to convert such hardcopies into a standard encoding scheme or to soft copies.

## III. VARIOUS TECHNIQUES

Handwritten Tamil Character recognition using wavelet is proposed by Dr.Amitabh et.al. [2014] [2]. In this paper the recognition of basic handwritten Tamil characters is done by Optical Character Recognition (OCR) system which can handle different font size and font types. Here wavelet is used for feature extraction purpose of various Tamil handwritten characters and neural classifiers have been used for the classification. The preprocessing step involves Binarization and noise reduction.

Binarization: It is a method of transforming a grey scale image into a black and white image through Thresholding .Otsu's method is the another approach used to perform histogram based thresholding

Noise removal: it median filtering, Wiener filtering and other morphological operation can be performed to remove noise.

In this paper the extracted features such as character height, width, horizontal and vertical lines, slope lines etc

Discrete wavelet transforms (DWT)

Mainly the transform levels of wavelet ranges from 1-5 .in order to reduce the image size and proceed to process wavelet decomposition is done.

In Continuous Wavelet Transform (CWT), the wavelets are not orthogonal and the data obtained by this transform are highly correlated. The four values of the decomposition are,

- CA – Accurate Value
- CD – Dimension value
- CV – Vertical Value
- CH – Horizontal Value

CA, CD, CH, CV are the values that act as feature extraction values. Normalization method is essential to achieve scale invariant and translation invariant. By transferring the image into center the translation invariant can be achieved.[1]

In [2] C.Sureshkumar et al have discussed about Tamil character recognition and conversion using neural network. The handwritten Tamil character is converted into Unicode Tamil character. Using spatial space detection technique the scanned

image is segmented into paragraphs. This paragraphs are further divided into lines using vertical histogram technique and these lines are divided into words using horizontal histogram technique. Support vector machine (SVM), RCS, Radial basis network, Fuzzy neural network, Self organizing map are used for the recognition of extracted features. Supervised learning algorithm is used for character classification.

*Support vector machine*

S.V.Rajashekaradhya et.al. [2009] [5] In this paper a zone based feature extraction algorithm is proposed. The recognition and classification of a numeral image are done with the help of nearest neighbor classifier (NNC), feed forward back propagation neural network (BPNN) and support vector machine (SVM) and this method can be tested for different languages like Kannada, Tamil, Telugu and Malayalam.

ICZVDD refers to the zone column distance feature in the vertical downward direction.

ICZHRD refers to zone row distance feature in the horizontal right direction. Both with reference to the image centroid.

Classification is done by Nearest Neighbor Classifier and Feed forward Back propagation neural network classifier.

Nearest neighbor classifier is a long employed approach and is used for large scale pattern matching. Feature vectors and class labels of the training samples are stored during the training phase.

ANN has been extensively used for both the recognition of Indian as well as non-Indian digits. Subsequent recognition and classification of numerical images are done with the help of feed forward back propagation neural network.

The back propagation neural network recognition performance will largely depend on the training algorithm and structure of the network. It consists of one input layer, one hidden layer and one output layer. Network performance will be deeply influenced by the number of hidden nodes.

J.Pradeep et.al. [2011] [4]The authors of this paper make use of Diagonal based feature extraction for the off-line handwritten English alphabet characters recognition system. The most important factor in achieving high recognition performance is probably based on selection of the appropriate feature extraction method. Off-line handwritten character recognition is done with the help of diagonal feature extraction. By moving along the diagonals the features are extracted from the pixels of each zone. In order to perform classification and recognition task the above features are used to train feed forward back propagation neural network.

In this paper [4] for classification purpose the feed forward back propagation neural network consisting of two hidden layer and with the architecture of 54-100-100-38. The hidden layers use log sigmoid activation function. The feature vector the output layer is referred as a competitive layer and one of the character is used to be identified. The feature vector is denoted as X where  $X = (f_1, f_2, \dots, f_d)$ . Here d denotes the number of zones in which each character is divided and f denotes features. Known data set is used for training neural network. Depending upon number of features used the number of input nodes will be selected. After training the network by using number of unknown datasets the recognition system was tested. The diagonal method of feature extraction is verified using a number of test images.

Mansi Shah et.al [2013] [6]In this paper a supervised learning technique can be utilized by the back propagation model or multilayer perceptron. There are number of parameters that regulates the performance of neural network, like

- Number of epochs
- Number of layers
- Learning algorithm
- Number of neurons consisting in each layer
- Transfer function used between multiple layers.

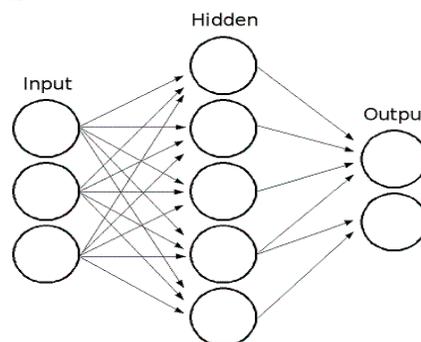


Fig. 2: Architecture Of Multilayer Neural Network

[6]Application in wear and manufacturing process: Properties like learning and non linear behavior of an ANN make them useful to model complex non-linear process. From the experimental data we can learn to improve the performance and reliability of mechanical equipment and manufacturing tools require accurate modeling and prediction of non-linear phenomena it can be done with the help of ANN.

Two main function of ANN are more beneficial

- 1) Firstly, the continuous approximation of a multivariable function.
- 2) Secondly, classification is used for recognition of the operation.

In order to reduce errors between the computed output and desired output back propagation algorithm is used and it aims at minimizing the learning time.

Kohonen Self Organization Neural Network: The output unit in the network arranged in two dimensional plane with weight between each unit and inputs. When the input vector is fed to the network only one output unit which has the best matching Weight is selected.

Adarsh Sarkar and Kritika Singh [7] in their report two neural network algorithms are used Deep auto encoders using the idea of restricted Boltzmann machine and back propagation using gradient descent method. After training the network for each of the classifier the accuracy of classification was calculated and estimated on the test set. To train and test the classifiers, datasets of more than 20000 Hindi numerals are used. Here work done by this project can be divided into two parts. Building the data base of Hindi numerals as there does not exist any standard database like MINIST for Hindi numerals is one of the major part of this report. Secondly identifying features and classification of algorithm which give efficient result for our data set.

In preprocessing step we converted the output image in RGB to grey scale. Using an average filter noise removal is done in this gray scale. Per-training, unrolling, fine tuning are the stages used here for testing. From this project we can see that chain code histogram have better feature-set than the projection histogram feature set. Only by pre trained with RBM and use of deep auto encoder network classifier on the data set, we can achieve 98.2% accuracy.

#### IV. APPLICATIONS

Handwritten Character Recognition finds various applications. To digitize historic paper based documents like books, newspapers, information material of any field, So that the documents can be preserved, make them fully and easily accessible, searchable and process able in digital form.

There are various business applications also like security control OCR systems etc. By using this recognition method a pre-trained system can identify any sort of damaged images.

It can also be used in Banks. Handwritten bank cheques are processed manually every day in developing countries. In such a verification written information including date, signature, legal and courtesy amounts present on each cheque has to be visually verified.

Every Sector in this world wants a system which is it intelligent to solve any problem according to the input.

#### V. CONCLUSION

This is detailed discussion about handwritten character recognition techniques and includes its use in different area by using Artificial Neural Network. From study of various papers we have seen that selection of relevant technique plays an important role in performance of character recognition rate. This material serves as a helpful guide and update for readers working in the Character Recognition area.

By studying artificial Neural Network we had concluded that as per as technology is developing day by day the need of Artificial Intelligence is increasing because of only parallel processing.

Parallel Processing is more needed in this present time because with the help of parallel processing only we can save more and more time and money in any work related to computers and robots. If we talk about the Future work we can only say that we have to develop much more algorithms and other problem solving techniques so that we can remove the limitations of the Artificial Neural Network.

#### REFERENCES

- [1] Dr.Amitabh wahi, Mr.Sundaramurthy.S, Poovizhi.P, "Recognition of Handwritten Tamil Characters using Wavelet", International Journal Computer Science and Engineering Technology (IJCSET), ISSN:2229-3345,Vol.5, No.04, 2014.
- [2] C.sureshkumar and Dr.T.Ravichandran, "Handwritten Tamil Character Recognition and Conversion using Neural Network", International Journal Computer Science and Engineering Technology (IJCSET), ISSN:0975-3397,Vol.02,No.7,2010.
- [3] J.Pradeep, E.Srinivasan and S.Himavathi, "Diagonal Based Feature Extraction for Handwritten Alphabets Recognition System using Neural Network", International Journal Computer Science and Engineering Technology (IJCSET), Vol.3, No.1, 2011.
- [4] S.V.Rajashekararadhya and Dr.P.Vanaja Ranjan, "Handwritten Neural/Mixed Numerals recognition of South Indian Scripts, The zone based Feature Extraction method", Journal of theoretical applied Information Technology(JATIT), Vol.7,No.1,pp 063-079,2009
- [5] Mansi Shah and Gordhan B Jethava, "A Literature Review on Handwritten Character Recognition", Indian Streams Research Journal, ISSUE-2,ISSN:2230-7850,Vol.3,2013.
- [6] Minodara Ripa and Laurentia frangu, "A survey of artificial neural network application in wear and manufacturing processes",The ANNALS of university "DUN AREA DE JOS of JALATIFASEICLE VIII,2004, ISSN 1221-4590.
- [7] Akarshan Sarkar and Kritika Singh, "Handwritten Hindi numeral recognition system".