iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** IV **Month of publication:** April 2017

DOI: http://doi.org/10.22214/ijraset.2017.4151

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 5 Issue IV, April 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

A Survey on Handwritten Devanagari Character Recognition using Neural Network

Prof. S. A. Dongare[1], Prof. N. J. Khapale[2], Prof. S. S. Dawange[3]
[1]*Computer Engg.* , [2]*E&TC*, [3]*Electrical Department, Pune University*

Abstract: Handwritten character can be converted in to the digital information using handwritten character Recognition, which is the ability of a computer to receive and interpret handwritten input from documents. Handwritten Devanagari Characters are more complex for recognition due to presence of header line, conjunct characters and similarity in shapes of multiple characters. For Handwritten Devanagari Character recognition using neural network various approaches has been proposed. In general the process involves phases as: Scanning, Preprocessing, Feature Extraction and Recognition. Preprocessing includes noise reduction, binarization, normalization and thinning. Feature extraction includes extracting some useful information out of the thinned image in the form of a feature vector. Artificial neural network is used for classification. In this survey, comparative study of various approaches has been presented.
Keywords*: Classification, feature extraction, handwritten character recognition, neural network, preprocessing.

I. INTRODUCTION

Information gathering is one of the most important things in the digital world, many organizations still use handwritten document for storing their information. So it is important task to convert this information in the digital world. This process helps user to convert their handwritten document in to the digital information. Handwritten character recognition is the ability of a computer to receive and interpret intelligible handwritten input from sources such as paper documents, photographs, touch-screens and other devices. Handwritten Marathi Characters are more complex for recognition than corresponding English characters due to many possible variations in order, number, direction and shape of the constituent strokes.

Important steps in any HCR are preprocessing, segmentation, feature extraction and recognition using neural network. Preprocessing includes series of operations to be carried out on document image to make it ready for segmentation. During segmentation the document image is segmented into individual character or numeric image. Feature extraction technique is applied on character image. Finally feature vector is presented to the neural network for recognition. Neural network can be used as a backend for performing recognition process. Success of handwritten Devanagari character recognition using neural network depends on size of database, if size of database used to train neural network is larger than probability of successful recognition is more. But larger databases put limitation on speed of recognition process.

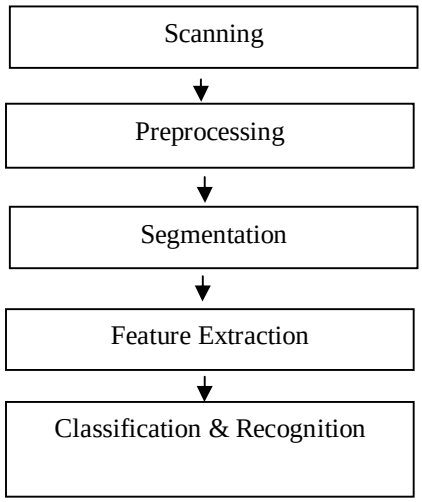


Fig. 1 Phases Involved in Handwritten Character Recognition

Volume 5 Issue IV, April 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

A. Image Preprocessing

The pre-processing is a series of operations performed on the scanned input image. It essentially enhances the image rendering it suitable for segmentation. Preprocessing includes noise reduction, binarization, normalization and thinning.

B. Segmentation

In the segmentation stage, an image of sequence of characters is decomposed into sub-images of individual character. In this system, the pre-processed input image is segmented into isolated characters. Each individual character is uniformly resized into n x n pixels for classification and recognition stage.

C. Feature Extraction

In feature extraction useful features are extracted to form feature vector . This feature vector fed to neural network for recognition.

D. Recognition of Character using Neural Network

An artificial neural network as the backend is used for performing classification and recognition tasks. In the offline recognition system, the neural networks have emerged as the fast and reliable tool for classification towards achieving high recognition accuracy [8].

II. DEVANAGARI SCRIPT

A. Properties of Devanagari Script

Devanagari script has features different from other languages. Devanagari character set has 13 vowels, 36 consonants and 10 numerals with optional modifier symbols. Characters are organized into three zones as upper, middle and lower zone. Core characters are positioned in middle zone, while optional modifiers in upper and lower zones. Two characters may be connected to each other. In Devanagari script, the concept of uppercase and lowercase characters is absent. Fig. 2 represents Devanagari character set. It represents Devanagari character modifier set. Modifiers are optional symbols arranged in upper and lower zones.

B. Issues Regarding Recognition of Devanagari Script

Some reasons that cause recognition of Devanagari characters difficult are as:

1) In Devanagari Script individual characters are connected by header line (Shirorekha) which makes segmentation of individual character is quite difficult.
2) Characters may be connected to form conjuncts for which separation is complex.
3) Presence of modifiers makes segmentation difficult.
4) Some Devanagari characters are similar in shape.

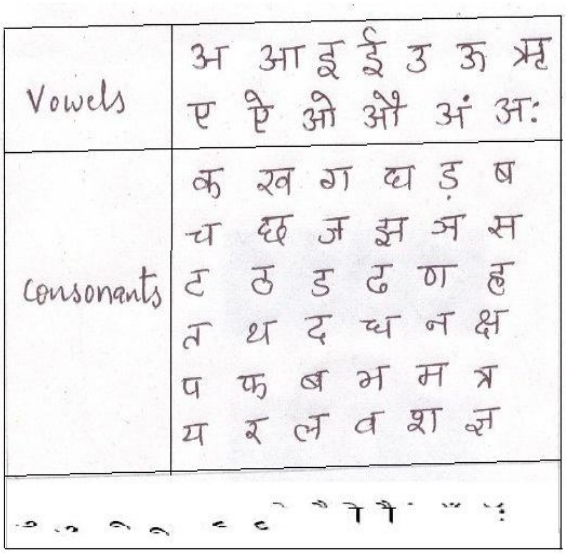


Fig. 2 Devanagari Script

III. LITERATURE REVIEW

Handwritten character recognition is an intelligent optical character recognition capable of handling the complexity of writing, writing environment, materials, etc. Many feasible techniques have been developed in this field. Techniques are based on extracting various features of handwritten characters and providing these features as input to an Artificial Neural Network. Handwriting recognition is classified into two types as offline and online handwriting recognition methods. In offline recognition, the writing is usually captured optically by a scanner and complete writing is available as an image. But in the on-line system two dimensional coordinates of successive points are represented as a function of time and the order of strokes made by the writer are also available. The offline methods have been shown to be superior to their online counterparts in recognizing handwritten characters due to the temporal information available with the former. Several applications including mail sorting, bank processing, document reading and postal address recognition require online handwriting recognition systems. As the result, the offline handwriting recognition continues to be an active area for research towards exploring the newer techniques that would improve recognition accuracy.

K. Y. Rajput and S. Mishra have proposed a system for recognizing handwritten Indian Devanagari script. In this paper, a system for recognizing handwritten Indian Devanagari script is presented. The system considers handwritten images as an input, separates the lines, words and then characters step by step and then recognizes the character using artificial neural network approach, in which creating a Character Matrix and a corresponding suitable network Structure is key. In addition, knowledge of how one is Deriving the Input from a Character Matrix must first be obtained. Feed Forward Algorithm gives insight into the entire working of a neural network; followed by the Back Propagation Algorithm which compromises Training, Calculation of Error, and Modifying Weights. Once the characters are recognized they can be replaced by the standard fonts.

In feature extraction character matrix as an array of black and white pixels of size 30X30 is prepared. Afterward, the Feed Forward neural network with back propagation is used in learning and recognition process. Disadvantage with this system is size of matrix is large, 900 no of input neurons are required.

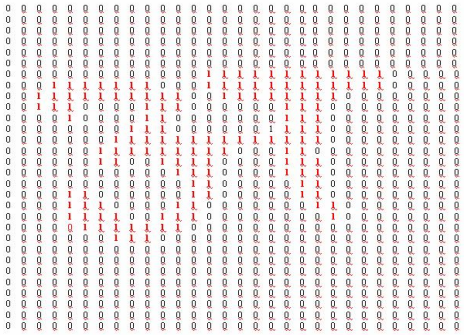


Fig. 3 Matrix size of 30X30 for Character Recognition

The network receives the 900 Boolean values as a 900 element input vector. It is then required to identify the letter by responding with a 49-element output vector. The 49 elements of the output vector each of them represent a letter. To operate correctly, the network should respond with a 1 in the position of the letter being presented to the network. All other values in the output vector should be 0. In addition, the network should be able to handle the noise. In practice, the network does not receive a perfect Boolean vector as input. Specifically, the network should make least possible mistakes while classifying vectors with noise of mean 0 and standard deviation of 0.2 or less. However, the success of this method lies totally on the size of database, i.e. larger the size of database used for training the neural network higher is the probability of successful recognition. Large data base places limit on the speed of recognition, and hence this method can be used only for off line recognition [1]**.**

S. Arora, D. Bhattacharjee [2] proposed two stage classification approach for handwritten Devanagari characters. The first stage is using structural properties like detection of shirorekha, spine in character and second stage exploits some intersection features of characters which are presented to a feed forward neural network. Each handwritten character can be adequately represented within 16 segments (each of size25 X 25 pixels) and hence 32 features for each character can be used as an input to neural network. In this paper differential distance based technique for identifying the shirorekha and spine has been used , as the simple histogram based technique does not identify the straight line for handwritten Devanagari character. Priority based search technique is used to identify the valid movement for shirorekha. A priority mask is moved from rightmost pixel of character, as the movement can go only in directions shown in mask for shirorekha. After identifying shirorekha and spine preliminary classification is done on basis of

Volume 5 Issue IV, April 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

features. For various grouping of characters intersection features are calculated and feature vector is fed to a back propagation feed-forward neural network.

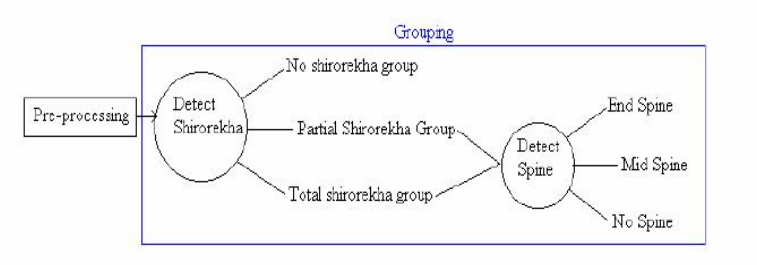


Fig. 4 Preliminary Classification

For each group characters, feature vector of size 32, consisting of intersection points and end points value in each segment is calculated and NN is designed using this input. A 3-layer feed forward architecture with back propagation is considered with one input layer of 32 neurons, one hidden layer, and one output layer. This system has been tested for different hidden layer neurons but maximum accuracy is achieved by 40 neurons in hidden layer. The Scaled conjugate gradient training method is applied. The training parameters are as: learning rate 0.01 momentum factor 0.95 and minimum gradient .00000001. The number of hidden layer neurons has been changed to get best results. This approach has been tested for 50000 samples with maximum accuracy of 89.12%. Complexity of this system lies in two stages first is detection of shirorekha and spine, then selection of feature vector as intersection points [2].

D. Singh, S. Singh and Dr. M. Dutta proposed twelve directional feature inputs depending upon the gradients. This technique can recognize all types of handwritten characters even special characters in any language. Features extracted from handwritten characters are directions of pixels with respect to their neighboring pixels. These inputs are given to a back propagation neural network with one hidden layer and one output layer. This approach provides better results in terms of recognition accuracy, training time and classification time.

In this paper the features of directions of pixels of the characters with respect to their neighboring pixels are extracted and given as input to the neural network. A back propagation feed-forward neural network is used to recognize the handwritten characters. After training the network with back-propagation learning algorithm, high recognition accuracy can be achieved. The direction has been divided into 12 regions with each region covering angle of 30 degree, hence direction value of any pixel may have only 12 values assigned from 1 to 12. This approach increases the information content and gives better recognition rate with reduced recognition time.

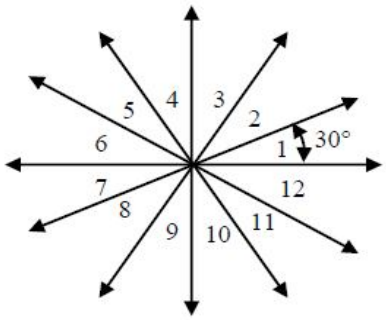


Fig. 5 Orientation of Direction

The topology of the network is 1024 input nodes, 36 hidden nodes and 5 output nodes. Since character is normalized into 32x32 pixel images and corresponding to each pixel directional value is as input, hence total numbers of inputs are 1024 to the neural network. The hidden layer has 32 neurons which are connected to all 1024 inputs. The output layer has 5 nodes since the system has been tested on five types of handwritten characters: three Hindi characters, one English character and one special character. Learning is implemented using the back-propagation algorithm with learning rate 0.2. Gradient is calculated after every iteration and compared with threshold gradient value 10-10. If gradient is greater than the threshold value then it performs next iteration. The batch steepest descent training function is trained. The weights and biases are updated in the direction of the negative gradient of the

Volume 5 Issue IV, April 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

performance function [4].

In [5] development of zone based method which is combination of image centroid zone and zone centroid zone of individual character or numerical image has been proposed. In feature extraction using zone based approach individual character or numerical image is divided into n equal sized zones then average distance of all pixels with respect to image centroid or zone centroid is computed. In combination of image centroid and zone centroid approach it computes average distance of all pixels present in each zone with respect to image centroid as well as zone centroid which gives feature vector of size 2xn features. This feature vector is presented to feed forward neural network for recognition.

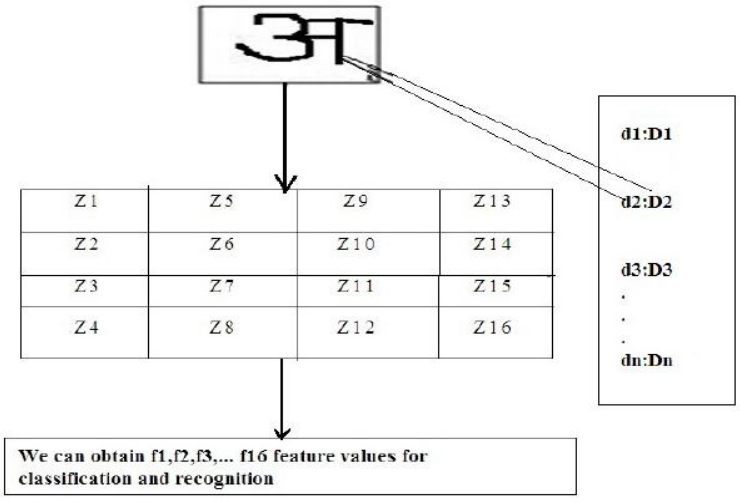


Fig. 6 Feature Extraction from Devanagari Numeral Image “six”

Here feed forward neural network with back propagation algorithm has been used for training and recognition. Architecture of neural network consists of 32 input neurons, 41 output neurons. 1 No of Hidden layer with 4 nodes, Error Back Propagation as training algorithm and Mean Square Error as performance function [5].

In [7] approach used to recognize Hindi characters, feature extraction includes extracting some useful information out of the thinned image in the form of a feature vector. The feature vector comprises of pixel values of normalized character image. A Back propagation neural network is used for classification. Experimental result shows the average accuracy of recognition for this approach is 93%.

Another approach used for recognition of Hindi character. In this work, three feature extraction techniques as histogram of projection based on mean distance, histogram of projection based on pixel value, and vertical zero crossing, have been used to improve the rate of recognition. These feature extraction techniques are powerful to extract features of even distorted characters or symbols. With this approach up to 90% correct recognition rate is achieved [10].

IV. CONCLUSION

Development of handwritten Devanagari Character Recognition is a challenging task in Pattern recognition area. It has been found that recognition of handwritten Devanagari characters is quite difficult task due to variations in writing styles, font size, color and thickness, similarity in shape, presence of modifiers, shirorekha and conjunct characters. Accuracy of the system depends on configuration of neural network, training algorithm. Success of handwritten Devanagari character recognition using neural network depends on size of database, if size of database used to train neural network is larger than probability of successful recognition is more. But larger databases put limitation on speed of recognition process. This system needs to be tested on a wider variety of images containing characters in diverse fonts and sizes.

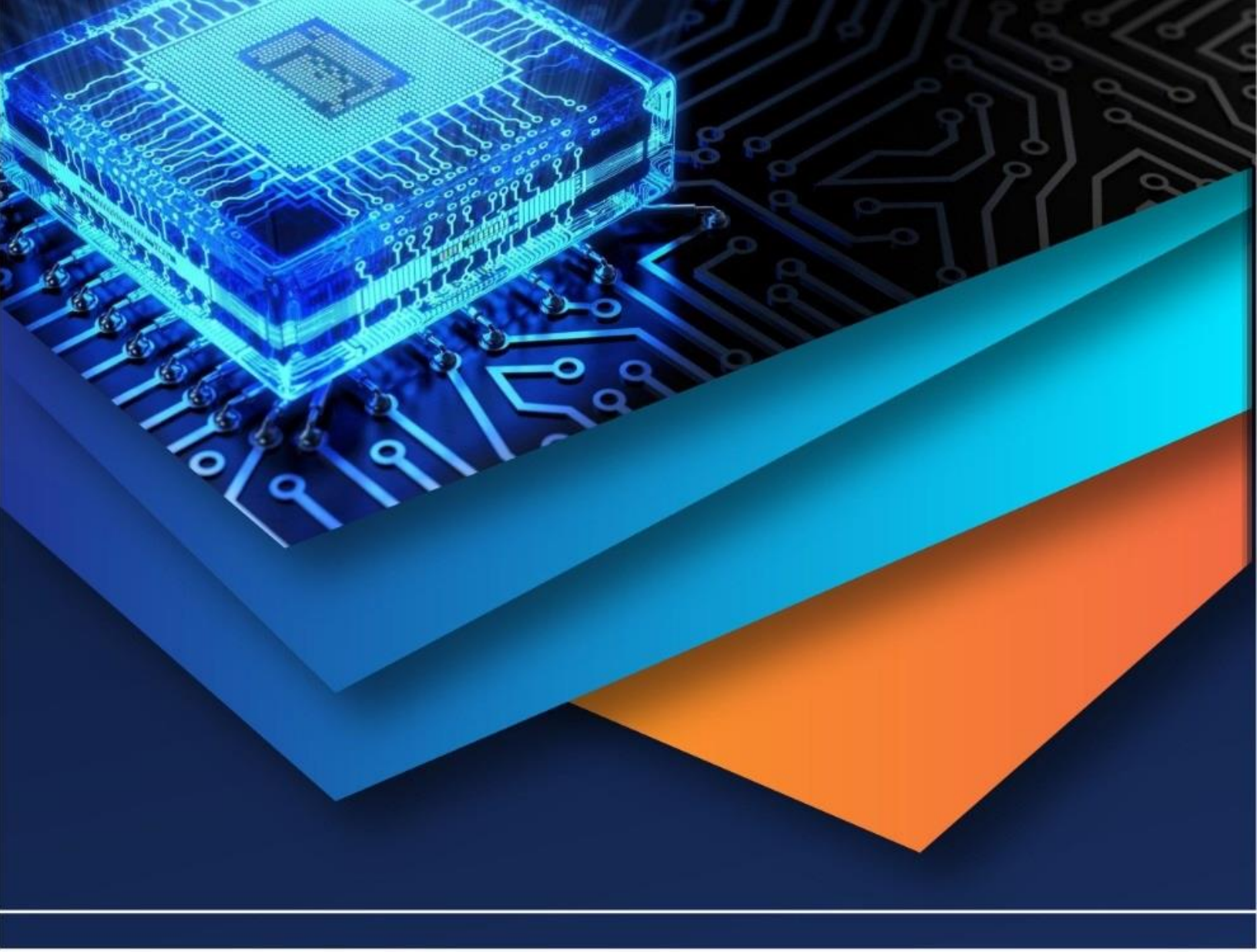
V. ACKNOWLEDGEMENT

First and the foremost I , express my deep sense of gratitude, sincere thanks to Head of Computer Engineering Dept. Prof. D. B. Kshirsagar, SRES' Sanjivani College of Engineering, Kopargaon. I am also thankful to all the faculty members, Computer Engineering Department, SRES' Sanjivani College of Engineering, Kopargaon for giving comments for improvement of work, encouragement and help during completion of this survey. Last but not the least, I should say thanks from my bottom of heart to my

Family and Friends for their never ending love, help, and support in so many ways through all this time.

REFERENCES

[1] K. Y. Rajput ,S. Mishra, “Recognition and Editing of Devnagari Handwriting Using Neural Network", Proceedings of SPIT-IEEE Colloquium and International Conference, Vol. 1.M.
[2] S. Arora, D. Bhattacharjee, M. Nasipuri, L. Malik and B. Portier, “Two Stage Classi_cation Approach for Handwritten Devanagari Characters".
[3] V. Agnihotri, “Offline Handwritten Devanagari Script Recognition", IJITCS,2012, 8, 37-42.
[4] D Singh, S. Singh and Dr. M. Dutta, “Handwritten Character Recognition Using Twelve Directional Feature Input and Neural Network", International Journal of Computer Applications, 0975-8887, Vol. 1.
[5] S. A. Dongare, D. B. Kshirsagar, S. V. Waghchaure,” Handwritten Devanagari Character Recognition using Neural Network “,IOSR, 2278-8727,Volume 16, Issue 2, Ver. X (Mar-Apr. 2014), PP 74-79
[6] N. Sharma, U. Pal, F. Kimura and S. Pal, “Recognition of Offline Handwritten Devanagari Characters Using Quadratic Classifier".
[7] G. Singh, S. Lehri, “Recognition of Handwritten Hindi Characters using Back-propagation Neural Network", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (4) , 2012, 4892-4895.
[8] Mitrakshi B. Patil, Vaibhav Narawade, “Recognition of Handwritten Devnagari Characters through Segmentation and Artificial neural networks", International Journal of Engineering Research Technology (IJERT), ISSN: 2278-0181, Vol. 1,Issue 6, August 2012.
[9] G. Sinha, Mrs. R. Rani, Prof. R. Dhir , “Recognition Using Zonal Based Feature Extraction Method and SVM Classifier", International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 2, Issue 6, June 2012.
[10] D. Yadav, S. Cuadrado and J. Morato, “Optical Character Recognition for Hindi Language Using a Neural-network Approach", J Inf Process Syst, Vol.9, No.1, March 2013.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY