# Recognition of Handwritten Hindi Characters using Backpropagation Neural Network

Gunjan Singh[1], Sushma Lehri[2]
*[1] Faculty of Management & Computer Applications,*
*R.B.S. College, Agra*
*[2] Institute of Computer & Information Sciences,*
*Dr. B.R. Ambedkar University, Agra*

*Abstract:* **Automatic recognition of handwritten characters is a difficult task because characters are written in various curved & cursive ways, so they could be of different sizes, orientation, thickness, format and dimension. An offline handwritten Hindi character recognition system using neural network is presented in this paper. Neural networks are good at recognizing handwritten characters as these networks are insensitive to the missing data. The paper proposes the approach to recognize Hindi characters in four stages— 1) Scanning, 2) Preprocessing, 3) Feature Extraction and, 4) 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. The feature vector comprises of pixels values of normalized character image. A Backpropagation neural network is used for classification. Experimental result shows that this approach provides better results as compared to other techniques in terms of recognition accuracy, training time and classification time. The average accuracy of recognition of the system is 93%.**
*Keywords: Neural Networks, Binarization, Normalization, Thinning, Feature Extraction.*

## I. INTRODUCTION

Handwritten character recognition is an important area in image processing and pattern recognition field. It is a wide field that covers all sort of character recognition via machine in various application domains. The goal of this area of pattern recognition is to translate human readable characters to machine readable characters. Today, we have automatic character recognizers that help humans in variety of practical and commercial applications [1]. A lot of research work has been done in this area, but still there is some space in the state of the art. Handwritten characters are non-uniform in nature, as a particular character can be written in different styles and sizes by different writers and even the same writer can write the same character in different styles at different times. Handwritten characters are also vague in nature as there may not be smooth curves or perfectly straight lines all the time [13]. It also increases the complexity of the system. Handwritten character recognition systems can be of two types-- online character recognition system and offline character recognition systems. On-line recognition systems utilize the digitizers which directly capture writing with the order of the strokes, speed, pen- up and pen- down information. In the offline character recognition system, the image of character is converted into bit pattern by an

optically digitizing device such as optical scanner or camera. The recognition is done on this bit pattern data. It is different from online character recognition in the sense because recognizer allows the previously written and printed text to be processed and recognized, while the online recognizer works only on real time data. The on-line methods have proved to be superior to off-line methods due to the temporal information available with the former [2] [3]. In this paper, we propose a handwritten Hindi character recognition method that uses a Back-propagation neural net to recognize the characters. This paper is organized in six sections: Section 2 gives introduction and features of Hindi language. Section 3 discusses the state of the art in the area of handwritten Hindi character recognition. Section 4 presents the proposed recognition system. In Section 5, we present the experimental results. Section 6 is devoted to conclusions and future work.

## II. HINDI LANGUAGE: A REVIEW

Hindi is an Indo-Aryan language and is one the official languages of India. It is the world's third most commonly used language after Chinese and English and has approximately 500 million speakers all over the world. It is written in Devnagari script. It is written from left to right along a horizontal line. The basic character set has 13 'SWARS' (vowel) and 33 'VYANJANS' (consonants) shown in the figure.
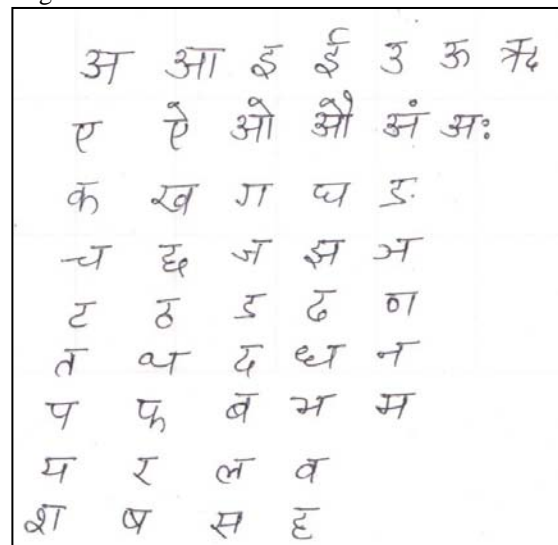


Figure 1: Hindi language basic character set

All the characters have the individual line at the upper part known as 'SHIRO-REKHA' or header line. In addition to the basic character set, it also has modifiers (figure 2). When the vowels are written by using a variety of diacritical marks which are written above, below, before and after the consonant they belong to, they are known as 'MATRAS' (modifiers). Modifiers can be classified as top and lower modifiers. Top modifiers are placed above the 'SHIROREKHA' and lower modifiers are placed below the Hindi character itself. Modifiers can also be placed to the left, to the right of the character or a combination of these.


Figure 2: Vowels and Corresponding Modifiers

Characters may also have a half form. A half form character is required to touch the character and these characters together make a composite character.

### III. RELATED WORK

OCR work on handwritten Hindi characters started early in 1977, when Sethi et al. [4] presented a system for handwritten Devnagari characters. Kumar et al. [5] proposed a Zernike moment feature based approach using ANN and achieved 80% accuracy. A two stage classification technique using ANN and minimum edit distance method was proposed by Arora et al. [6] and obtained 90.74% recognition rate. Two sets of characters were used in each stage. A system using multilayer perceptron network and Radial Basis function network was presented in [7]. A fuzzy model based recognition system has been proposed in [8] with accuracy of 90.65%. Sharma et al. [9] presented a system using quadratic classifier for Devnagari characters. Till date, there is no complete OCR for handwritten Hindi characters that gives 100% success rate.

### IV. PROPOSED RECOGNITION SYSTEM

The proposed Hindi character recognition system consists of 4 stages – scanning, pre-processing, feature extraction and recognition.
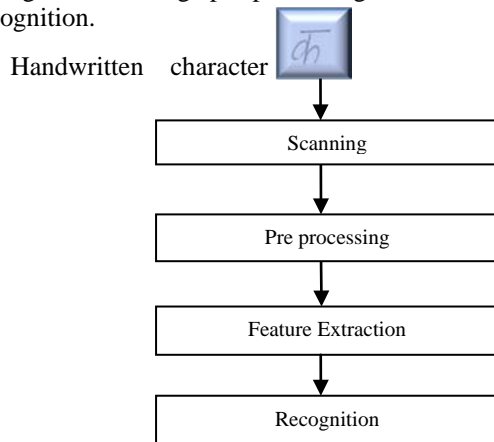


Figure 3: Flow diagram of the proposed system

### IV-A. Scanning

Handwritten character data samples are acquired on paper from various people. These data samples are then scanned from paper through an optically digitizing device such as optical scanner or camera. A flat-bed scanner is used at 300dpi which converts the data on the paper being scanned into a bitmap image.

### IV-B. Preprocessing

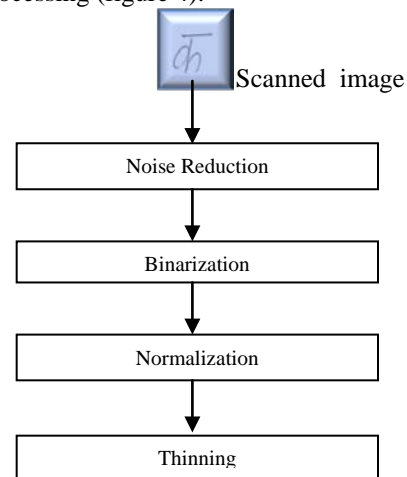A series of operations are performed on scanned image during preprocessing (figure 4).



Figure 4: Preprocessing of the handwritten character

The operations that are performed during preprocessing are:

(i) Applied median filtering to reduce noise from the introduced to the character image during scanning. It is usually taken from a template centered on the point of interest. To perform median filtering at a point values of the pixel and its neighbors are sorted into order based upon their gray levels and their median is determined [12].

(ii) Global thresholding is applied to convert image from gray scale to binary form.

(iii) Image is normalized into 7X7.

(iv) Thinning is performed by the method proposed in [10].

### IV-C. Feature Extraction

Transform the thinned image of size 7x7 into a one-dimensional 49x1 vector form. This vector is to be fed to a backpropagation neural network for training and recognition purpose.

### IV-D. Recognition Using Backpropagation Neural Net

In the proposed system backpropagation neural network is used for classification. A backpropagation neural net is a multi layer feed forward network trained with extend gradient-descent based delta- learning rule or backpropagation learning rule [11] [14]. A multilayer neural network consists of an input layer, an output layer and one or more hidden layer. Hidden layer/s lies between input layer and output layer. There are two or more layers of connection

weights. Information flows in forward direction only. Backpropagation learning rule minimizes the total squared error of the output computed by the net.

The back-propagation algorithm works as follows —

(i) All the weights are initialized to some small random values.
(ii) Input vector and desired output vectors are presented to the net.
(iii) Each input unit receives an input signal and transmits this signal to each hidden unit.
(iv) Each hidden unit calculates the activation function and sends the signal to each output unit.
(v) Actual output is calculated. Each output unit compares the actual output with the desired output to determine the associated error with that unit.
(vi) Weights are adjusted to minimize the error.

In this paper, the proposed backpropagation neural net is designed two hidden layers as shown in figure 7. The input layer contained total 49 nodes as 49 features were extracted for each character. The output layer contained total 5 nodes (1 node for each class). Size of the output was 5X5, where each character is represented as a 5X1 output vector. Number of nodes in both hidden layers was set to 7.
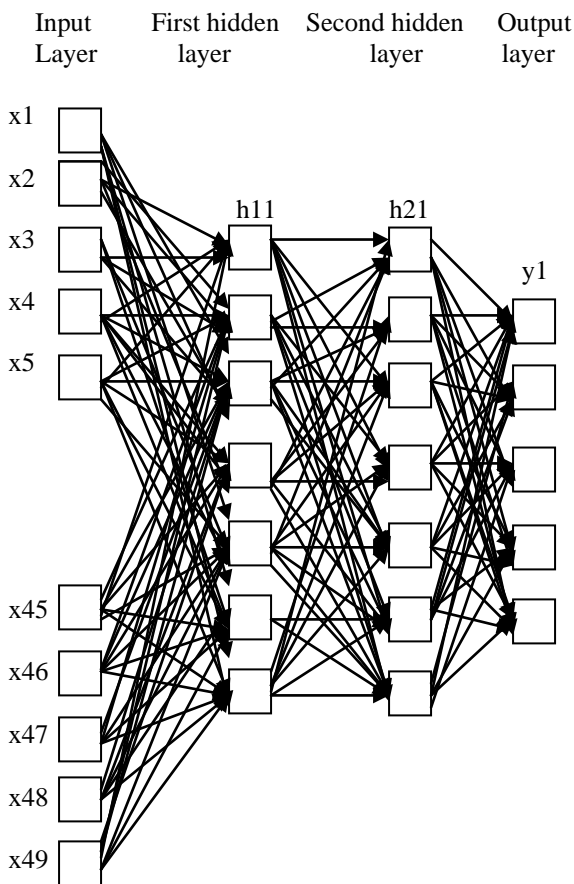


Figure 5: Backpropagation network used for the proposed system

The backpropagation neural network was created using MATLAB 2009a and the following parameters were used to train the net:

- Transfer function used for 1st layer: *tansig*
- Transfer function used for 2nd layer: *purelin*
- Training function : *trainlm*
- Maximum number of epochs: 1000
- Performance function: Mean squared error
- Mu : .00100
- Max Mu : 1.00e+10
- Error goal : .004
- Max. no. of validation checks : 6

Training stops when one of the following conditions occurs:

(i) The maximum numbers of epochs is reached.
(ii) Performance is minimized to the goal.
(iii) Mu exceeds the maximum value.
(iv) Number of validation checks exceeds the maximum value.

## V. EXPERIMENTAL RESULTS

Experiment has been carried out on first 5 'VYANJANS' (consonants) of Hindi character set. 1000 handwritten samples (200 samples for each) written by five different people were used as the dataset. One such character set sample is shown in the following figure.
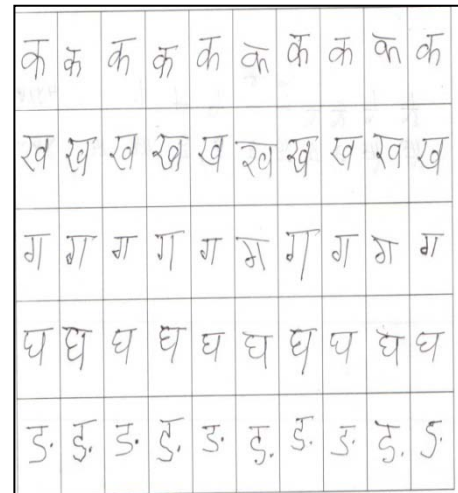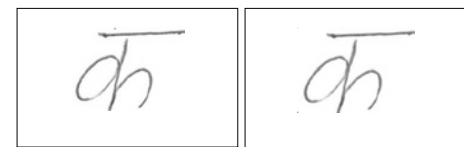


Figure 6: A part of the character sample taken for experiment

Results of operations performed on scanned image of character 'd' during preprocessing stage are as shown in figure 7—



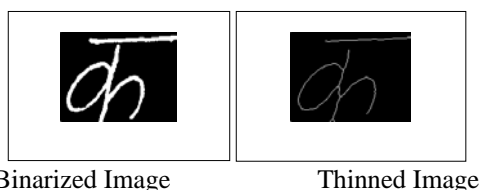Original Image               Filtered image

Binarized Image      Thinned Image

Figure 7: Preprocessing operations applied on character 'd'

Results of feature extraction stage on the normalized image of character 'd' are:



Normalized Image Character      7X7 matrix of pixels (1 – pixel, 0 – Background pixel)

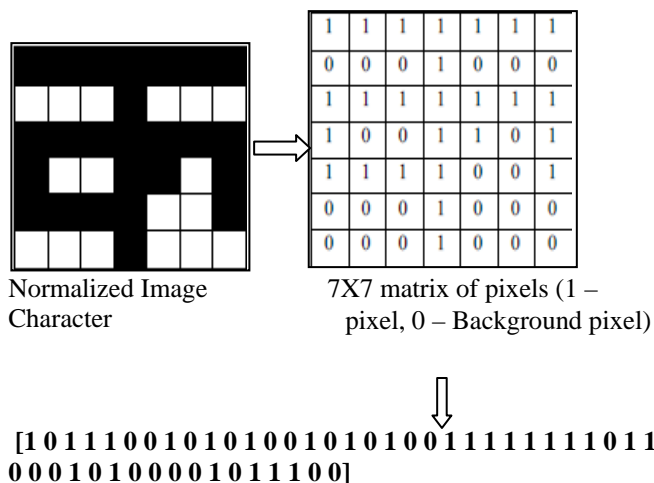**[1 0 1 1 1 0 0 1 0 1 0 1 0 0 1 0 1 0 1 0 0 1 1 1 1 1 1 1 1 0 1 1 0 0 0 1 0 1 0 0 0 0 1 0 1 1 1 0 0]**

Figure 8: Result of feature extraction algorithm on character 'd'

This process is repeated for all the five characters. Each character is entered as an input vector of 49X1. Total 1000 (200 for each character) samples were used for classification. Out of these 1000 samples 60 % were used for training, 20% for validation and rest 20% were used for testing.

Results of recognition stage are shown in the following Table I.

| Character | Output code | No. of epochs | Recognition accuracy |
|---|---|---|---|
|  | **10000** | **840** | **95%** |
|  | **01000** | **920** | **93%** |
|  | **00100** | **840** | **96%** |
|  | **00010** | **972** | **90%** |
|  | **00001** | **862** | **91%** |
| **Total recognition accuracy** | | | **93%** |

Table I: Recognition accuracy of characters

## VI. CONCLUSION

In this paper, we have presented a system for handwritten Hindi character recognition. Features are extracted by counting the number of character pixels and background pixels of the normalized character image. Experimental results show that back-propagation network yields recognition accuracy of 93%. Presented method takes an edge over other methods as there is no need for skew correction and header line removal. Further, the two-hidden layers in backpropagation neural network are able to recognize any nonlinear pattern with great accuracy. This is one of the strength of the proposed method. In future, we are going to include other features like wavelets to increase the recognition rate further.

## REFERENCES

1. Govindan V.K. and Shivprasad A.P., *"Character Recognition: A Review "*, Pattern Recognition, vol. 23, no. 7, pp. 671-683, 1990.
2. Plamondon R. and Srihari S.N., *"On-Line and Off- Line Handwritten Character Recognition: A Comprehensive Survey,"* IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 22, no. 1, pp. 63-84, 2000.
3. Arica N. and Yarman F., *"An Overview of Character Recognition Focused on Off-line Handwriting"*, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 31(2), pp. 216 – 233, 2001.
4. Sethi I.K. and Chatterjee B., *"Machine Recognition of Constrained Hand printed Devnagari"*, Pattern Recognition, vol.9, pp. 69-75, 1977.
5. Kumar S. and Singh C., *"A Study of Zernike Moments and its use in Devnagari Handwritten Character Recognition"*, Proc. International Conference on Cognition and Recognition, pp. 514-520, 2005.
6. Arora S., Bhattacharjee D., Nasipuri M, Basu D.K. and Kundu M., *"Combining Multiple Feature Extraction Techniques for Handwritten Devnagari Character Recognition"*, Proc. IEEE Region 10 Colloquium and Third Intl. Conf. Industrial & Information Systems, Kharagpur, 2008.
7. Verma B.K., *"Handwritten Hindi Character Recognition using Multilayer Perceptron and Radial Basis Function in Neural Networks"*, IEEE International conference on Neural Networks, vol. 4, pp. 2111-2115, 1995.
8. Hanmandlu M., Murthy O.V.R. and Madasu V.K., *"Fuzzy Model Based Recognition of Handwritten Hindi Characters"*, Proc. Ninth Biennial Conf. Australian Pattern Recognition Society on Digital Image Computing Techniques & Applications, Glenelg (Australia), pp. 454–461, 2007.
9. Sharma N., Pal U, Kimura F. and Pal S., *"Recognition of Off-Line Handwritten Devnagari Characters using Quadratic Classifier"*, Proc. of Indian Conference on Computer Vision Graphics and Image Processing (ICVGIP), India, pp. 805 - 816, 2006.
10. Pokhriyal A. and Lehri S., *" MERIT : Minutiae Extraction Using Rotation Invariant Thinning"*, International Journal of Engineering Science & Technology, vol. 2(7), 3225-3235, 2010.
11. Yegnanarana B., *" Artificial Neural Networks"*, Prentice Hall India, 2004.
12. Gonzales R.C. and Woods R.E., *"Digital Image Processing"*, Second Edition, Prentice Hall, 2008.
13. Bunke H. and Wang P.S. P., *"Hand Book of Character Recognition and Document Image Analysis"*, World Scientific, 1997.
14. Fausett L.V., *"Fundamentals of Neural Networks: Architectures, Algorithms and Applications"*, Pearson Education, 2004.