Offline Handwritten Character Recognition using Neural Network

Pankaj Rani
M.Tech Student
ECE Department
BGIET, Sangrur

Sushil Kakkar
Assistant Professor
ECE Department
BGIET, Sangrur

Shweta Rani
Associate Professor
ECE Department
BGIET, Sangrur

ABSTRACT
The offline character recognition is very useful software in the field of research. Authoritative field of research has made by character recognition due to its need in various fields of research as in banks, post offices to fulfill all recognition requirements. This paper is an exploration on the different scripts including Mathematical digits, Hindi consonants and vowels, Gurumukhi characters as well as numerals. Different handwritten samples of all these language scripts are taken from different persons in a good environment. First of all pre-processing including different operations for the reduction of noise is operated. The pre-processed documents are then segmented for the feature extraction. In this technique, different features like lines, corners etc. are to be extracted. Number of hidden layers taken are 15. Then neural network is trained for the testing of all these handwritten samples. Out of 100% samples, 90% are used for training, 5% are for testing and the remaining of the samples are used for validation. Comparison of recognition rates achieved on different scripts is accessible. Out of all samples taken, highest accuracy is achieved on the Gurumukhi script which is 99.9%.

Keywords
Handwritten Recognition, Pattern Recognition, Neural Network and Feature Extraction

1. INTRODUCTION
OCR is the way to convert the handwritten data into machine editable form. The character recognition is further divided into offline and online type of recognition system [1-3]. In online character recognition the live recognition of the data is done and two dimensional coordinates of point sequences of the writing as a function of time are available while in the offline OCR, only the completed writing is available as an image. So it is difficult to implement offline recognition system than online recognition system. The character recognition may be handwritten or printed. The handwritten character recognition is more difficult than that of printed recognition by virtue of individual styles of writing, speed of writing, size of letters, physical and mental condition of the writer, overlap of letters etc. [5]

Except all these, there are some physical devices which also affect the recognition rate as acquisition devices, pen width, pen ink color etc.

P M Kamble et al in [5] proposed a Handwritten Marathi basic character recognition using statistical method. In their work the eccentricity, orientation and mass of character feature and minimum distance was calculated and used for classification. R Verma et al [7], proposed a neural network and a surface feature extraction technology for character recognition enhancement, back-propagation neural network and surface feature technology have been used. The reduction in noise and image quality improvements were exhibited the proposed technique using neural network. In [9] Deshmukh et al. proposed a recognition system for handwritten Devanagari Compound Character based on Legendre moment feature descriptor. The proposed system was trained and tested on 27000 handwritten samples collected from different people. For classification they have used Artificial Neural Network and the overall recognition rate for basic was up to 98.25% and for all compound character is 98.36%. Gupta A. at el [12] focused especially on offline recognition of handwritten English words by first detecting individual characters. The holistic approach was used in recognition of limited size vocabulary where global features extracted from the entire word image were considered. It was found that as the size of the terminology increased, the complexity of holistic based algorithms also increased and correspondingly the recognition rate decreased rapidly. P Kaur et al [14], proposed a recognition system of handwritten numerals of Gurumukhi script, they considered handwritten samples of 10 digits from different persons.

2. OCR PROCEDURE AND ALGORITHM USED
The different stages that are elaborated in the process of Handwritten Character Recognition are shown by the following figure 1 [6]. Different techniques are convoluted in the character recognition for the complete recognition of any handwritten document as shown in the flow chart.

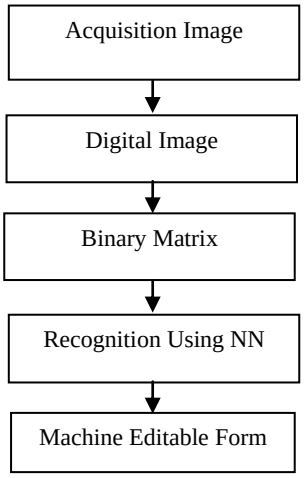

Fig 1: Conceptual model of HCR

The procedure for the recognition process is shown in the figure 1 step by step. Character Recognition may be achieved by capturing the image of the handwritten character using a simple web camera and then processing it for recognition. The intelligent problems are solved out by neural network which

cannot be solved by mathematical formulae or algorithms [6,7]. Neural networks have remarkable ability to derive meaning from complicated or imprecise data therefor can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques [7-9]. The main objective of neural network is to develop a computational device for modeling the human brain so that the various computational tasks are performed faster than traditional systems [7]. This is much helpful in segmentation, in which the input image is segmented into individual characters and then each character is recognized into 10*5 pixels towards the training network [8].

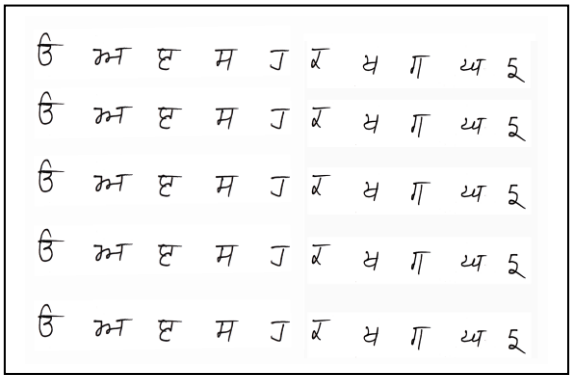

Fig 2: Resized image of Gurumukhi Characters

In figure 2, the original scanned image is resized in the format of matrix of 10*5.

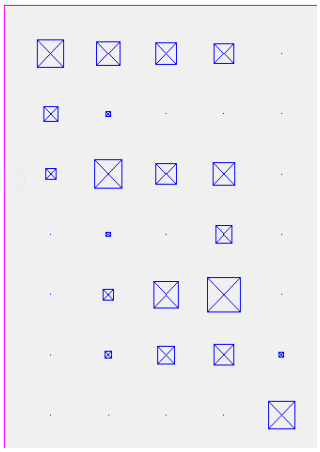

Fig 3: Extracted features of nayia

As shown in the above figure, the features of the character naiya are extracted into different lines and corners.

After resizing the images, pre-processing including different for the reduction of noise and for other basic operations is applied to the image. In this technique, the variations within the classes are minimized and the variations between classes maximized by the extraction of information from the raw data. This type is the most relevant for the classification purposes. To select a good feature extraction method is probably the single most important factor in achieving high recognition performance in HCR [11]. Different features like lines, corners, writing style of a person, etc. are extracted. Feature extraction is based on character geometry. Different line types that form a particular character are to be extracted. It also concentrates the positional features of the same. To extract the features of these different scripts neural network training tool is used. The direction and magnitude of maximum variation in intensity with change in given variable especially per unit distance in a specified direction is measured by calculating the gradient [17].

$$g(x, y) = \sqrt{g_x^2(x, y) + g_y^2(x, y)} \qquad (1)$$

$$g_x(x, y) = h*i\ (x, y)$$

$$g_x\ (x, y) = h^t*i\ (x, y)$$

Where, (x,y) is the cropped single character and h and h^t are the mask which calculate gradient of g_x and g_y.

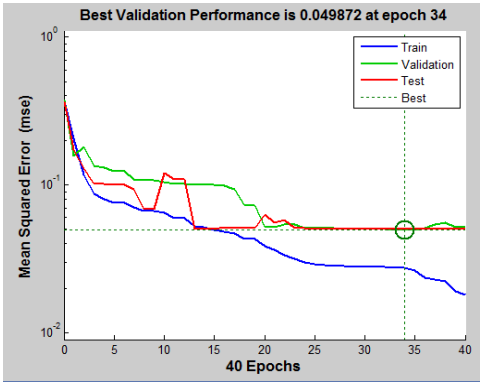

Fig 4(a): performance plot of the neural

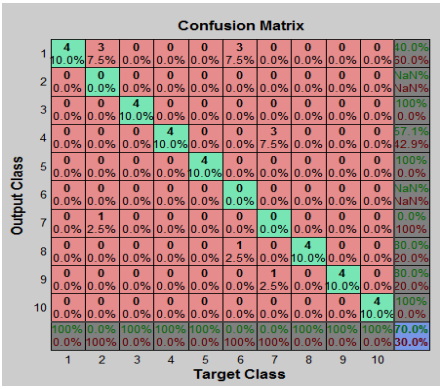

Fig 4(b): Accuracy achieved

For training the samples, a neural network toolbox is used. Performance of the network may be improved by a learning process also known as testing. The pattern recognition type is used to generate a neural network output that most closely matches the output value in the training data.[12].

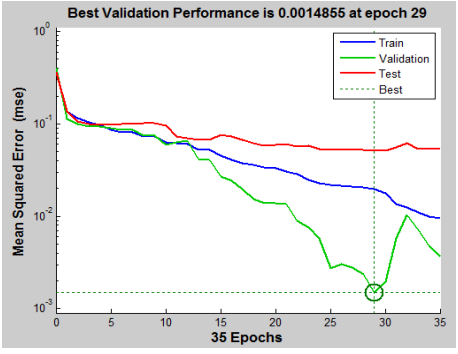

Fig 5(a): Performance plot of the neural network created for characters

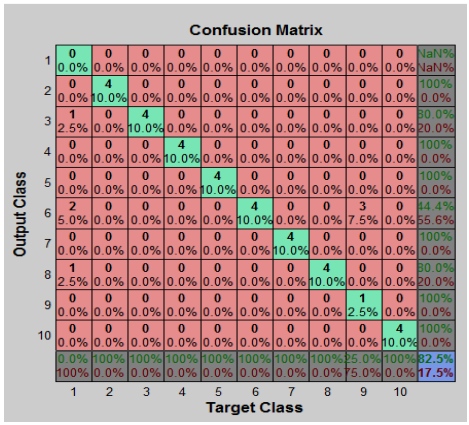

Fig 5(b): Accuracy achieved

When training is complete, it is needed to check out the network performance. The final command creates three regression plots for training, testing and validation. In the testing performance, it is needed to test the created network to see if it has found a good balance between normalization (accuracy) and generalization.

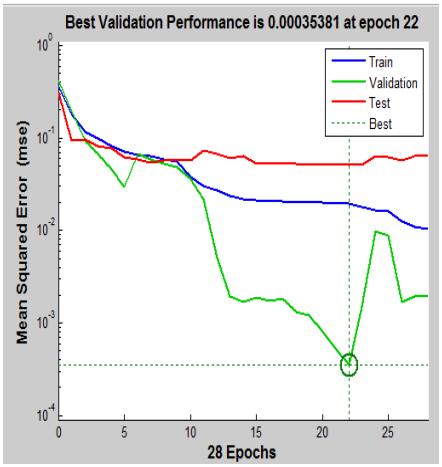

Fig 6(a): Performance plot of the NN created for Hindi vowels

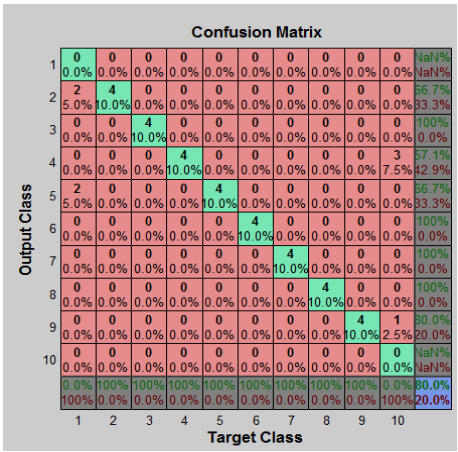

Fig 6(b): Accuracy achieved

For all the implemetation in our work, the software MATLAB abberiviates for matrix laborarty is used. Different data sets of different languages have been considered and it is shown that all the three performances i.e. training, validation and testing varies for difffrent datasets

Table 1. Results obtained from different scripts

Language Script	Gurumukhi Characters	Hindi Vowels	Alphabets
Total no.of epochs	35	28	40
Epoch no at best validation is achieved	29	22	34
Best validation performance	0.0014855	0.00035 381	0.049872
Gradient	0.0028008	0.01003	0.0056749

Table 2. Comparison of training, testing and validation performance

Language script	Gurumukhi Characters	Hindi Vowels	Alphabets
Performance	0.0204	0.0202	0.0296
Training Performance	0.0198	0.0196	0.0273
Testing Performance	0.0514	0.0516	0.0500
Validation Performace	0.0015	3.5381e-04	0.0499
Accuracy	82.5%	80%	70%

In table 1, results obtained from different samples are discussed and it is concluded that the best validation performance is achieved on recognition of alphabets at epoch no. 34 which is 0.049872.

Table 2 shows the training, testing and validation performance fro all the charaters and the total numner of epochs have been taken for each different script and the epoch number at which the best validation performance is achieved. The epoch is an instant in time chosen as the origin of a particular era

3. CONCLUSION AND FUTURE DIRECTIONS

In this paper, a pattern recognition type in neural network for classification purposes is applied. The proposed system is assessed on different datasets of different scripts. It has been found that the extracted features are effective in this system and good recognition rates have been achieved on each of the script. Highest accuracy has been achieved on Gurumukhi characters is 82.5%. The proposed system can further be directed on complete word recognition. The most important future direction of the proposed system is on the banking process to locate serial numbers.

4. REFERENCES

[1] R Plamondon and S.N. Srihari, "Online and Offline Handwritten Character Recognition: A Comprehensive Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.22, No.1, pp. 63-84, 2000.

[2] U Pal and B B Chaudhuri, "Indian Script Character Recognition: A Survey," Pattern Recognition, Elsevier, Vol.37, pp.1887-1899, 2004.

[3] Sutha, J.; Ramaraj, N., "Neural Network Based Offline Tamil Handwritten Character Recognition Systems", Conference on Computational Intelligence and Multimedia Applications, Vol.2, pp. 446-450, 2007.

[4] S N Nawaz., M. Zidouri, A., Al-Khitab, W. G., "An Approach to Offline Arabic Character Recognition Using Neural Network,"10th International Conference, IEEE, vol. 3, pp-1328-1331, 2003. P M. Kamble and R S. Hagadi, "Handwritten Marathi Character Recognition using R-HOG feature,"International Conference on ACT and A, 2015.

[5] P M. Kamble and R S. Hagadi, "Handwritten Marathi Character Recognition using R-HOG feature,"International Conference on ACT and A, 2015.

[6] S.G Mallat,., "A Theory of Multiresolution Signal Decomposition of Wavelet Representation", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 11, issue 7, pp. 674-693, July1989.

[7] R Verma and R Kaur "An Efficient Technique for Character Recognition using Neural Network and Surf Feature Extraction," International Journal of Computer Science and Information Technology, vol. 5(2), 2014, 1995-1997.

[8] S Nath and RS, "Handwritten Character Recognition: A Review," International Journal of Scientific and Research Publication, Vol. 5, issue 3, March 2015.

[9] K. V. Kale, P.D. Deshmukh, S. V. Chavan, M. M. Kazi, and Y. S. Rode, "Zernike Moment Feature Extraction for Handwritten Devanagari (Marathi) Compound Character Recognition,"International Journal of Advance Research in Artificial Intelligence, Elesvier, Vol. 3, No. 1, pp. 68-78, 2014.

[10] 10. G Sinha,. R Rani, and R Dhir, "Handwritten Gurumukhi Character Recognition using K-NN and SVM classifier,"International Journal Of Advance Research in Computer Science and Software Engineering,Vol.2, Issue 6, June 2012.

[11] A Roy and N. R. Manna, "Handwritten Character Recognition with Feedback Neural Network," International Journal of Computer Science & Engineering Technology (IJCSET), ISSN : 2229-3345, Vol. 5 No. 01 Jan 2014.

[12] Gupta A. and M Srivastava., "Offline Handwritten Character Recognition Using Neural Network," Computer Applications And industrial Electronics (ICCAIE), 7th international conference IEEE, 2011.

[13] V J Dongre , "A Review of Research on Devnagari Character Recognition,"International Journal of Computer Application, Vol. 12, No. 2, November 2010.

[14] P Kaur and B Singh, "Recognition of Signboard Imagesof Gurumukhi", Vol. 3, No. 6, June 2012.

[15] S A. Vaidya and B R. Bombade, "A Novel Approach of Handwritten Character Recognition using Positional Feature Extraction,", International Journal of Computer Science and Mobile Computing,Vol. 2, Issue. 6, pp.179-186, June 2013.