ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 4, April 2016

Comparison Study on Recognition of Gurumukhi Characters Using Neural Network with Different Number of Hidden Layers

Pankaj Rani, Sushil Kakkar, Shweta Rani

M. Tech Student, Dept. of ECE, BGIET Sangrur, Punjab, India

Assistant Professor, Dept. of ECE, BGIET Sangrur, Punjab, India

Associate Professor, Dept. of ECE, BGIET Sangrur, Punjab, India

ABSTRACT: This paper studies a comparison on recognition process of Gurumukhi Characters with different number of neurons. Use of neural network with different number of hidden layers is examined to accomplish optical character recognition. In this process of recognition two phases, one is segmentation of characters and other is feature extraction is included. A special method of feature extraction is introduced for extracting the features of the handwritten characters of Gurumukhi Script and the extracted data is used to train the neural network. Same data set is recognized four times by changing the number of neurons at every turn. It is done so that the recognition rate can be increased. The minimum number of neurons taken is five and it was observed that as we increase the number of neurons, the accuracy is decreased. Maximum accuracy achieved on Gurumukhi characters with 5 numbers of neurons is 90%

KEYWORDS: Character Recognition, Designed Neural Network and Handwritten Data Samples.

I.INTRODUCTION

Character recognition, usually abbreviated to optical character recognition or shortened OCR, is the mechanical or electronic translation of images of handwritten, typewritten or printed text (usually captured by a scanner) into machine editable text [1]. It is a field of research in pattern recognition, artificial intelligence and machine vision. The goal of Optical Character Recognition (OCR) is to classify optical patterns (often contained in a digital image) corresponding to alphanumeric or other characters. The process of OCR involves several steps including segmentation, feature extraction, and classification [2].The neural network technology can be used to analyse the stroke edge, the line of discontinuity between the text characters, and the background [3]. Allowing for irregularities of printed ink on paper, each algorithm averages the light and dark along the side of a stroke, matches it to known characters and makes a best guess as to which character it is. To implement the industrial optical character recognition, different steps like image acquisition, pre-processing, character segmentation, character recognition and post processing have to be carried out by using the special types of algorithms. The text can then be modified, searched, or copied as in a standard text document [9]. The OCR software then averages the results from all the algorithms to obtain a single reading [2].

OCR technology is now employed in a wide variety of fields to digitize documents normally received or maintained in hard copy [7-8]. All these different techniques results into one single column as the final accuracy using the confusion matrix plot and final accuracy has been observed on Gurumukhi Script. Plot confusion (targets, outputs) returns a confusion matrix plot for the target and output data in targets and outputs, respectively [10]. In the confusion plot, the predicted i.e. output class is corresponded by the rows and the true class i.e. target class is shown by the columns. The diagonal cells represent how many and what percentage of the trained network correctly estimates the classes of observations [9].

II. METHODOLOGY

MATLAB computation software with Neural Network Toolbox and Image Processing Toolbox is used to recognize the defined data. All the implementation and simulations are carried out with the use of pattern recognition in neural

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 4, April 2016

network .The proposed system used for all the implementations of recognition process is the neural network. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems [10].

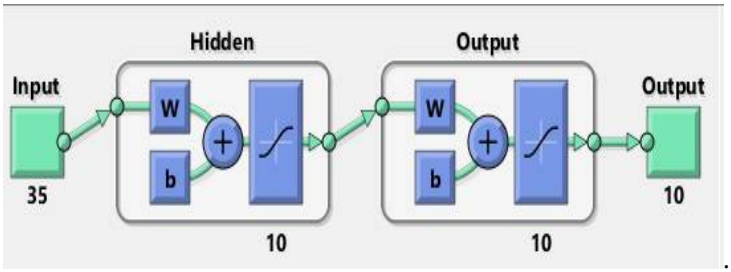

Figure 1: Pattern Recognition Using Neural Network

This is the most important setting through which the number of hidden layers in network, and number of neurons in each hidden layer is determined. Hidden layers are layers between input and output layer. The best try in this work is to have the smallest possible number of layers and neurons by which the training set can successfully learnt [5]. The smaller number of neurons - the faster learning, better generalization [5]. The number of input and output neurons also plays an important role in this process upon which the suitable number of hidden layers depends. This is performed on 10*5 images and one hidden layer with 10 neurons, which is the default setting. We can also increase or decrease the number of neurons by just entering the number 15 instead of 10 neurons.

III.RESULT AND DISCUSSION

The software MATLAB abbreviates for matrix laborarty is used for all the implementation purposes. Around 100 of samples are taken for proposed system. Out of which 90% of the samples are used for training, 5% are for validation and rest of the samples are used for testing. It is shown that all the three performances i.e. training, validation and testing varies at different turns with different number of neurons.

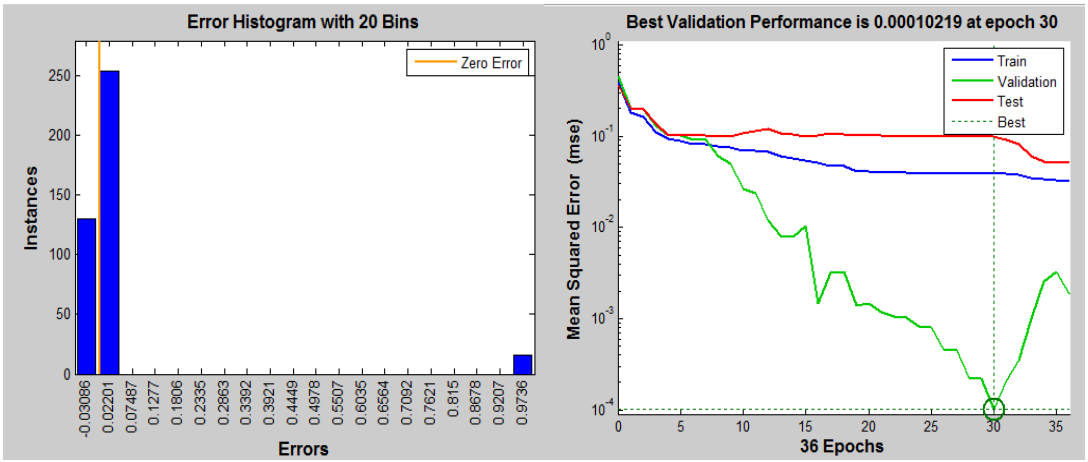

Figure 2(a) Figure 2(b)

Figure 2(a), 2(b): Error histogram and Best Validation performance with 20 no. of neurons

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 4, April 2016

In figure 2, the error histogram with 20 bins for all types and the best validation performance for the Gurumukhi Characters is shown with 20 number of neurons in the hidden layers. The best validation performance is 0.00010219 at epoch no. 30.

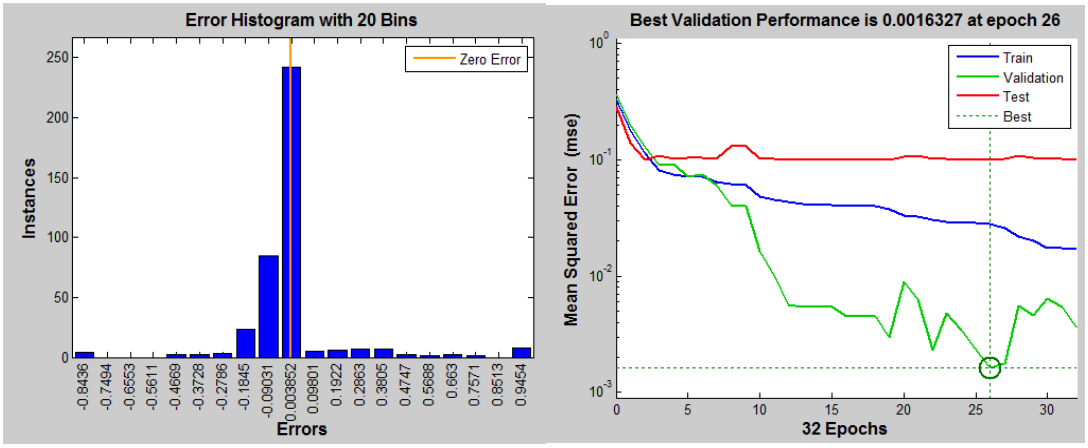

Figure 3(a) Figure 3(b)

Figure 3(a), 3(b): Error histogram and Best Validation performance with 15 no. of neurons

Same as shown in figure 2, it is also shown here the best validation performance achieved with having 15 number of neurons. Here the best validation performance achieved is 0.0016327 at epoch 26 whereas total number of epochs taken are 32. From both of these graphs it is concluded that as we decreased the number of hidden layers we have achieved the better accuracy.

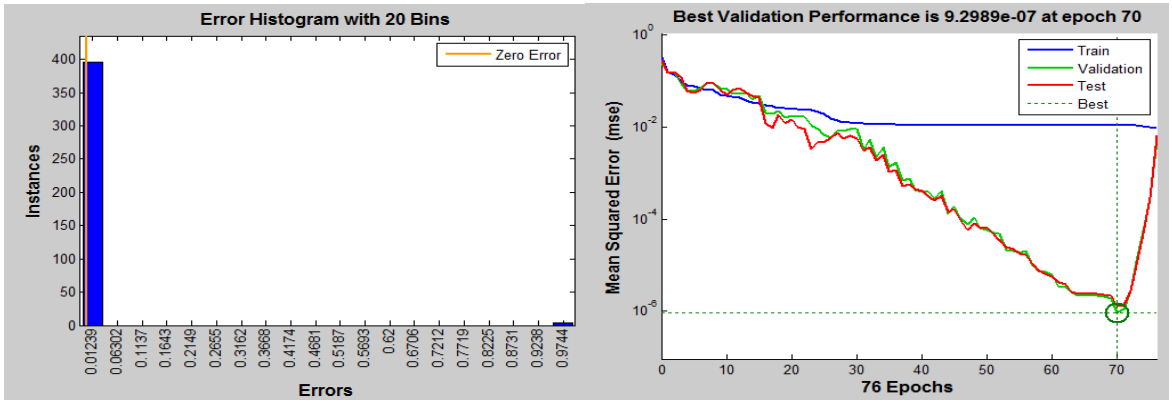

Figure 4(a)Figure 4(a)

Figure 4(a), 4(b): Error histogram and Best Validation performance with 5 no. of neurons

Training consists in learning a relation between data and attributes from a fraction of the training dataset, and testing consists in testing predictions of this relation on another part of the dataset. If prediction is known to us, we can compare the output of the relation and the real attributes.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 4, April 2016

Table 1: Accuracy achieved on Gurumukhi Characters with Different no. of neurons

S. No.	No. of Hidden Layers	Accuracy achieved in %age	No. of Epochs	Epoch no. at which best validation is achieved	Best Validation Performance
1	20	60	36	36	0.00010219
2	15	70	32	26	0.0016327
3	10	80	34	30	0.050453
4	5	90	76	70	9.2989e-07

The table no. 1 represents the detailed version of total numner of epochs taken for Gurumukhi script and the epoch number at which the best validation performance achieved . The epoch is an instant in time chosen as the origin of a particular era [6]. Further a reference point is served by it from which time is measuredand accurecy achieved with different number of hidden layers.

Table 2: Results obtained on different perfromances Gurumukhi Script

No of Hidden Layer	Training	Testing	Validation	Performance
20	0.0389	0.0999	1.0219e-04	0.0401
15	0.0424	0.0205	0.0559	0.0420
10	0.0204	0.0597	0.0505	0.0239
5	0.0111	1.2425e-06	9.2989e-07	0.0100

Table 2 shows the training, testing, validation performance and overall performance achieved on gurumukhi Script with differenr number of neurons. It is shown that the overall perormance obtained is going to vary as we decrese the number of neurons in the hidden layers.

IV.CONCLUSION

This paper presented a character recognition system for handwritten Gurumukhi characters. All the implementations of this systemare simulated andanalysed by the use of neural network toolbox in MATLAB. Same data set is optimized several times until satisfactory recognition rate on Gurumukhi script is achieved. This is done by varying the number of neurons in hidden layers at every turn. Maximum numbers of neurons taken in this research are 20 and accuracy achieved with this set of neurons is 60% which is not a satisfied result. To improve the obtained result, the system is again optimized by decreasing the number of neurons. The minimum number of neurons taken is 5 and with this set accuracy achieved is 90%. Hence, in this work it is concluded that the recognition rate goes to increase as the number of neurons in hidden layers goes to decrease.

REFERENCES

[1] S. Mori, C. Y. Suen and K. Kamamoto, "Historical review of OCR research and development," Proc. of IEEE, vol. 80, pp. 1029-1058, July 1992.
[2] M. Blumenstein and B. Verma, "Neural-based solutions for the segmentation and recognition of difficult handwritten words from a benchmark database," in Proceedings of the Fifth International Conference on Document Analysis and Recognition, ICDAR '99, pp. 281 – 284, September 1999.
[3] A. Rajavelu, M. T. Musavi and M. V. Shirvaikar, "A Neural Network Approach to Character Recognition," Neural Networks, vol. 2, pp. 387-393, 1989.
[4] Suray Nath RS, "Handwritten Character Recognition: A Review," International Journal of Scientific and Research Publication, vol. 5, issue 3, March 2015.
[5] U Pal and BB Chaudhuri, "Indian Script Character Recognition: A Survey," Pattern Recognition, Elsevier, vol.37, pp.1887-1899, 2004.
[6] A. Gupta and M. Srivastava, "Offline Handwritten Character Recognition Using Neural Network," Computer Applications And industrial Electronics (ICCAIE), 7th international conference IEEE, 2011.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 4, April 2016

[7] 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, Elsevier, Vol. 3, No. 1, pp. 68-78, 2014.
[8] Puneet kaur ,Balwnt singh "R of signboard images of Gurmukhi '',vol. 3 ,issue. 6, June 2012.
[9] Kulvir kaur ,Harpreet kaur ,"MGCR'' International journal of research science and management , ISSN:2349.5137, Jan2015.
[10] Vikas J Dongre et al, "A Review of Research on Devnagari Character Recognition," International Journal of Computer Application, vol. 12, No. 2, November 2010.
[11] P.M. Kamble and R.S. Hegadi, "Handwritten Marathi Basic Character Recognition Using Statistical Method," Emerging Research in Computing, Information, and Application, Elsevier, vol. 3, pp. 28-33, March 30, 2015.
[12] Anoop Rekha, "Offline Handwritten Gurumukhi Character and Numeral Recognition using Different Feature Sets and Classifiers-A survey," vol. 2, issue: 3, pp. 187-191, May-June 2012.