International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 04 | Apr -2017 www.irjet.net p-ISSN: 2395-0072

# FEATURE EXTRACTION METHOD BASED ON COMBINE CLASSIFIER FOR MARATHI HANDWRITTEN CHARACTER RECOGNITION

**Dr. Ratnashil N Khobragade**

*Assistant Professor, P G Dept of Computer Science, SGB Amravati University, Amravati, Maharashtra, India.*

---

**Abstract:** *In this paper the MLPNN based classifier is used for classification of Marathi handwritten characters. For MLP NN, various transfer functions and learning rules are investigated for different number of hidden layers and processing elements are set. The Scale Conjugate Gradient (SCG) algorithm is used as learning rule. Self generated database of 3700 characters in training set, test set and validation set is used. The Combine Feature Set Classifier (CFSC) with 240 features is proposed. The training performance of 98.61% is observed using 96 processing elements at 186 epochs with SSE of 1 and the testing performance of 98.42% is observed using SCG training on 50 characters. The training performance of 88.22% is obtained using 880 processing elements at 5000 epochs with SSE of 941 and best testing performance of 84% is observed training using SCG on 3700 characters.*

***Key Words:*** *Feature Extraction, Handwritten Devanagari character, Marathi character Recognition, Combine Feature Set Classifier, Zoning Density Features (ZDF), Artificial Neural Network.*

## 1. INTRODUCTION

Handwritten character recognition is a frontier area of research for the past few decades and there is a large demand for OCR on handwritten documents. This paper concentrated on the MLPNN based classifier for classification of Marathi handwritten characters. For MLPNN, various transfer functions and learning rules are investigated for different number of hidden layers and processing elements are set. The database is generated from the varying group of people and handwritten symbols are collected. The database of 50 and 3700 characters in training set, test set and validation set is used. The Line Classifier Zonal Features (LCZF) obtained 85 features, Row Column Features (RCF) 120 features, Zoning Density Features (ZDF) 35 features and Combine Feature Set Classifier (CFSC) 240 features are calculated and applied to MLPNN. The neural network is trained and tested on two sets of database, one database is of group of 50 characters and other database is of 3700 characters by varying parameters.

## 2. Characteristics of Marathi Script :

Marathi is an official language of Maharashtra and it is derived from Devanagari script. It is the 4th most spoken language in India and 15th most spoken language in the world. Marathi script consists of 16 vowels and 36 consonants making 52 alphabets. Marathi is written from left to right and has no upper and lower case characters. Every character has a horizontal line at the top called as the header line. The header line joins the characters in a word. Vowels are combined with consonants with the help of specific characteristic marks. These marks occur in line, at the top, or at the bottom of a character in a word [13]. Marathi also has a complex system of compound characters in which two or more consonants are combined forming a new special symbol. In India huge volumes of historical documents and books (handwritten or printed in Devanagari script) remain to be digitized for better access, sharing, indexing etc[14]. The objective of this research is to study the handwritten character recognition and explore a multi-feature multi-classifier scheme for handwritten Marathi characters.

## 3. LITERATURE REVIEW

Recognition of the text like human is still a challenging task for machine. Handwritten characters recognition has been a popular research area for many years because of its various application potentials. The detail literature review for the development in optical character recognition of Marathi handwritten character and Devnagari character recognition such as image preprocessing, segmentation, feature extraction, neural network classifiers and their implementation [13] etc. have been discussed here. In [1] M. Hanmandlu et al. presents the modified exponential membership function fitted to the fuzzy sets for recognition of handwritten Hindi characters based on features consisting of normalized distances obtained using the Box approach. Sandhya Arora et al. [2] suggested combine multiple feature extraction techniques for handwritten devnagari character recognition. Satish Kumar[3] suggested three tier strategy to recognize the hand printed characters of Devanagari script using multiple features and multi-stage classifier. The recognition rate of 94.2% is achieved with this scheme on database consisting of more than 25000 characters belonging to 43 alphabets. Invariant Moments for handwritten devnagari vowels recognition presented by R. J. Ramteke [4], is independent of size, slant, orientation, translation and other variations in handwritten vowels. In [5] O V Ramana Murthy, M Hanmandlu considered the character image divided into predefined number of zones and a feature is computed from each of these zones based on the pattern (black) pixels contained in that zone. Some of such features are sum squared

distance, histogram average pixel density. In [6] Shailendra Kumar Shrivastava and Pratibha Chaurasia use the energy features of segment characters for the classification of Devanagari character using SVM. The best result obtained in DATASET1 Linear kernel 96%, Quadratic kernel 100%, RBF kernel 97%, and polynomial kernel 100%. Dinesh V. Rojatkar et. al. investigates LRTB feature based classifier using single hidden layer feed-forward neural network with five fold cross validation applied to handwritten Devanagari consonant characters in [7]. They found best network at fold 5 with 80 neurons at trial 3. Networks analyzed on account of confusion matrix, reveals the greater details for individual classes. Average classification accuracy on training, validation, test and combined dataset is 99.40%, 97.38%, 97.05% and 98.98% respectively on the total dataset size of 8224 samples distributed uniformly within 32 classes of typical Devnagari consonants. In [8] Sushama Shelke, Shaila Apte presents a novel approach for recognition of unconstrained handwritten Marathi characters using multistage feature extraction and classification scheme and achieved 95.40% recognition rate. Zoning based feature extraction is propose by O. V. Ramana Murthy and M. Hanmandlu [9] in which character image is divided into predefined number of zones and a feature is computed from each of these zones. In [10] optimal classifier for the categorization of handwritten Marathi consonant characters using a single hidden layer feed-forward neural network with five fold cross validation is proposed by D V Rojatkar et. al. and obtained overall, classification accuracy on training, validation, test and combined dataset is 99.58%, 97.88%, 97.62% and 99.05% respectively on the total dataset size of 8224 samples distributed uniformly within 32 classes of typical Devnagari consonants. A detail survey of preprocessing, segmentation, feature extraction, classification and matching techniques for optical character recognition of general scripts presented by Ratnashil N Khobragade et. al. in [11]. In [12] Compound characters are one of the features of Marathi script Mrs. Snehal S. Golait, Dr. L.G. Malik present a short review on feature extraction for Marathi handwritten compound character recognition.

**4. THE PROPOSED APPROACH AND EXPERIMENTAL RESULT**

The Combine Feature Set Classifier (CFSC) with *240 features is proposed in this work.* The Scale Conjugate Gradient (SCG) algorithm and Gradient Descent & Momentum Adaptive Learning Algorithms (GDMALR) are used for training and learning the network multiple times. Initially 32 hidden neurons are chosen for experimentation. Then these hidden neurons are increased up to 128 in the step of 16 i.e.{32, 48, 64, 80, 96, 112 and 128}. In some of the training the neurons are increased up to 1016. For the classification of

characters different types of Artificial Neural Network (ANN) models are studied and the optimal neural network model is selected in experimentation. In the experimental setup the different combination of parameters such as varying number of processing elements, training algorithms, feature extraction methods and varying datasets are tested.

### 4.1 COMBINE FEATURE SET CLASSIFIER (CFSC)

Combine Feature Set Classifier (CFSC) feature extraction method is proposed and implemented in this work. Line Classifier Zonal Feature 85, Row Column Feature 120 and the Zoning Density Feature 35 are combined to form 240 features vector called Combine Feature Set Classifier (CFSC). These combine 240 features are used for training of neural network and then used to classify 37 consonants character of Marathi manuscript. The observations from the results obtained from the training and testing of CFSC method are described as follows.

The training and testing performance details for CFSC feature extraction method using SCG training on 50 characters is shown in table 1 and graph is shown in figure 1. It is observed that best training performance of 100% is obtained using 112 processing elements at 5000 epochs with SSE of 0.00099 and best testing performance of 99.1% is observed. This is highest training and testing performance observed in all methods for 50 characters using SCG algorithm.

**Table 1:Training and Testing Performance details for CFSC using SCG training on 50 characters.**

| Feature Extraction Method | Processing Elements | Best Epochs | Max Epochs | SSE | Gradient | Performance of Training | Performance of Testing |
|---|---|---|---|---|---|---|---|
| CFSC 240 | 32 | 5000 | 5000 | 1.05 | 1.41 | 99.16 | 98 |
| | 48 | 5000 | 5000 | 0.115 | 0.59 | 99.93 | 98 |
| | 64 | 5000 | 5000 | 0.0069 | 0.032 | 99.99 | 99 |
| | 80 | 5000 | 5000 | 0.0021 | 0.003 | 99.99 | 99 |
| | 96 | 5000 | 5000 | 19.44 | 5.9 | 68.29 | 40 |
| | **112** | **5000** | **5000** | **0.00099** | **0.0053** | **100** | **99.1** |
| | 128 | 5000 | 5000 | 0.00099 | 0.0028 | 100 | 99 |

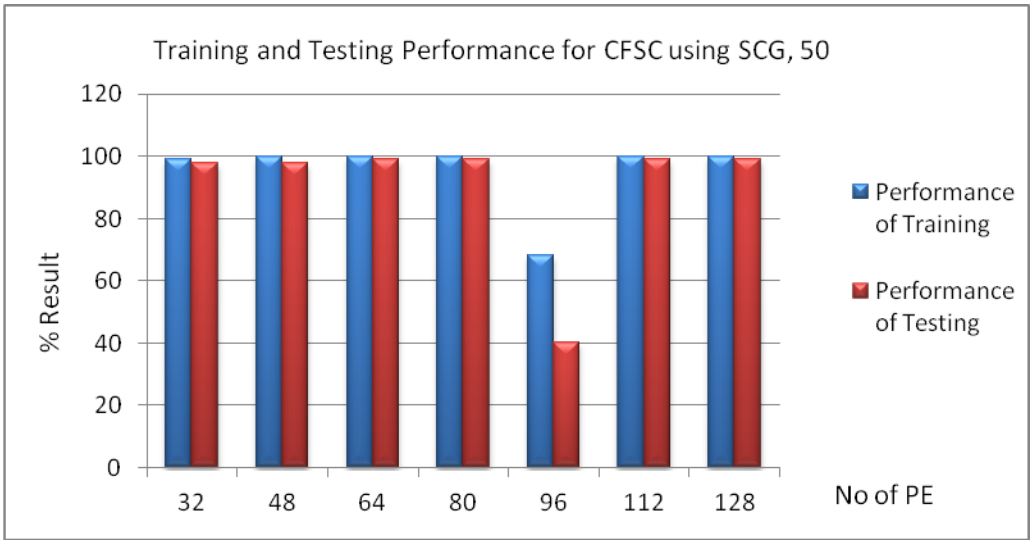


**Figure 1: Graph for Training and Testing Performance graph for CFSC method using SCG training algorithm on 50 characters.**

The training and testing performance details for CFSC feature extraction method using SCG training on 3700 characters is shown in table 2 and graph is shown in figure 2.

**Table 2:Training and Testing Performance details for CFSC using SCG training on 3700 characters.**

| Feature Extraction Method | Processing Elements | Best Epochs | Max Epochs | SSE | Gradient | Performance of Training | Performance of Testing |
|---|---|---|---|---|---|---|---|
| CFSC 240 | 32 | 5000 | 5000 | 2432 | 110 | 39.53 | 30 |
| | 48 | 5000 | 5000 | 2216 | 204 | 48.11 | 46 |
| | 64 | 5000 | 5000 | 2067 | 82.71 | 53.19 | 53 |
| | 80 | 5000 | 5000 | 2697 | 71.19 | 25.25 | 20 |
| | 96 | 5000 | 5000 | 2496 | 174 | 36.83 | 32 |
| | 112 | 5000 | 5000 | 2133 | 1218 | 51.01 | 48 |
| | 128 | 5000 | 5000 | 1483 | 199 | 69.77 | 70 |
| | 256 | 1 | 5000 | 2960 | 0.0024 | 0.002 | 0 |
| | 384 | 5000 | 5000 | 310 | 6.18 | 94.48 | 94 |
| | 512 | 5000 | 5000 | 237 | 17.48 | 95.81 | 95 |
| | 624 | 5000 | 5000 | 184 | 1.13 | 96.75 | 96 |
| | 752 | 5000 | 5000 | 175 | 1.01 | 96.92 | 96 |
| | 880 | 5000 | 5000 | 145 | 1.07 | 97.45 | 97 |
| | **1016** | **5000** | **5000** | **144** | **1.13** | **97.47** | **98.21** |

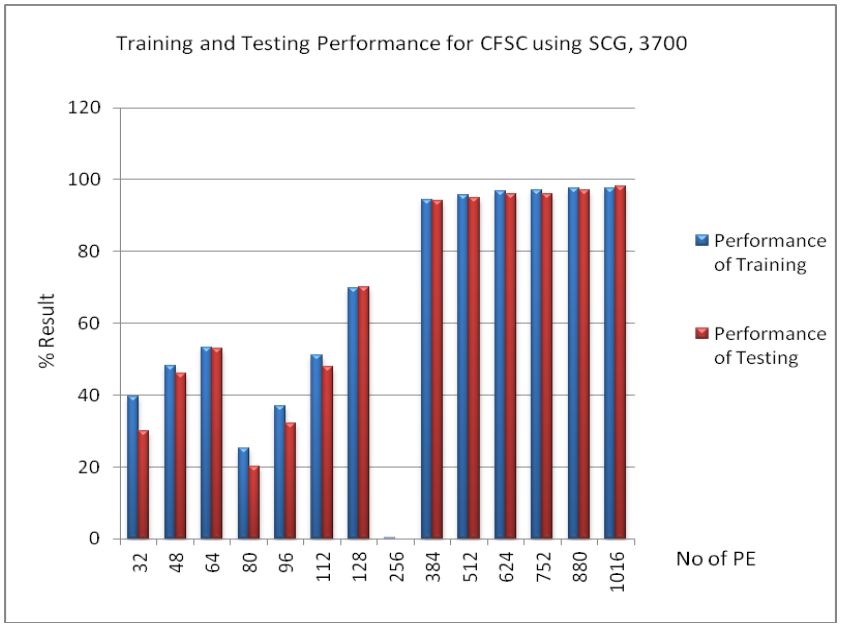


**Figure 2: Graph for Training and Testing Performance graph for CFSC method using SCG training algorithm on 3700 characters.**

The best training performance of 97.47% is obtained using 1016 processing elements at 5000 epochs with SSE of 144 and best testing performance of 98.21% is observed. This is the highest performance of testing observed in all methods for 3700 characters using SCG algorithm.

### 4.2 BEST PERFORMANCE METHOD

After training and testing of methods using various combinations of parameters the best training performance is shown in table 3, figure 3 and figure 4.

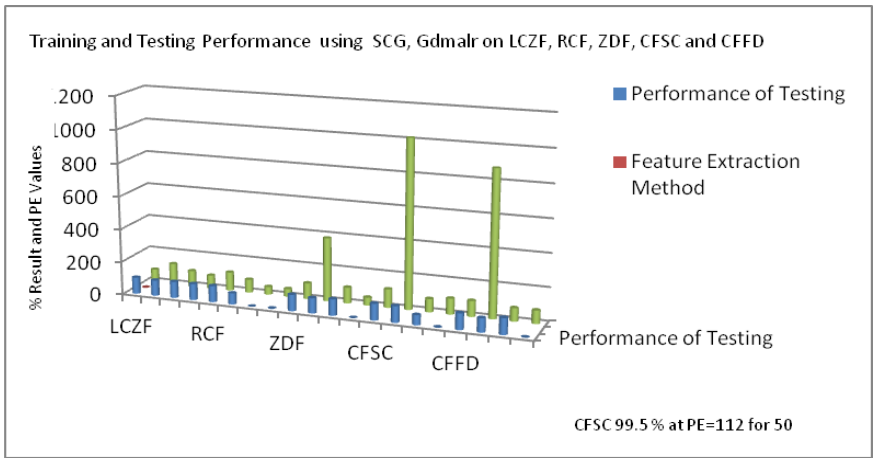


**Figure 3:Graph for Best Training and Testing Performance observed using SCG training algorithm on 50 and 3700 characters for all methods**

**Table 3: Best Training and Testing Performance using SCG and GDMALR training algorithm on 50 and 3700 characters for all methods**

| Feature Extraction Method | Training Algorithm | Characters | Processing Elements | Best Epochs | Max Epochs | SSE | Gradient | Performance of Training | Performance of Testing |
|---|---|---|---|---|---|---|---|---|---|
| LCZF | SCG | 50 | 80 | 609 | 5000 | 0.00099 | 0.022 | 99.99 | 99 |
| | SCG | 3700 | 128 | 5000 | 5000 | 445 | 2.1 | 92.23 | 94 |
| | GDMALR | 50 | 96 | 235 | 5000 | 0.98 | 0.19 | 99.99 | 99 |
| | GDMALR | 3700 | 80 | 5000 | 5000 | 0.48 | 0.63 | 99.98 | 98 |
| RCF | SCG | 50 | 112 | 2689 | 5000 | 0.001 | 0.004 | 100 | 99 |
| | SCG | 3700 | 80 | 7000 | 7000 | 1661 | 333 | 65.21 | 68 |
| | GDMALR | 50 | 48 | 10 | 5000 | 35.69 | 7.75 | 0.2 | 0 |
| | GDMALR | 3700 | 48 | 1 | 5000 | 2960 | 1.17 | 9.8 | 2 |
| ZDF | SCG | 50 | 96 | 58 | 5000 | 0.00077 | 0.00091 | 100 | 99 |
| | SCG | 3700 | 384 | 5000 | 5000 | 433 | 2.72 | 92.21 | 92 |
| | GDMALR | 50 | 96 | 233 | 5000 | 0.00099 | 0.00029 | 99.99 | 99 |
| | GDMALR | 3700 | 48 | 1 | 5000 | 2960 | 7.861 | 0.0014 | 0 |
| CFSC | **SCG** | **50** | **112** | **5000** | **5000** | **0.00099** | **0.0053** | **100** | **99.1** |
| | **SCG** | **3700** | **1016** | **5000** | **5000** | **144** | **1.13** | **97.47** | **98.21** |
| | GDMALR | 50 | 80 | 122 | 5000 | 35.99 | 7.98 | 63.02 | 60 |
| | GDMALR | 3700 | 96 | 1 | 5000 | 2960 | 2.24 | 3.51 | 1 |

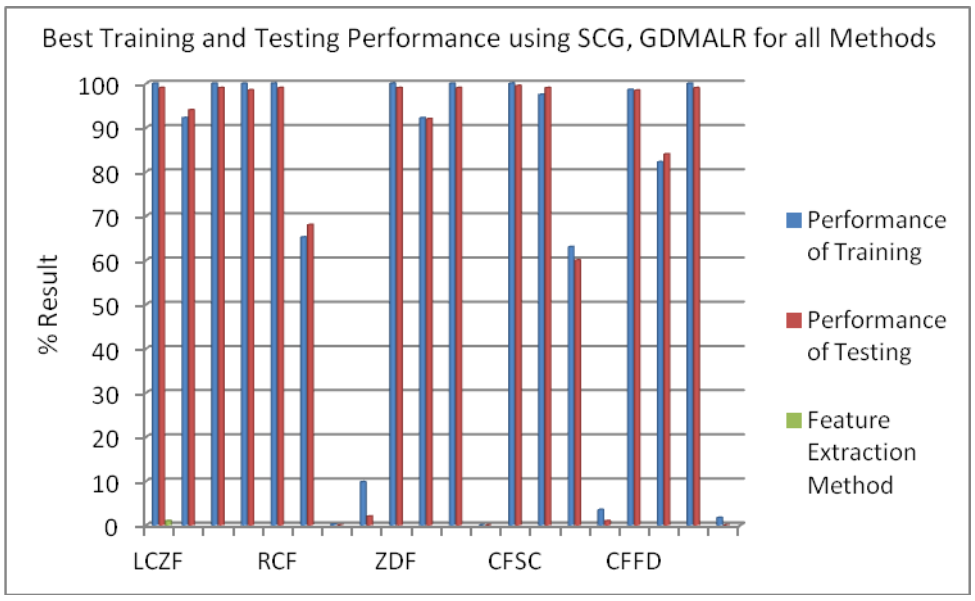


**Figure 4: Graph for Best Training and Testing Performance using SCG and GDMALR training algorithm on 50 and 3700 characters for all methods.**

Best training performance of 100% and testing performance of 99.1% is observed on 112 PE at 5000 epochs with SSE of 0.00099 for CFSC feature extraction method using SCG training on 50 characters. The best training performance of 97.47% and best testing performance of 98.21% is observed at 1016 PE on 5000 epochs with SSE of 144 for CFSC feature extraction method on 3700 characters using SCG algorithm shown in table 4 and figure 5.

**Table 4: Best Training and Testing Performance using SCG training algorithm on 50, 3700 characters.**

| Feature Extraction Method | Training Algorithm | Characters | Processing Elements | Best Epochs | Max Epochs | SSE | Gradient | Performance of Training | Performance of Testing |
|---|---|---|---|---|---|---|---|---|---|
| CFSC | SCG | 50 | 112 | 5000 | 5000 | 0.00099 | 0.0053 | 100 | 99.1 |
| | SCG | 3700 | 1016 | 5000 | 5000 | 144 | 1.13 | 97.47 | 98.21 |

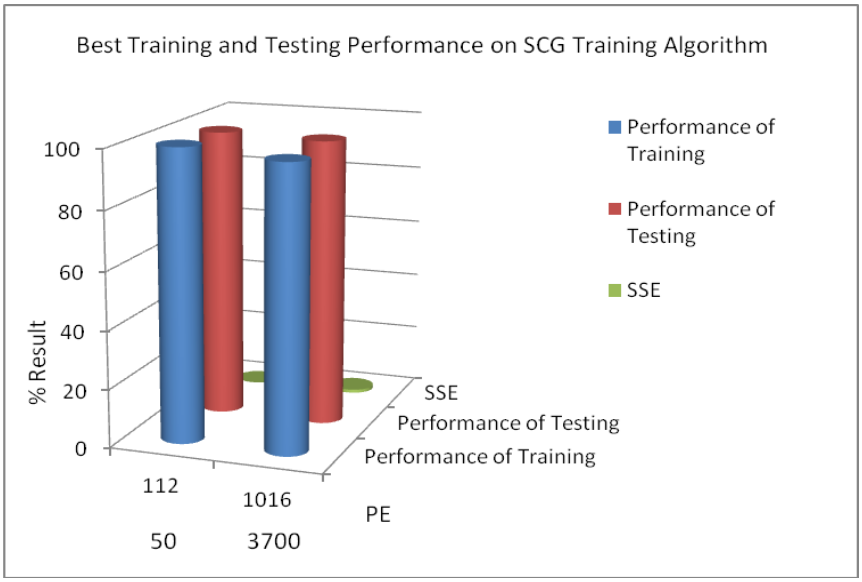


**Figure 5: Graph for Best Training and Testing Performance using SCG training algorithm on 50 and 3700 characters.**

## 5. CONCLUSION AND FUTURE SCOPE

In this work we have evaluated the performance of the MLPNN based classifiers for classification of Marathi handwritten characters using Combine Feature Set Classifier (CFSC) feature extraction method and results are found to be satisfactory. For MLPNN, various learning rules and transfer are investigated for different number of hidden layers and processing elements are set. The Scale Conjugate Gradient algorithm is used as default-learning rule. The training and testing performance of proposed feature extraction method using SCG training algorithms on 50 and 3700 characters are observed. The training performance of 98.61% is observed using 96 processing elements at 186 epochs with SSE of 1 and the testing performance of 98.42% is observed using SCG training on 50 characters. The training performance of 88.22% is obtained using 880 processing elements at 5000 epochs with SSE of 941 and best testing performance of 84% is observed training using SCG on 3700 characters. The features extraction method could be further change and analyze to reduce the dimensionally and computational complexity. The work can be extended for recognition of words and complete sentences for Marathi manuscript and for the other scripts.

**REFERENCES**

[1] M. Hanmandlu, O.V. Ramana Murthy Vamsi Krishna Madasu, "Fuzzy Model based recognition of handwritten Hindi characters", Digital Image Computing Techniques and Applications, 0-7695- 3067-2/07 $25.00 IEEE 2007.

[2] Sandhya Arora et.al., "Combining Multiple Feature Extraction Techniques for Handwritten Devnagari Character Recognition", Region 10 Colloquium and the Third ICIIS, Kharagpur, INDIA December 8-10, IEEE 2008.

[3] Satish Kumar, "Devanagari Hand-printed Character Recognition using Multiple Features and Multi-stage Classifier", International Journal of Computer Information Systems and Industrial Management Applications IJCISIM ISSN: 2150-7988 Vol.2, pp. 039- 055, 2010.

[4] R. J. Ramteke, " Invariant Moments Based Feature Extraction for Handwritten Devanagari Vowels Recognition", International Journal of Computer Applications 0975 - 8887 Volume 1 – No. 18, 2010.

[5] O V Ramana Murthy, M Hanmandlu, "Zoning based Devanagari Character Recognition", International Journal of Computer Applications 0975 – 8887 Volume 27– No.4, August 2011.

[6] Shailendra Kumar Shrivastava, Pratibha Chaurasia,"Handwritten Devanagari Lipi using Support Vector Machine", International Journal of Computer Applications (0975 – 8887) Volume 43– No.20, April 2012.

[7] Dinesh V. Rojatkar, Krushna D. Chinchkhede, G.G. Sarate , "Design and Analysis of LRTB feature based Classifier applied to Handwritten Devnagari Characters : A Neural Network Approach", International Conference on Advances in Computing, Communications and Informatics ICACCI pp. 97-101, 2013.

[8] Sushama Shelke, Shaila Apte, "A Novel Multi-feature Multi-Classifier Scheme for Unconstrained Handwritten Devanagari Character Recognition", 12th International Conference on Frontiers in Handwriting Recognition 2010.

[9] O. V. Ramana Murthy, M. Hanmandlu, "Zoning based Devanagari Character Recognition", International Journal of Computer Applications 0975 –8887 Volume 27–No.4, pp 21-25, August 2011.

[10] Dinesh V. Rojatkar, Krushna D. Chinchkhede, G.G. Sarate, "Handwritten Devnagari Consonants Recognition using MLPNN with Five Fold Cross Validation" International Conference on Circuits Power and Computing Technologies ICCPCT pp. 1222-1226, IEEE 2013.

[11] Ratnashil N Khobragade, Dr. Nitin A. Koli, Mahendra S Makesar, "A Survey on Recognition of Devnagari Script", International Journal of Computer Applications & Information Technology, Vol. II, Issue I, IJCAIT January 2013.

[12] Mrs. Snehal S. Golait, Dr. L.G. Malik," Review on Feature Extraction Technique for Handwritten Marathi Compound Character Recognition", 2013 Sixth International Conference on Emerging Trends in Engineering and Technology 978-1-4799-2560- 5/13, IEEE 2013.

[13] Ratnashil N Khobragade, Dr. Nitin A. Koli, Mahendra S Makesar, "Zoning Density Based Feature Extraction For Recognition Of Marathi Handwritten Character", International Research Journal Of Engineering And Technology (IRJET), Volume: 02 Issue: 04, Pp. 1819-1823, July-2015.

[14] Ratnashil N Khobragade, Dr. Nitin A. Koli, Mahendra S Makesar, "Analysis of Methods for Recognition of Devnagari Script", International Journal of Pure and Applied Research in Engineering and Technology, Volume 2 (8) pp. 27-38, IJPRET, 01-04-2014.