

Devanagari Handwritten Numeral Recognition using Feature Selection Approach

Pratibha Singh

Institute of Engineering and Technology, Indore, 452017, India
Email: prat_ibh_a@yahoo.com

Ajay Verma, Narendra S. Chaudhari

Institute of Engineering and Technology¹, Indian Institute Technology Indore², 452017, India
Email: ajay_rt@rediffmail.com, nsc183@gmail.com

Abstract— In this paper novel feature selection approach is used for the recognition of Devanagari handwritten numerals. The numeral images used for the experiments in the study are obtained from standard benchmarking data-set created by CVPR (ISI)Kolkata. The recognition algorithm consists of four basic steps; pre-processing, feature generation, feature subset selection and classification. Features are generated from the boundary of characters, utilizing the direction based histogram of segmented compartment of the character image. The feature selection algorithm is utilizing the concept of information theory and is based on maximum relevance minimum redundancy based objective function. The classification results are obtained for a single neural network based classifier as well as for the committee of Neural Network based classifiers. The paper reports an improvement in recognition result when decision combiner based committee is used along with class related feature selection approach.

Index Terms— Ensemble, MLP, Feature Selection, MRMR, MI

I. INTRODUCTION

Recognition of offline handwriting is the process of automatic reading of handwritten information by computer system on a piece of paper. The process undergoes steps like Pre-processing, Segmentation, Dimensionality reduction and classification. The reason for the poor result of recognition is mainly due to variation in writing style, the dependency on the writing tool used by the writer, the variation in size and orientation of the written characters. Pre-processing is the process of data conditioning operation so that it qualifies the initial level for the succeeding steps. Dimensionality reduction is needed to improve the performance of the recognition results in terms of accuracy and speed. There are two methods of dimensionality reduction: Feature extraction and Feature Selection. In the real world classification domain the class probability and class conditional probability is often unknown a priori. Also often the feature suitable for particular classification test is not known in advance. Therefore feature selection [1] [2] [3] [4] become important in most of the classification algorithms. By reducing the number of features, one can both reduce over fitting of learning methods, and increase the computation speed of prediction. The feature selection is the process of selection, from the whole set of available features, the subset allowing the most

discriminative power. The choice of a good feature subset is crucial in any classification problem. In Most of the classification algorithms we extract features out of which some of the features have more discriminating power for particular class than the others. These other features may be important recognition of some other class. So keeping this idea in mind a **class wise feature selection algorithm** is proposed. Feature selection for a feature set Y of cardinality N involves the exponential growth of the search space of the order 2^N . Therefore, heuristic algorithms become necessary for finding near-optimal solutions. Feature selection methods can be classified into two types, *filters* and *wrappers*. The first type is classifier independent, as they are not dedicated to a specific type of classification method. On the contrary the *wrappers* rely on the performance of one type of classifier to evaluate the quality of a set of features.

The rest of the paper is organized as follows. Section II describes the major related attempt made for the recognition of Devanagari script. The proposed methodology is explained in section III, where the information about the feature generation, the criterion used for feature selection and the method of combination of classifiers is explained. Section IV presents the experimental results and section V describes the conclusion and future scope for improvement.

II. RELATED WORK

The research in Devanagari started in the year 1977 by Sethi and Chatterjee [5]. Authors proposed a decision tree based approach for constrained hand printed Devanagari characters using primitive features. Hanmandlu and Murthy [6] used fuzzy model based recognition for handwritten Hindi numerals, where they used Normalized distance as feature for individual boxes and obtained 92.67% accuracy. Sharma [7] used Directional chain codes features and calculated it for blocks in bounding boxes which are applied to MQDF based classifier for obtaining total likelihood of the character and obtained 98.86 % accuracy for numerals and 80.36 % for characters.. Bajaj et al [8] employed set of different features namely density, moment and descriptive component feature for classification of Devanagari numerals. They used multi-classifier connectionist

architecture for increasing reliability and obtained 89.6% accuracy. Bhattacharaya et al [9] proposed a multilayer perceptron(MLP) neural network based classification approach for Devanagari numerals and obtained 91.28 % results. They used multiresolution features based on wavelet transforms. Rajan et al [10] proposed an efficient zone based feature extraction method for numeral recognition of four popular south Indian scripts. They used Zone centroid and image centroid based distance metric for feature extraction. Deshpande et al [11] proposed a regular expression based approach to Devanagari character recognition for increasing efficiency. In this approach the characters are first converted into sequences and then geometrical properties of characters are converted into regular expression which is used for matching purposes. Now the regular expression is tested with minimum edit distance. Chaudhari et. al [12] presented a pioneering effort for the development of handwritten numeral database of Indian scripts. For Devanagari characters Ensemble of classifier has been implemented by Satish Kumar et al [13]. In the paper author suggested a three tier approach wherein the first stage classify bar and not bar group of characters. After this separation, the further classification is done using different feature sets like Histogram, Profile, Edge based features etc. The overall recognition result obtained was 94.2 %. Agnihotri et al [14] implemented a system for Devanagari handwritten recognition using diagonal features and genetic algorithm for a dataset of 1000 samples and obtained 85.78 % match score.

III. PROPOSED METHODOLOGY

For the recognition of Devanagari characters the attempts made by various researchers cannot be compared because the dataset used by them were not common. The data in most of the previous efforts are generated by the research groups themselves. The benchmarking datasets were not there, but recently some dataset have been developed [12] for isolated numerals. So this study makes use of the above mentioned dataset. The method used by this study is materialized using steps of the block diagram shown in Fig. 1. The Input images from training and test dataset undergo the Preprocessing, Feature extraction, Feature ranking according to 'class to predict' criterion and classification using single as well as combination of classifier.

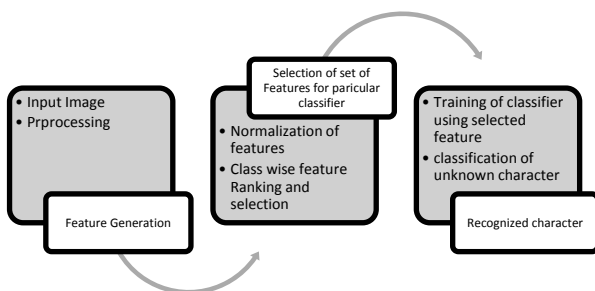


Fig. 1. Proposed Approach

A. Data Collection and Pre-processing

The first benchmark for the offline Hindi Numeral was published in the year 2009 [12]. The Devanagari character benchmark dataset from CVPR unit is still not available for the research community. CVPR Unit of ISI Kolkata is the first unit affiliated to Internationally Recognized association IAPR (International Association for Pattern Recognition) and is providing one of the major efforts towards the recognition of Indian languages. The numeral dataset from CVPR unit consists of 22,556 images stored in 'tif format collected from 1049 writers. From the dataset 200 images of each class are randomly selected and used for experimentation in our study. Some sample images from the dataset are shown in Fig. 2. The grayscale images are converted to two tone images. The bounding box of character image is obtained and then boundary is extracted. The details about Pre-processing and feature generation can be found in [15]. The output of pre-processing steps are shown in Fig. 3

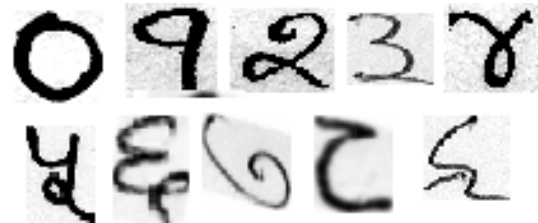


Fig. 2. Image samples from ISI data set

After obtaining the boundary, features are generated for dimensionality reduction using histogram of gradient directions. The boundary image is divided into 9 compartments as shown in Fig. 4. Histogram is obtained as an accumulation done for each of these 9 compartments

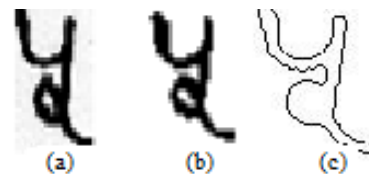


Fig. 3. (a). Gray-scale image of number '5' (b) Black & white image(c) boundary of the image

B. Feature Extraction using Gradient Direction

The gradient feature decomposition originally proposed for online character recognition by A. Kawamura et al [16] which reported a high recognition efficiency for Japanese characters. This method is applicable to machine printed/handwritten, binary/ grayscale as well as low resolution images. The calculated gradient of the image is decomposed into four, eight or sixteen directional planes. For our analysis we have taken eight directional plane. The Fig. 4 shows the gradient vector decomposition into their nearest vector plane. We have accumulated the magnitude of gradient in sampled eight directions for each of the subsection of original image.

$$x(x, y) = f(x + 1, y - 1) + 2f(x + 1, y) + f(x + 1, y + 1) - f(x - 1, y - 1) - 2f(x - 1, y) + f(x - 1, y + 1) \tag{1}$$

$$gy(x, y) = f(x - 1, y + 1) + 2f(x, y + 1) + f(x + 1, y + 1) - f(x - 1, y - 1) - 2f(x, y - 1) + f(x + 1, y - 1) \tag{2}$$

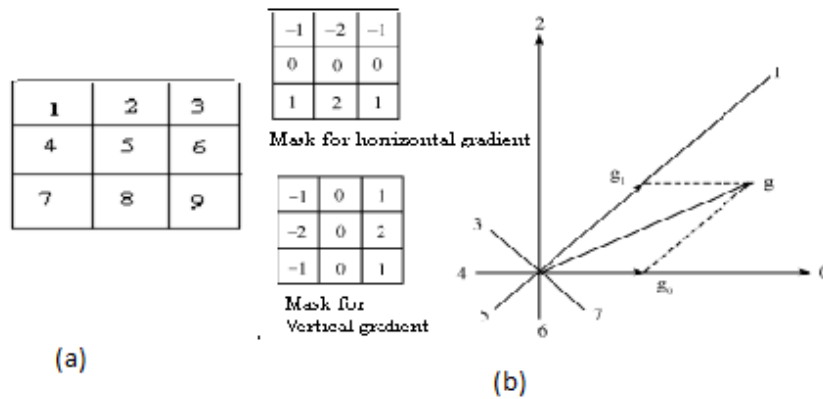


Fig. 4. (a) Zoning of bounding box of image(b) Gradient Feature Calculation Method

Conventionally the gradient is calculated on each pixel of the image. In our analysis we have applied ‘sobel’ edge detection algorithm to calculate gradient vector at each image pixel of preprocessed grayscale image. The gradient vector can be quantized into eight directions using angle vector quantization or vector decomposition using parallelogram rule. In the first method the magnitude of gradient in each image pixel is assigned to directional plane nearest to it and in the second method the gradient vector is decomposed into two nearest directional planes using parallelogram vector division rule. The parallelogram quantization method gives less quantization error so we have applied this method for quantizing directions of gradient vector.

C. Feature selection

The goal of feature selection is to search the most relevant features from a list of large number of features of a particular problem domain. Let X be the original set of features, with cardinality n . The continuous feature selection problem refers to the assignment of weights to each feature in such a way that the order corresponding to its theoretical relevance is preserved. Feature selection algorithms are characterized on the basis of the search organization, generation of successors and on the basis of evaluation function [17]. The classification of feature selection methods is given in Fig. 5 and more details about the methods and the criterion used for evaluation can be found in the survey paper by A. Belanche et al. [17]. For this study Filter method of feature selection is used whose evaluation criterion is based on mutual information. Some of the core definitions related to mutual information are explained in the following subsection.

D. Feature selection using Entropy and Mutual Information

In feature selection problem, the relevant features have important information regarding the output, whereas the irrelevant features contain little information regarding the output. An experimental comparison of various feature selection algorithm applied to different classifiers is

reported by Bhalla et al [18] for the classification of Microarray data. The objective of feature selection is to find those features that contain as much information about the output as possible. For this purpose, Shannon’s information theory [19] provides a feasible way to measure the information of random variables with entropy and mutual information. MI is capable of measuring a general dependence between two features without assuming the distributions of the features, and case based reasoning requires no assumption on the different project features to derive the solutions.

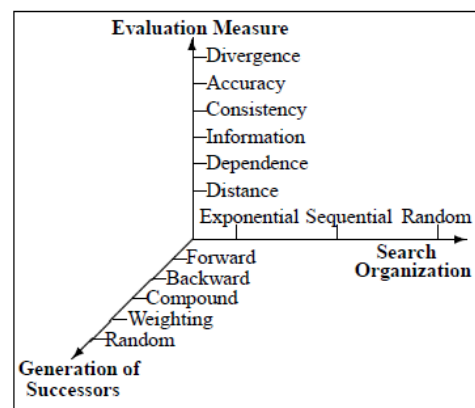


Fig. 5. Classification of feature selection algorithms [17]

The entropy $H(X)$ is a measure of the uncertainty of a random variable X . For a discrete random variable X , with the probability density $p(x)$, the entropy of X is defined as

$$H(X) = - \sum_{x \in X} p(x) \log p(x) \tag{3}$$

Here, the log base is 2 and entropy is expressed in bits. The joint entropy of X and Y with joint pdf $p(x,y)$,

$$H(X, Y) = - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x, y) \tag{4}$$

When certain variables are known and others are not, the remaining uncertainty is measured by the conditional entropy,

$$H(Y|X) = -\sum_{x \in X} p(x)H((Y|X = x) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x|y) \quad (5)$$

Therefore, the joint entropy and conditional entropy has the following relation:

$$H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y) \quad (6)$$

The information found shared by two random variables is important in our work and it is defined as the mutual information between two variables:

$$I(X; Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \frac{\log p(x|y)}{p(x)p(y)} \quad (7)$$

If the mutual information is large, the two variables are closely related. If the mutual information becomes zero, the two variables are independent. The probability of the individual i being selected is given by:

$$S_i = \frac{1}{n} \sum_{i=1}^n I(X_i; Y) - \frac{1}{n^2} \sum_{i=1}^n I(X_i; X_j) \quad (8)$$

where X_i is the input and Y is the output. The above mention criterion is known as Maximum relevance minimum redundancy based feature selection which was introduced by Peng et al [20].

E. Artificial Neural Network based classifier (ANN-10, ANN-20, ANN-30)

A standard feed forward network with one hidden layer of 10 neurons and sigmoid transfer function is used as classifier. The network is trained by the back-propagation learning algorithm using the MATLAB Neural Network toolbox [21]. The 10 output values are normalized and used as posterior probabilities. Again experiments were conducted with 20 and 30 neurons in the hidden layer.

F. Classifier combination module

According to the level for a pattern recognition task, classifier combination can be grouped as: Feature level combination and the decision or score level combination out of which the score level combination is most popularly used. This is due to the fact that the knowledge of the internal structure of classifiers and their feature vectors is not needed in the score level. Though there is a possibility that representational information is lost during such combinations, this is usually compensated by the lower complexity of the combination method and superior training of the final system. After calculating the measurement level output using posterior probabilities $\{p_{ij}(x), i = 1, m; j = 1, c\}$ for a m classifiers and c classes, fixed combining rules are used as decision combiner [22]. The confidence $q_j(x)$ for class j is computed by

$$q'_j(x) = \text{Rule}_i(p_{ij}(x)) \quad (9)$$

$$q_j(x) = \frac{q'_j(x)}{\sum_j q'_j(x)} \quad (10)$$

The following combiners are used for rule given in equation (9): *Majority Voting, Maximum, Sum, Minimum, Average, Product.*

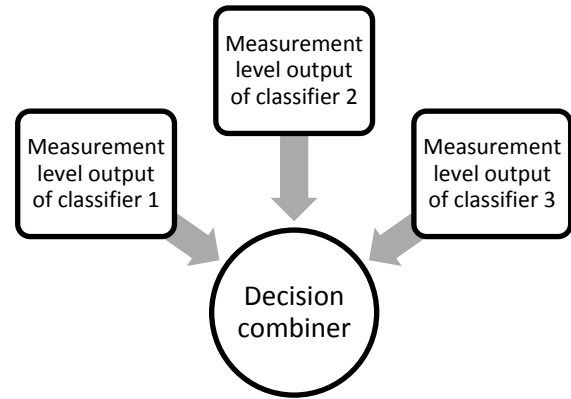


Fig. 6. Process of combination of classifier

Although the classifier ensemble had been used by many researchers, the novelty of our work is that we applied a MRMR Feature selection based classifier ensemble method for the recognition of Hindi Numerals.

IV. EXPERIMENTAL RESULTS

Experiments are firstly performed on numeral images. Individual classifier recognition rate as well as using committee of classifier is obtained without feature selection. The results are shown in table 1. The time mentioned in the table include the time for training, the time for testing as well as the time for feature selection (when used). The time for pre-processing and feature extraction is not included in the results. Also the feature selection method uses the class dependent feature selection method for obtaining features for specific class. The feature vectors are combined to form the resultant feature vector. The Results are showing the effectiveness feature selection based ensemble over the single classifier. The Experiments are conducted on 2000 image samples from ISI dataset, out of which 70% used for training and 30% used for testing. The Proposed strategy produces the results comparable to [23] with only 200 sample image per class. This strategy will be effective for dataset of small size, as in Hindi only a limited dataset is available in public domain.

From the results presented in table - I, it is clear that the recognition rate is low if all 72 features are used with a average processing time of 1.9 sec for a single image. However when the feature selection criterion with selection on the basis of class to predict is used the recognition performance improves by 5%. The time needed for processing of a single image is 2 second.

The result of combination of classifier as shown in Fig. 7-9. It is clear from the figures that the recognition performance of proposed method with 48, 56, 64 features is considerably good. Also recognition performance of the proposed method with 48 features gives 98% and above recognition efficiency with processing time reduced to half of the time required for features greater than or equal to 64 with ANN-30.

Table 1. Recognition Performance by varying number of features

Number of Feature	Recognition Rate for Single classifier	Recognition Rate of combiner	Time for the Recognition (sec)	Recognition Rate for single classifier	Recognition Rate of combiner	Time for the Recognition (sec)	Recognition Rate for single classifier	Recognition Rate of combiner	Time for the Recognition (sec)
	Using 10 nodes in hidden layer			Using 20 nodes in hidden layer			Using 30 nodes in hidden layer		
10Feat	0.80805	0.8355	332	0.8424	0.867	490	0.8612	0.879	711
20Feat	0.89433	0.9207	348	0.9091	0.9328	686	0.9135	0.938	1346
36Feat	0.9459	0.9655	431	0.9467	0.9677	852	0.954	0.9758	2032
48 Feat	0.9615	0.9793	569	0.9599	0.981	1182	0.9636	0.9827	2250
56 Feat	0.9644	0.9832	645	0.9652	0.9827	1257	0.968	0.9843	2470
64Feat	0.9622	0.9808	904	0.9699	0.9843	1735	0.9692	0.9853	4078
All Feat	0.8796	0.9322	574	0.8952	0.941	1449	0.8833	0.9375	3812

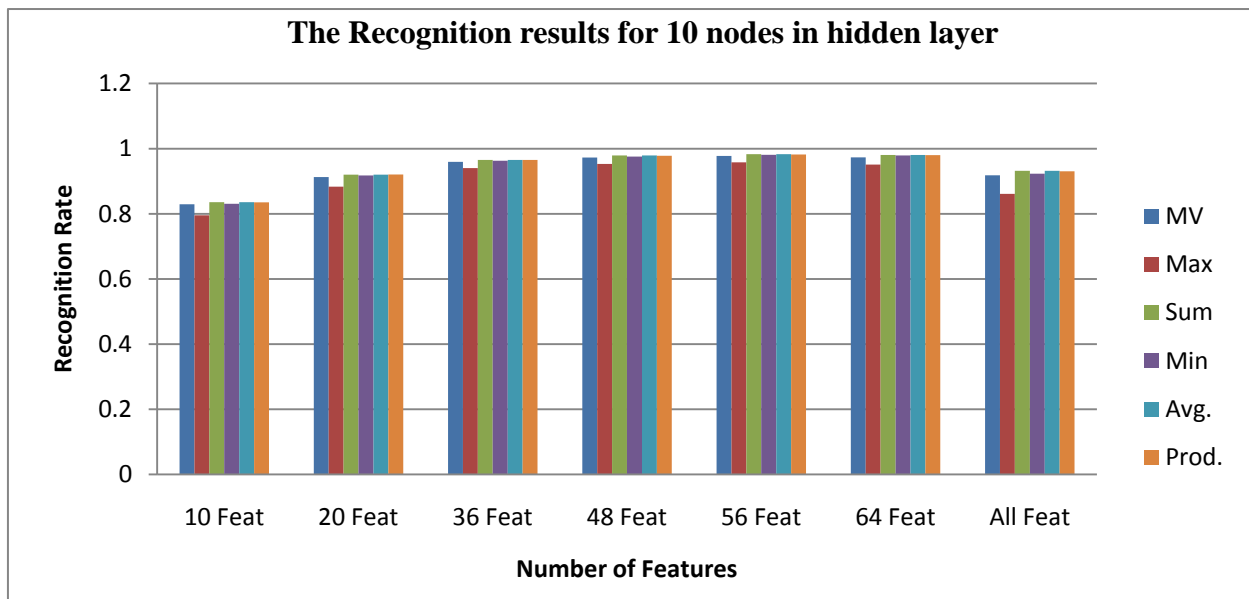


Fig. 7. Performance comparison of proposed approach with the single classifier with 10 nodes in hidden layer.

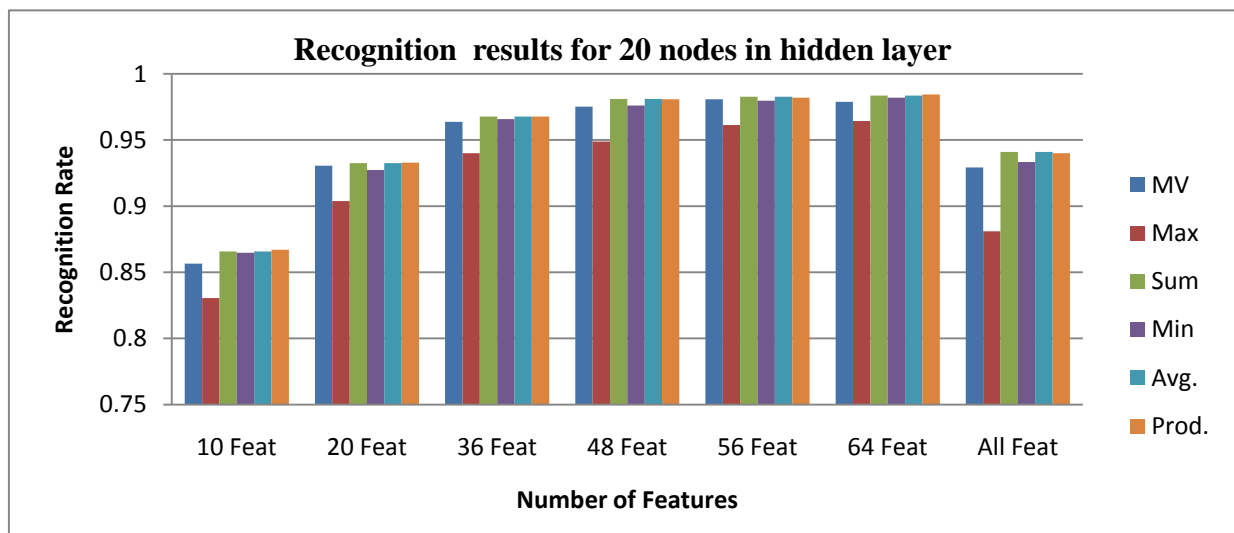


Fig. 8. Performance comparison of proposed approach with the single classifier with 20 nodes in hidden layer.

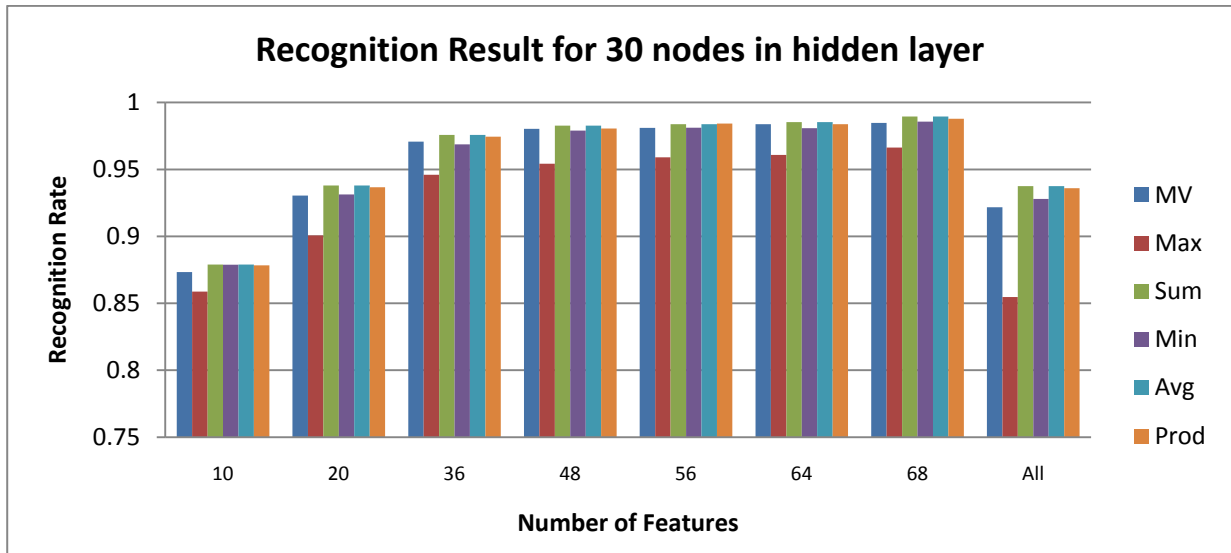


Fig. 9. Performance comparison of proposed approach with the single classifier with 30 nodes in hidden layer.

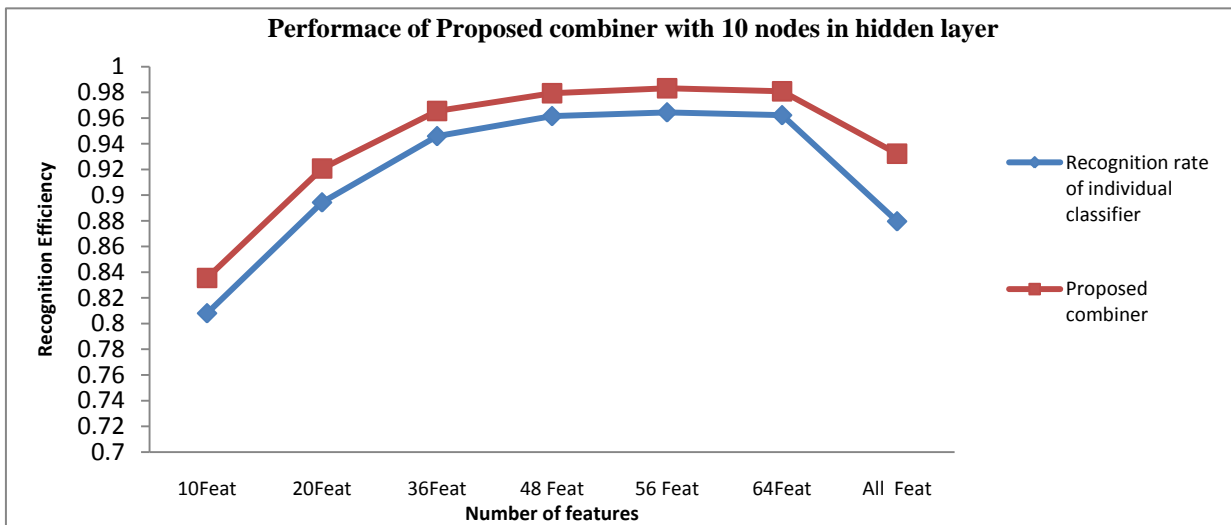


Fig. 10. Performance comparison for feature selection based ensemble for ANN-10

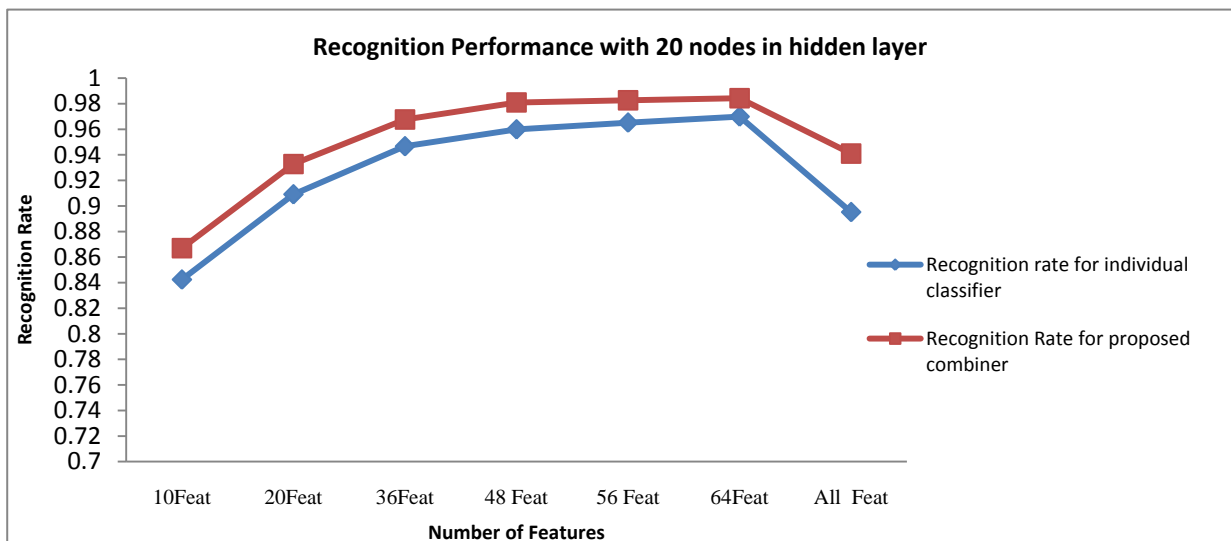


Fig. 11. Performance of feature selection based classifier ensemble for ANN-20

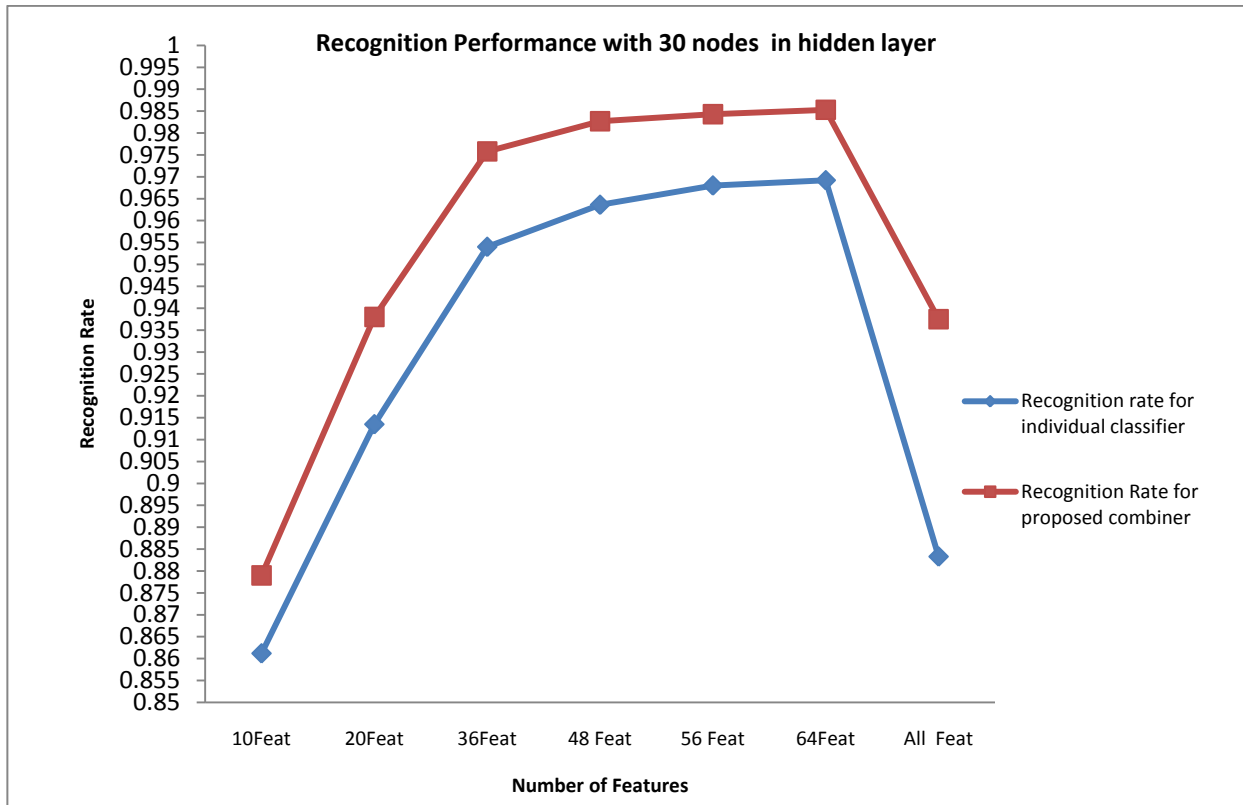


Fig. 12. Performance of feature selection based classifier ensemble for ANN-30

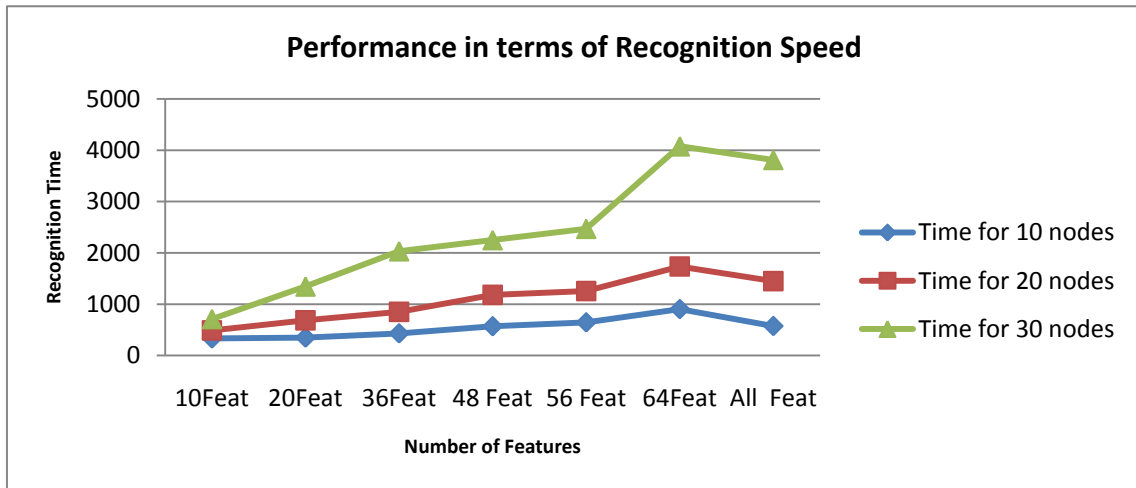


Fig. 13. Processing time for various patterns excluding preprocessing

V. CONCLUSION

The contribution of the paper is to propose a **feature selection based classifier combination algorithm** for the handwritten Devanagri numeral recognition. The obtained results are verifying that (1) the combination of classifier is improving the recognition result by 2-3 % than the single classifier. (2) Selected feature based on proposed “**class specific feature selection using MRMR**” based optimization criterion gives better recognition result. The combination rules used belongs to low complexity combining algorithm results in faster

recognition results. The experiments are ongoing and the method using complex combination rules are underway. In future the better recognition performance can be approached by increasing the complexity in combination algorithms.

REFERENCES

[1] P Langley and A L Bium, "Selection of Relevant Features and Examples in Machine Learning," *Artificial Intelligence*, vol. 97, Artificial Intelligence, no. Artificial Intelligence on Relevance, pp. 245-271, 1997.

- [2] Guyon Isabelle and Andr'e Elisseeff, "F, 2003. An introduction to variable and feature selection," *Journal of Machine Learning Research*, 3, pp. 1157–1182., 2003.
- [3] A. Jain and D. Z Guyon, "Feature selection: Evaluation, application, and small sample performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 2, pp. 153–158, 1997.
- [4] John Kohavi and Ron H George, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97(1–2), pp. 273–324, 1997.
- [5] I.K. Sethi and B. Chatterjee, "Machine Recognition of constrained Hand printed Devnagari," *Pattern Recognition*(9), pp. 69-75, 1977.
- [6] O. V. Ramana Murthy and M. Hanmandlu, "'Fuzzy Model Based Recognition of Handwritten Hindi Numerals," in *Intl. Conf. on Cognition Recognition*, 2005, pp. 490-496.
- [7] N Sharma, U Pal, F Kimura, and S. Pal, "Recognition of Off-Line Handwritten Devnagari Characters Using Quadratic Classifier," in *ICVGIP 2006, LNCS 4338*, 2006, pp. 805 – 816.
- [8] Reena Bajaj, Lipika Dey, and S Chaudhary, "Devnagari numeral recognition by combining decision of multiple connectionist classifiers," *Sadhana, Vol.27, part. 1*, pp. 59-72, 2002.
- [9] B. B. Chaudhuri, R. Ghosh and M. Ghosh U. Bhattacharya, "On Recognition of Handwritten Devnagari Numerals," in " , In *Proc. of the Workshop on Learning Algorithms for Pattern Recognition (in conjunction with the 18th Australian Joint Conference on Artificial Intelligence)*.
- [10] P. V. Rajan and S. V. Rajashekaradhy, "Efficient Zone based feature extraction method for handwritten numeral recognition of four popular south Indian scripts," *JATIT 2005-2008*.
- [11] P.S. Deshpande, L. Malik, and S. Arora, "Fine classification and Recognition of Handwritten Devnagri characters with regular expressions and minimum edit distance method," *Journal of Computers*Vol.3 No 5, may2008.
- [12] U. Battacharya and B. B. Chaudhari, "Handwritten numeral databases of Indian scripts multistage recognition of mixed numerals," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 31(3), pp. 444-457., 2009.
- [13] Satish Kumar, "A Three Tier Scheme for Devanagari Hand-printed Character Recognition," in *World Congress on Nature & Biologically Inspired Computing*, 2009, pp. 1016-1021.
- [14] Ved Prakash Agnihotri, "Offline Handwritten Devanagari Script Recognition," *I.J. Information Technology and Computer Science, MECS Press*, vol. V-4, pp. 37-42, July 2012.
- [15] P. Singh, A. Verma, and N. S. Chaudhari, "Performance Evaluation of Classifiers Applying Directional Features for Devnagri Numeral Recognition," *AMR , Smart Nano systems*, vol. 403-408, pp. 1042-1048, 2012.
- [16] A. Kawamura et al., "On-Line Recognition of Freely Handwritten Japanese Characters Using Directional Feature Densities," in *11th IAPR conference in Pattern Recognition Methodology and systems*, 1992, pp. 183-186.
- [17] A Belanche, L Nebot, and C Molina, "Feature selection algorithms: a survey and experimental evaluation," in *306 - 313, IEEE International conference on data mining*, pp. 306 - 313.
- [18] Aarti Bhalla and R. K. Agrawal, "Microarray Gene expression Data classification using less gene expression by combining feature selection methods and classifiers," *I.J. Information Engineering and Electronic Business*, vol. V5, pp. 42-48, November 2013.
- [19] C. E., & Weaver, W. Shannon, *The mathematical theory of communication*. Urbana, IL: University of Illinois Press., 1949.
- [20] Hanchuan Peng, Fuhui Long, and Chris Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226-1238, 2005.
- [21] H. Demuth and M. Beale. (1998) Mathworks Natick, MA, USA.
- [22] J. Kittler, M. Hatef, and R.P. W. Duin, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226-239, 1998.
- [23] Tarun Jindal and Ujjwal Bhattacharya, "Recognition of Offline Handwritten Numerals Using an Ensemble of MLPs Combined by Adaboost," in *Proceedings of the 4th International Workshop on Multilingual OCR*, Washington, 2013.

Authors' Profiles



Pratibha Singh has done B.E. in Electrical Engineering in the year 1999 and M.E. in Electrical Engg(with specialization in Digital Techniques and Instrumentation) in year 2001 respectively from SGSITS Indore. She is working as Assistant Professor in IET DAVV Indore. She has contributed papers in 8 conference proceedings and 5 peer review journals. Her research area includes computer vision, Pattern Recognition, Image processing.



Ajay Verma received his Ph.D. degree in Year 2002 from Devi Ahilya University. His research interests include nonlinear dynamics and system theory, Image Processing and nonlinear control system. He had published articles more than 15 in peer review journals and around 25 in national and international conference proceedings



Narendra S. Chaudhari has completed his undergraduate, graduate, and doctoral studies at Indian Institute of Technology (IIT), Mumbai, India. He is currently with IIT, Indore, India as a Professor of Computer Science and Engineering. Since Engineering, Nanyang Technological University (NTU), Singapore. He has been a referee and reviewer for various premier conferences and journals including IEEE Transactions, Neurocomputing, etc. He has more than 240 publications in top quality international conferences and journals. His current research work includes network protocols, security, optimisation algorithms, game AI, parallel computing, and theoretical computer science