Improving Performance of PNN using Clustered ICs for Gender Classification

Sunita Kumari, Sambit Bakshi, Banshidhar Majhi
sunita.nitrkl@gmail.com, sambitbaksi@gmail.com, bmajhi@nitrkl.ac.in
National Institute of Technology Rourkela
Odisha-769008, India

Abstract—The research presented in this paper proposes a novel gender classification approach using face image. The approach extracts features from grayscale face images through Infomax ICA and subsequently selects features using k-means clustering and classifies the clustered features employing PNN. All the experimental evaluations are done on cropped face images from FERET database using 280 faces for training and 120 different faces for testing. The approach, when features are not clustered gives maximum accuracy of 93.33%. However the proposed approach yields 95% accuracy through employing clustering on features, which is significant for gender classification using low resolution (118 × 97) face images.

Keywords- gender classification, PNN, K-mean clusterings, ICA.

I. INTRODUCTION

The most important aspect of practical deployment of a biometric system lies in its response time, which further depends on the matching time of probe and gallery images. When the dataset is as large as the population of a nation, then the response time falls down rapidly due to 1:*N* nature of matching. The response time can be improved by partitioning the whole dataset by classification method and to search only a partition of the total dataset when a query comes. There are several classification criteria viz. gender, ethnicity, age. Gender classification drags the attention because of the fact that it mostly partitions the database into two equal halves, and hence the worst case matching time is also minimized. The authors in [1] have discussed about few classifiers but finding an optimum classifier is still an open issue. In [2-4], the authors have discussed about some of the feature extraction techniques. In this paper, a thrust is given on the classifier based the neural network which uses the features derived from independent component analysis.

II. RELATED SOFT COMPUTING TECHNIQUES

The major challenge for gender classification is two-fold: (a) to find discriminating features, and (b) an optimum hyperplane for separating male and female classes. The following subsections elaborate these issues.

A. Feature Extraction using ICA

Face image contains a lot of redundant information within itself. Hence, a technique is required to remove this spurious information. Independent Component Analysis (ICA) is one of such approaches [5]. ICA is a computational method for separating a multivariate signal into additive subcomponents. It can be defined as: For the given set of input sample *x*, ICA finds a linear transform $s = Wx$ such that the components, s is as independent as possible. Here, W is an unknown separating matrix and needs to be determined. There exist several algorithms for determining *W* like *Jade*, *Information Maximization (infomax)* and *Fast fixed point (fast)* ICA. The scheme proposed in the paper is based on the infomax ICA which is described in the following subsection.

1) Information Maximization ICA

Infomax is a gradient based Infomax is a gradient based neural network and it maximizes information from input to the output network as proposed by Bell et al. [6]. The information maximization is achieved by maximizing the joint entropy of transformed vector where is a sigmoidal function. The joint entropy is given by:

$$H(y) = -E[\ln f(y)]$$

Where $f(y)$ is the multivariate joint density function of y.

$$f(y) = \frac{f(x)}{|J_W|}$$

Here, $|J_W|$ denotes the absolute value of Jacobian matrix J_W, which is defined as

$$J_W = \det\left[\frac{\partial y_i}{\partial x_i}\right] i$$

On combining the above equations, $H(y)$ can be written as

$$H(y) = E[\ln|J_W|] + H(x)$$

Maximization of $H(y)$ can be achieved by adapting W and can be achieved using only the first term.

2) Probabilistic Neural Network

The Probabilistic Neural Network (PNN) is developed by D. Specht [7] is a network formulation of probability density functions. It is defined as an implementation of Kernel Discriminant Analysis in which the operations are organized into multilayer feed forward neural network (shown in Fig. 1). It uses the bayes's rule for classification. The main advantages of PNN which discriminate from others are: Fast training process, an inherently parallel structure, the training set can be added or removed without any extensive training. The basic operation performed by the PNN is an estimation of probability density function of features of each class from the provided training samples using Gaussian Kernel. The PNN structure consists of neurons in four different layers.

- Input layer: The number of nodes in input layer corresponds to the dimension of face image where each node consists of pixel values.
- Pattern layer: Each pattern node corresponds to one training example and it gives an estimation of probability density functions for the test image with respect to an individual train image. The PDF for each training node can be calculated as: $\frac{1}{\sigma}W(\frac{x-x_k}{\sigma})$ where, x is an unknown (test) sample and x_k is the k_{th} training sample, W is the weighting function and σ is the smoothing parameter.
- Summation layer: It represents the PDF for each class i.e. PDF for male and female class which is the average PDFs of male samples and same for PDF for female class also. Thus, summation layer consists of two nodes: First represents the PDF for male class and second node represents PDF for the female class. It can be calculated as: $\frac{1}{n\sigma}\sum W(\frac{x-x_k}{\sigma})$. Gaussian function is used as an activation function here.
- Output layer: The output nodes are two input neurons which produces binary outputs related to w_r, w_s when $r \neq s$ and $r,s = 1,2\cdots q$. It uses the classification criteria:

$$\sum \exp[-(W_i - x)^T * (W_i - x)/\sigma^2] > \sum \exp[-(W_j - x)^T * (W_j - x)/\sigma^2]$$

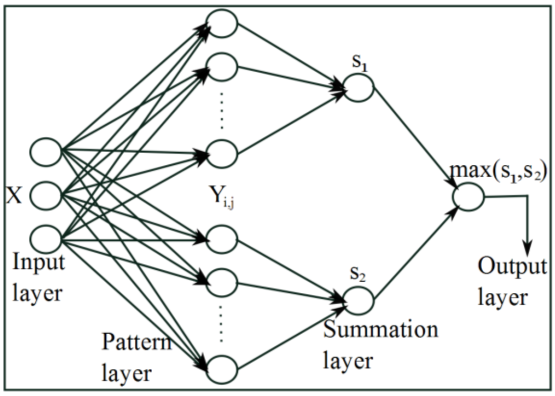

Figure 1. Architecture of Probabilistic Neaural Network

B. K-means Clustering

K-means clustering is a method of cluster analysis which partitions n observations into k clusters in which each observation belongs to the cluster with the nearest mean [8]. The algorithm contain following steps:

1. Choose K clusters. It represents initial group centroids.
2. Assign each object to the group that has the closest centroid.
3. When all objects are assigned, recalculate the positions of the K centroids.
4. Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

III. PROPOSED APPROACH

In this paper, authors have contributed their efforts in testing the performance of PNN for gender classification. Later, its performance is improved using k-means clustering. K-means clustering acts as a feature selection technique as it finds the specified number of clusters among independent components. A block diagram of the proposed approach is shown in Fig. 2.

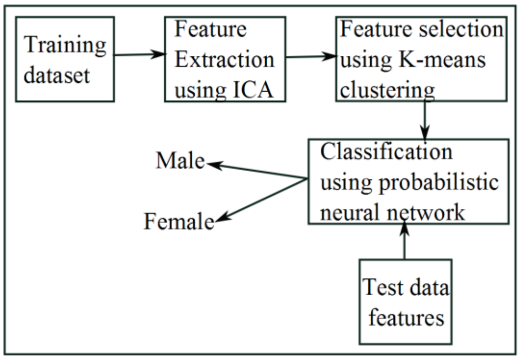

Figure 2. Proposed gender classification approach

IV. DATABASE DESCRIPTION AND INPUT SOURCE

Evaluation of the proposed method is carried out using FERET database [9]. Some of the sample images from this are shown in Fig. 3. The database consists of 500 images consisting 250 male and 250 female. 280 images (140 male and 140 female) are randomly selected to form training set, and another 120 face images (60 male and 60 female) are chosen for testing.

The original size of images is 384×256. These images first have been aligned, and then conventional triangularization approach is applied to localize the region of interest as shown in Fig. 4. Size of each cropped face image is 293×241. Further, to reduce the complexity of the system, all face images were resized to 118×97. ICA is applied on these low resolution face images to give input to the classifier.

Figure 3. Sample images from FERET database

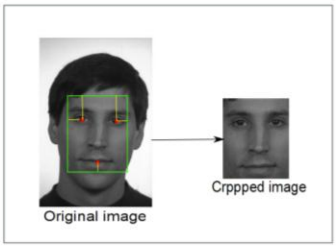

Figure 4. Conventional triangularization approach for face cropping

TABLE I. RESULTS FROM CLUSTERED PNN (INDEPENDENT COMPONENTS ARE VARIED IN THE RANGE OF 10-150 AND K IS VARIED FROM 2-20 AND CORRESPONDING ACCURACIES ARE SHOWN (- INDICATES THE FAILURE OF CLUSTER CREATION)

K# ↓	**Number of Independent Components (ICs) →**														
	10	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**150**
2	65.37	90	86.67	70	63.33	60	73.33	70	71.67	71.67	65	68.33	70	71.67	68.33
3	61.67	86.67	85	66.67	68.33	75	68.33	70	85	65	70	70	65	93.33	90
4	88.33	95	83.33	70	63.33	65	71.67	63.33	78.33	85	86.67	65	60	85	90
5	81.67	83.33	90	75	63.33	76.67	61.67	66.67	73.33	68.33	68.33	80	66.67	86.67	83.33
6	-	73.33	83.33	73.33	63.33	60	73.33	75	71.67	63.33	63.33	68.33	66.67	76.67	73.33
7	-	80	85	70	76.67	71.67	63.33	61.67	88.33	76.67	76.67	68.33	61.67	76.67	70
8	-	80	-	68.33	68.33	65	70	75	75	68.33	68.33	70	68.33	80	80
9	-	-	-	80	78.88	68.33	78.33	78.33	73.3	78.33	78.33	70.33	73.33	83.33	78.33
10	-	-	-	78.33	75	73.33	63.33	76.67	75	73.33	73.33	63.33	73.33	83.33	78.33
11	-	-	-	-	71.67	68.33	88.33	78.33	76.67	76.67	76.67	68.33	68.33	86.67	76.67
12	-	-	-	-	71.67	75	83.33	85	65	68.33	68.33	73.33	65	75	75
13	-	-	-	-	76.67	75	78.33	85	73.33	75	75	75	65	61.67	86.67
14	-	-	-	-	76.67	75	85	76.67	80	68.33	68.33	76.67	81.67	61.67	81.67
15	-	-	-	-	76.67	66.67	83.33	80	75	66.67	66.67	80	81.67	83.33	81.67
16	-	-	-	-	-	65	83.33	86.67	73.33	83.33	83.33	76.67	81.67	83.33	81.67
17	-	-	-	-	-	78.33	73.33	86.67	-	83.33	83.33	78.33	81.67	86.67	81.67
18	-	-	-	-	-	80	73.33	86.67	-	81.67	81.67	83.33	90	86.67	83.33
19	-	-	-	-	-	-	71.67	73.33	-	81.67	81.67	83.33	90	88.33	83.33
20	-	-	-	-	-	-	71.67	73.33	-	81.67	81.67	81.67	81.67	88.33	83.33
*****	93.33	90	83.33	80	90	83.33	83.33	71.67	71.67	86.67	80	78.33	86.67	83.33	75.83

* Accuracy without clustering

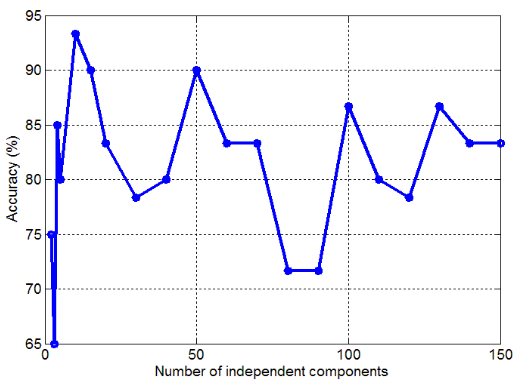

Figure 5: Accuracy obtained by probabilistic neural network with varying number of independent components

V. EXPERIMENTAL EVALUATION

A. *Results from PNN*

PNN is used to classify male and female classes from the independent components derived from those images. The maximum accuracies obtained from PNN are shown in Fig. 5. The different maximum accuracy values are obtained on varying the independent components from 2–150. But PNN works well with low dimensional dataset and fails with the high dimensional dataset. Hence the maximum accuracy obtained is 93.33% at 10 ICs but as the ICs are increased the accuracy is degrading.

B. *Results from Clustered PNN*

As PNN doesn't work well with the high dimensional data set, thus a feature selection technique is used to select the most significant coefficients among the all present independent coefficients. In this approach, k-means clustering is used to find the specified number of cluster center of independent coefficients. The centers are chosen from 2-20 on ICs 10-150. The outcome of this analysis is given in Table. I.

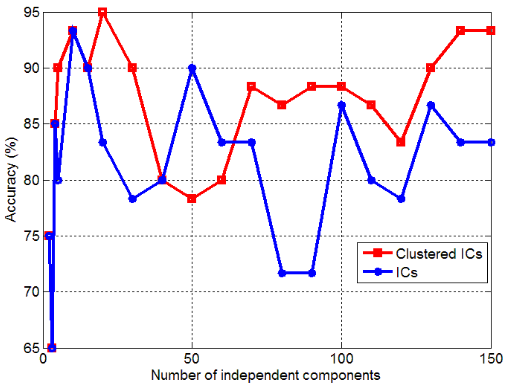

Figure 6: A comparison between PNN and clustered PNN with the variance of no. of independent components

The comparison between these two techniques is shown in Fig. 6. The maximum accuracy obtained is 95% at number of ICs = 20 and number of clusters (K#) = 4. Here, the cluster centers are varied from 2-10 as shown in Fig. 7.

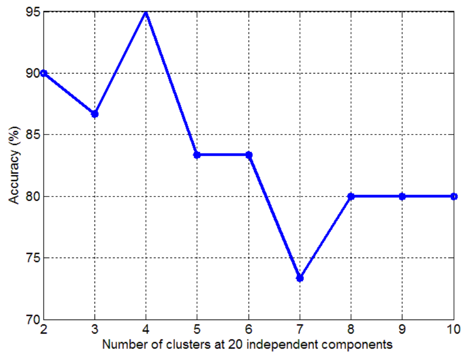

Figure 7: Accuracy obtained by probabilistic neural network using 20 independent components

σ is the smoothing parameter of PNN. The accuracy of PNN is obtained on varying this in the range of 0–1 and at 1.5–2.4; the maximum accuracy is obtained as shown in Fig. 8.

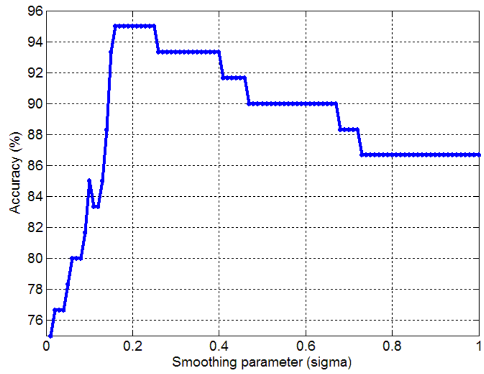

Figure 8: Nature of PNN with varying σ

VI. Conclusions and Future Work

The authors have proposed a method for gender classification using probabilistic neural network as classifier and gender features are extracted using infomax ICA. However this approach does not yield satisfactory accuracy. So, the improvement in performance is achieved through using k-means clustering as a feature selection technique. Clustered ICs are subjected to PNN for reducing the classification time. The approach also produces more accuracy than non-clustered ICs. But the no. of clusters for which accuracy will be highest cannot be mathematically modeled, and has to be tested.

References

[1] B. Moghaddam, and M. H. Yang, "Learning gender with support faces," IEEE *Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, no. 5, pp. 707 – 711, May 2002.

[2] M. Bartlett, J. Movellan, and T. Sejnowski, "Face recognition by independent component analysis," *IEEE Transactions on Neural Networks*, Vol. 13, No. 6, pp. 1450 – 1464, Nov 2002.

[3] A. Jain, and J. Huang, "Integrating independent components and support vector machines for gender classification," Vol. 3, pp. 558 – 561, Aug 2004.

[4] M. Turk, and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71–86, 1991.

[5] A. Hyvrinen, and E. Oja, "Independent component analysis: algorithm and applications," *IEEE Transactions on Neural Networks*, Vol. 13, pp. 411 – 430, 2000.

[6] A.J. Bell, and T. J. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," Neural Computation, Vol. 7, pp. 1129 – 1159, 1995.

[7] D. F. Specht, "Probabilistic neural networks," Neural Networks, Vol. 3, pp. 109 – 118, Jan 1990.

[8] S. Haykin, "Neural networks: A comprehensive foundation," 2nd Ed., *Prentice Hall*, 1998.

[9] P.J.Phillips H.Moon, S.A.Rizvi, and P.Z.Rauss, "*The FERET evaluation methodology for face recognition algorithms," Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 1997, pp. 137 – 143.