World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:2, No:12, 2008

A New Weighted LDA Method in Comparison to Some Versions of LDA

Delaram Jarchi, and Reza Boostani

Abstract**—**Linear Discrimination Analysis (LDA) is a linear solution for classification of two classes. In this paper, we propose a variant LDA method for multi-class problem which redefines the between class and within class scatter matrices by incorporating a weight function into each of them. The aim is to separate classes as much as possible in a situation that one class is well separated from other classes, incidentally, that class must have a little influence on classification. It has been suggested to alleviate influence of classes that are well separated by adding a weight into between class scatter matrix and within class scatter matrix. To obtain a simple and effective weight function, ordinary LDA between every two classes has been used in order to find Fisher discrimination value and passed it as an input into two weight functions and redefined between class and within class scatter matrices. Experimental results showed that our new LDA method improved classification rate, on glass, iris and wine datasets, in comparison to different versions of LDA.

Keywords**—**Discriminant vectors, weighted LDA, uncorrelation, principle components, Fisher-face method, Bootstarp method.

I. Introduction

RECENTLY, many linear discriminant methods are proposed and applied to face recognition problem [1]-[2]-[4]. Image is a high-dimension data, therefore, other dimension reduction techniques are also combined with linear discrimination methods [4]. For example, The Fisher-face method [4] combines PCA and the Fisher criterion to extract the information that discriminates the classes of a sample set [1]. Different versions of LDA have suggested improving classification rate. In an improved LDA approach by Yuan Jing[1], Three weaknesses of traditional LDA were investigated and then an improved LDA (ILDA) introduced that could decrease influence of those weaknesses. In another approach, a weighted LDA method is proposed [2]-[3]. The aim of weighted LDA approach is to alleviate influence of the outlier class (a class that is far away from other classes). There are two versions of weighted LDA, one redefines between class matrix by incorporating a weight function [2], the other, redefines within class matrix by incorporating a weight function [3]. In this paper, it has been tried to compare different versions of LDA method on standard datasets (multi class datasets) such as glass, iris, wine. It has been shown that those linear methods that are proposed for face recognition must change a little for doing well on other datasets. This paper is structured as follows: In section 2, a generalized improved LDA that is based on genetic algorithm is proposed which is a generalization form of improved LDA [1] to improve classification rate on other datasets. Then, in section 3, a new weighted LDA approach that uses a simple weight criterion to improve classification rate is suggested. Next, experimental results on three datasets: glass, iris and wine are shown. Finally, in the conclusion part different versions of LDA are compared.

II. Generalized Improved LDA

A. Improved LDA

In this section, an improved LDA method [1] is briefly described. Improved LDA method increased classification rate of traditional LDA method in three ways:

1) **Selection of discriminant vectors**: in this improvement, the discriminate vectors that possess more between-class than within class scatter information will be selected. Suppose the followed matrix:

$W_{opt}=[\phi_1, \phi_2, ..., \phi_r],$

where r is the number of discrimination vectors, and ϕ_i is the i'th discriminant vector.

For $\forall \phi_i$ (i=1,…,r), we have :

$$\phi_i^T S_t \phi_i = \phi_i^T S_b \phi_i + \phi_i^T S_w \phi_i .$$

If $\phi_i^T S_b \phi_i > \phi_i^T S_w \phi_i$, Then

$$F(\phi_i) = \frac{\phi_i^T S_b \phi_i}{\phi_i^T S_t \phi_i} > 0.5.$$

After obtaining the discriminate vectors, it should be found the Fisher value of each discriminant vectors, then, select those discrimination vectors whose Fisher discrimination values are more than 0.5, and discard the others.

2) **Statistical uncorrelation of discriminant vectors**: discrimination vectors should be made to satisfy the statistical uncorrelation, a favorable classification property. Although UODV satisfies this requirement, it also uses more computing time than the Fisher-face method, since it respectively calculates every discrimination vector satisfying the constraint of uncorrelation. This improvement will take advantages of both the Fisher-face method and UODV [1].

The routine to make discriminant vectors satisfy the statistical uncorrelation is as follows:

Step 1) Use the Fisher-face method to obtain the discrimination vectors:

$(\phi_1, \phi_2, ..., \phi_r)$.

If the corresponding Fisher values $(\lambda_1, \lambda_2, ..., \lambda_r)$ are unequal mutually, over; else go to the next step.

Step 2) For 2<=k<=r , if $\lambda_k \neq \lambda_{k-1}$, then keep ϕ_k , else replace ϕ_k by φ_k from UODV.

3) **Selection of principle components**: an automatic strategy for selecting principal components should be established. This would effectively improve classification performance and further reduce feature dimension [1]. The routine to select effective principle components is as follows:

Step 1) Discard the smallest C components (C is the number of classes).

Step 2) Compute the Fisher discrimination values of the remainder components according to equation below:

$$J_i = \frac{\beta_i^T S_b \beta_i}{\beta_i^T S_t \beta_i}, (1 \leq i \leq p)$$

β_i is i'th principle component and p is the number of principle components. Then, rank them in descending order and calculate the sum of their Fisher discriminability values J_{all} .

Step 3) Select the components with the first largest J_i values until a threshold T is satisfied, where T is the ratio of the sum of their values to J_{all} .

This method is used for face recognition problem and could improve classification rate. But when it applied for another dataset, such as glass dataset, some problems are appeared. One of the problems is, what should be the best value of T. It has been thought that with simple programming techniques the best value of T can be chosen. So, we started to find the best T to gain the best classification rate for glass dataset. Our experiment on the glass dataset showed that it can not lead to the best recognition rate in comparison with other linear methods even with the best T. The key was to change some of the other parameters' value. Improved LDA method discards the smallest C principle components (C is the number of classes). But in all of applications this may not works well. Because some of the principle components with smaller eigen values may have discriminant information and it should not be discarded. Then, it is decided to discard D smalles principle components (D <= C). In this way, by getting the best value for T and D, an improved LDA can apply to any other dataset in order to improve classification rate. Our experiment showed that after training phase, in order to classify test samples, using k-nearest-neighbor classifier instead of nearest neighbor classifier will improve classification rate.

To classify a test sample based on k-nearest neighbor classifier, first k nearest training samples to the test sample (based on Euclidean distance) will be determined, then, a class that has the most training samples, among the k nearest training samples from all classes, is considered as label of the test sample.

The problem is how to find value of the best k that will result in the best recognition rate. Before, we also had two parameters (D and T) that their values should be chosen so that the best recognition rate will be resulted, and now a third parameter, k, is added too. It has been suggested to employ genetic algorithm in order to determine value of D, T and K. Each individual in the population consists of three values that all are encoded into a unit binary code. Each individual's value corresponds to value of T , D and K. We set population size equal to 2 times of number of examples.

B. Genetic Algorithm

Genetic algorithm has been proposed Goldberg [8]-[9], which is a versatile tool for many applications. This algorithm is briefly described as follows: First parameters that are going to be optimized are converted to binary codes which are called chromosomes. Then a random population is created aligned with our chromosome size. Cross over operation is applied in order to generate a new generation. The best chromosomes in each generation are selected by the fitness function. Here, fitness function is selected as the value of classification rate. Mutation rate is considered 0.1. Selection method is K way-Tournament method (K is 0.1 of population size).

C. Validation

Our data validation method is a resampling method, because the number of samples in some classes is low. We used bootstrap method [7] that is a resampling technique with replacement. Bootstrap method can be described in the following steps:

Step 1) From a dataset with N examples Randomly select (with replacement) N examples and use this set for training.

Step 2) The remaining examples that were not selected for training, are used for testing. This value is likely to change from fold to fold.

Step 3) Repeat this process for a specified number of folds (K) .

Step 4) The true error is estimated as the average error rate on test examples.

In our research, bootstrap method with K=10 is used.

III. New Weighted LDA Method

Traditional LDA in a multiclass problem has some weaknesses. One of the problems is that it dose not consider every two class conjunctions and just cares about those pairs that are far away from each other. As a result , a suboptimal solution will be obtained. Second, it is based on the assumption that all the classes have the same within-class scatter covariances [3], which is the mean of the all the actual within-class scatter covariances. From the statistical point of view, even when the covariances are different from class to class, using the mean as a uniform within-class scatter covariance is optimal[3]. But still a problem exists. All the classes including the outlier class have the same impact on estimating the uniform within-class scatter covariance.

The aim in weighted LDA method is to alleviate the role of an outlier class. Some researches [2]-[3] have been done to achieve this aim, one of them is redefining between class scatter matrix. In this research those classes that are well separated should have little influence on classification. So, in redefining between class scatter matrix, those classes that are far away from each other will assign a smaller weight on their between class scatter matrix.

The other idea is to redefine within class scatter matrix. Here, the role of an outlier class must be alleviated too. This can be done by incorporating a weight into within class scatter. So, if the classes are more concentrated (their within class are small), it must be assigned a smaller weight into their within class scatter matrix. Our new weighted LDA method considers both redefining between class and within class scatters matrices. The question is how to use an appropriate weight. In the previous researches several weight measures have tested: Euclidean distance, Mahalanobis distance and estimated Bayesian error. In this paper, we introduce a new and simple method for the weight.

A. Weight Measure

We want to determine how far every two classes are, then use this distance measure as a weight. The weight which tells us how distinct can be two classes; it might be a better weight measure. The previous weights measure, that were based on the distance might not be always the best measure, here, it is used a criterion that measures the class separability. For every two classes it is used the traditional LDA method to find a W which maps the original space into a new space. Because LDA used for two classes so that W maps every two classes into a line that those two classes are well separated. We can use the Fisher criterion of the determined W as a measure that says how much separable are two classes. It is sufficient to find Fisher criterion of the determined W. This value is a representation of class separability measure and is used as an input to the weight function. Now, we introduce two weight functions, one for between class and the other for within class scatter matrices.

B. Redefining Between Class Scatter Matrix

We use the between class scatter matrix definition that is based on Loog's approach [5]:

$$S_b = \sum_{i=1}^{C-1}\sum_{j=i+1}^{C} P_i\, P_j\, (\mu_i - \mu_j)(\mu_i - \mu_j)^T$$

P_i is a prior probability for class i and P_j is a prior probability for class j.

By incorporating weight into between class scatter matrix we can have a new formula for between class scatter matrix:

$$S_b = \sum_{i=1}^{C-1}\sum_{j=i+1}^{C} n_i\, n_j\, w(\Delta_{ij})(\mu_i - \mu_j)(\mu_i - \mu_j)^T$$

$$w(\Delta_{ij}) = \frac{1}{\Delta_{ij}}$$

Δ_{ij} is Fisher discrimination value of the resulted W^*

$$W^* = \underset{w}{\operatorname{argmax}}\{\frac{W^T S_B W}{W^T S_t W}\} = S_t^{-1}(\mu_i - \mu_j)$$

$$\Delta_{ij} = \frac{W^{*T} S_B W^*}{W^{*T} S_t W^*}$$

C is the number of classes, n_i is number of training sample for class i, n_j is number of training sample for class j and

$$S_t = S_B + S_W\,.$$

Note that we can use the following formula for between class scatter matrix, but we need to multiply the weight fuction by N:

$$S_b = \frac{1}{N}\sum_{i=1}^{C-1}\sum_{j=i+1}^{C} n_i\, n_j\, w(\Delta_{ij})(\mu_i - \mu_j)(\mu_i - \mu_j)^T$$

$$w(\Delta_{ij}) = \frac{N}{\Delta_{ij}}$$

where, N is number of training samples.

C. Redefining within Class Scatter Matrix

Within class covariance matrix of class i is defined as:

$$S_{W_i} = \frac{1}{k}\sum_{j=1}^{k}(x_j - m_i)(x_j - m_i)^T$$

Total within class covariance is brought below:

$$S_W = \sum_{i=1}^{C} S_{W_i}\;,$$

where, C is the number of classes.

By incorporating weight into within class scatter matrix we can have a new formula for within class scatter matrix:

$$S_w = \sum_{i=1}^{C} P_i\, r_i\, S_{W_i}\;,$$

where, r_i is :

$$r_i = \frac{1}{\sum_{j\neq i}\Delta_{ij}}$$

Δ_{ij} is Fisher discrimination value of the resulted W^*

$$W^* = \underset{w}{\operatorname{argmax}}\{\frac{W^T S_B W}{W^T S_t W}\} = S_t^{-1}(\mu_i - \mu_j)$$

$$\Delta_{ij} = \frac{W^{*T} S_B W^*}{W^{*T} S_t W^*}$$

where, $S_t = S_B + S_W$.

Based on an improved LDA [1] to overcome small sample size problem, instead of S_W, S_t (S_t is sum of S_B and S_W) is used in denominator of Fisher criterion. Now, by having a new formula for between class scatter matrix and within class scatter matrix, discriminant vectors matrix (W) can be obtained as follows:

$$W = \underset{w}{\operatorname{argmax}}\{\frac{W^T S_b W}{W^T S_t W}\}$$

$$S_t = S_b + S_w\,.$$

In our new method, experimental results showed that using all the discriminant vectors of W can lead to a better recognition rate, The reason is that the rank of W is increased in the weighted LDA method.

Note that S_B and S_W are traditional between class and within class scatter matrices, whereas, S_b and S_w are new between class and within class scatter matrices.

IV. Results

We tested our two new methods on different datasets such as glass, iris and wine. For having a comparison between different versions of LDA, we also tested other methods simultaneously. As explained in part II the data validation method is bootstrap method. We tested other linear methods such as:

LDA+Knn, LDA+PCA+Knn, Generalized improved LDA+Knn, New weighted LDA+PCA+knn and New weighted LDA+knn .

Results are shown in Fig. 1, 2 and 3 for glass, iris and wine dataset respectively.

For generalized improved LDA in each bootstrap iteration resulted T, D (number of discarded eigen vectors with small eigen values) and K are summarized in Table I, Table III and Table V, for glass and iris and wine dataset. Recognition rate of each method can be estimated as the average recognition rate of 10 bootstrap iterations. Table II, Table IV and Table VI shows recognition rate of all methods on glass, iris and wine dataset.

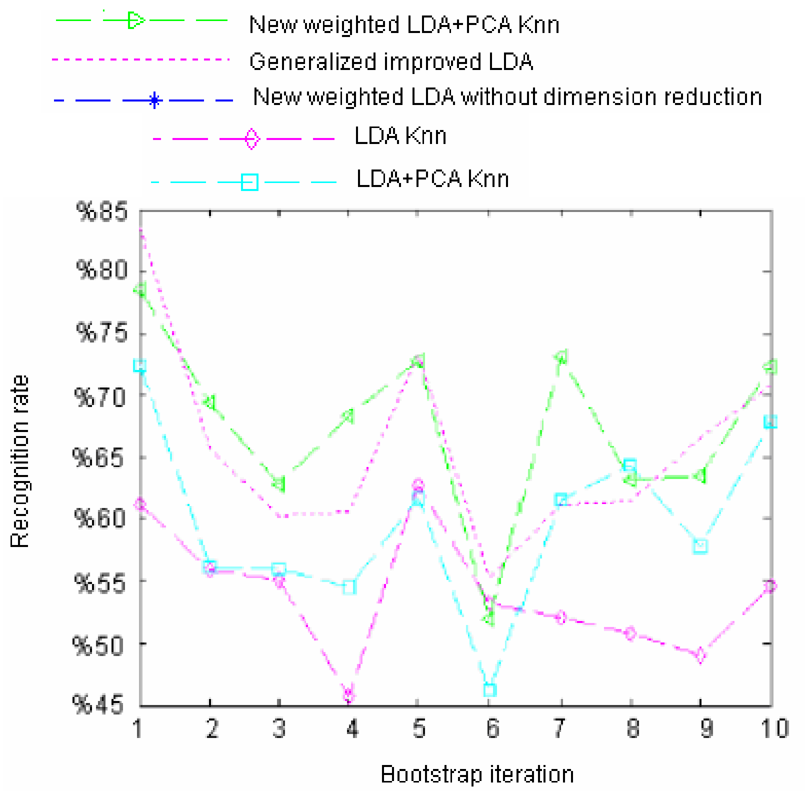

Fig. 1 Comparison of different LDA methods' recognition rate on glass dataset using bootstrap method

TABLE I
Resulted T, D and K in each Bootstrap Iteration for Glass Dataset

	1	2	3	4	5	6	7	8	9	10
T	0.9531	0.7969	0.9688	0.9375	0.8594	0.9688	0.9531	0.8906	0.7188	0.9219
Num. of discarded eigen vectors (D)	2	0	2	2	2	4	1	3	3	2
K (Knn)	1	8	1	1	1	1	1	1	1	1

TABLE II
Estimated Recognition Rate of Different Methods on Glass Dataset

	Recognition rate
New weighted LDA+PCA+knn	67.59863
Generalized improved LDA	65.88589
LDA+PCA+Knn	59.86484
LDA+Knn	54.06780

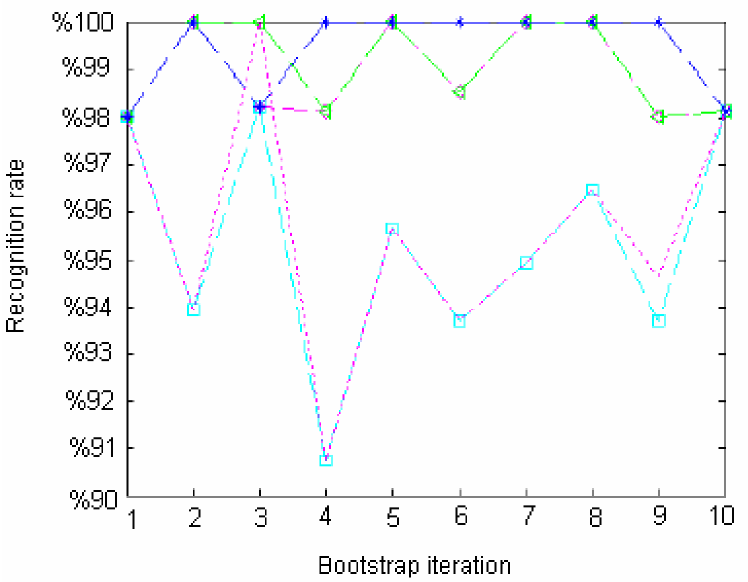

Fig. 2 Comparison of different LDA methods' recognition rate on iris dataset using bootstrap method

TABLE III
RESULTED T , D AND K IN EACH BOOTSTRAP ITERATION FOR IRIS DATASET

	1	2	3	4	5	6	7	8	9	10
T	0.8594	0.2969	0.9844	0.0781	0.5625	0.0313	0.3906	0.3594	0.9844	0.2656
Num. of discarded eigen vectors (D)	2	1	1	1	0	1	3	2	0	1
K(Knn)	24	28	3	54	17	45	10	60	11	27

TABLE IV
ESTIMATED RECOGNITION RATE OF DIFFERENT METHODS ON IRIS DATASET

	Recognition rate
New weighted LDA+knn	99.44330
New weighted LDA+PCA+knn	99.09255
Generalized improved LDA	95.62437
LDA+PCA+Knn	95.35516
LDA+Knn	98.91711

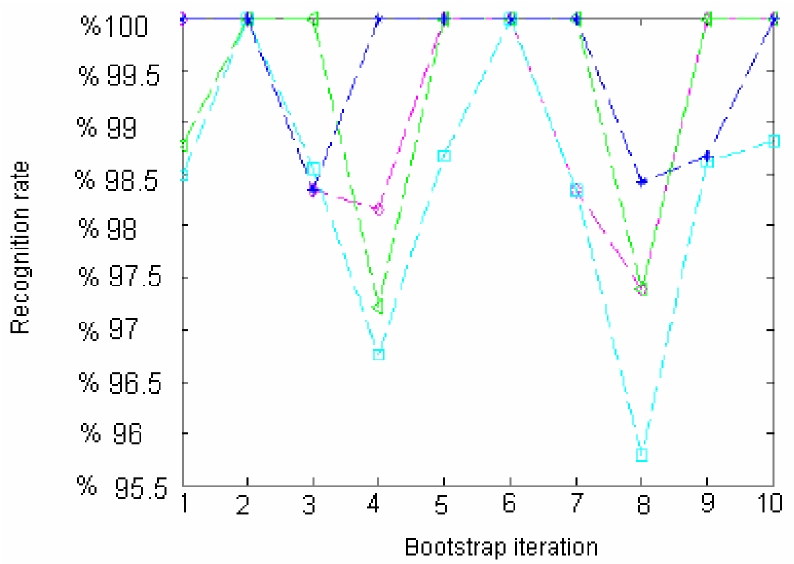

Fig. 3 Comparison of different LDA methods' recognition rate on wine dataset using bootstrap method

TABLE V
RESULTED T, D AND K IN EACH BOOTSTRAP ITERATION FOR WINE DATASET

	1	2	3	4	5	6	7	8	9	10
T	0.9063	0.9531	1	1	0.8594	0.8438	0.9219	0.8906	0.9844	0.0938
Num. of discarded eigen vectors (D)	3	1	2	0	1	1	2	2	1	2
K(Knn)	2	39	31	47	19	7	42	22	3	33

TABLE VI
ESTIMATED RECOGNITION RATE OF DIFFERENT METHODS ON WINE DATASET

	Recognition rate
New weighted LDA+knn	99.54127
New weighted LDA+PCA+knn	99.33587
Generalized improved LDA	81.31251
LDA+PCA+Knn	98.39992
LDA+Knn	99.21858

V. CONCLUSION

In this paper a new weighted LDA method is proposed that could improve recognition rate based on increasing rank. K-nearest neighbor classifier is used in our research and the best value of k is determined by genetic algorithm. Glass dataset contains 6 classes while, iris and wine contains 3 classes. It can be seen that our new weighted LDA could improve recognition rate of classification on glass dataset more than the two other datasets. So, when we have a large number of classes, using weighted LDA method can be led to better results.

Iris, glass, and wine datasets have 4, 9, and 13 dimensions respectively. Results show that when the number of dimensions is low using PCA to reduce dimension can not lead to a better results. For glass dataset new weighted LDA with dimension reduction (PCA) is the best method. For iris and wine datasets, new weighted LDA without dimension reduction is the best, however, improvement in comparison with traditional LDA is not remarkable.

REFERENCES

[1] Xiao-Yuan Jing, David Zhang, and Yuan-Yan Tang ," An improved LDA Approach", IEEE Transaction on Syatems, Man, And Cybernetics—Part B: Cybernetics, VOL. 34, NO. 5, October 2004.

[2] Yu Bing. Jin Lianfu. Chen Ping,"A new LDA-based method for face recognition", Proceedings of 16th International Conference on Pattern Recognition, Volume 1, 11-15 Aug. 2002 Page(s):168 - 171 vol.. 1.

[3] Tang, E.K. Suganthan, P.N. Yao, X, "Generalized LDA Using Relevance Weighting and evolution strategy", Congress on Evolutionary Computation, 2004. CEC2004. Volume 2, 19-23 June 2004 Page(s): 2230 - 2234 Vol. 2.

[4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherface: Recognition using class specific linear projection," IEEE Trans. Pattern Anal. Machine Intell., vol. 19, pp. 711–720, July 1997.

[5] M. Loog, R.P.W. Duin and R. Haeb-Umbach, "Multiclass linear dimension reduction by weighted pairwise fisher criteria", IEEE Transaction on Pattern Analysis and Machine Intelligence,23(7),2001,00.762-766.

[6] A. M. Martinez and A. C. Kak, "PCA versus LDA," IEEE Trans. Pattern Anal. Machine Intell., vol. 23, pp. 228–233, Feb. 2001.

[7] Chernick M.R., *Bootstrap Methods: A Practitioner's Guide*, John Wiley and Sons, New York, 1999.

[8] Goldberg, D.E. "Genetic algorithms as a computational theory of conceptual design". In Applications of Artificial Intelligence in Engineenng. Vol. 6, 1991, pp, 3-16.

[9] David E. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning", Addison-Wesley Longman Publishing Co., Inc., Boston, MA, 1989.

Delaram Jarchi was born in Tehran, Iran, in 1983. She received her B.Sc. degree in Computer Engineering from Iran University of Science and Technology (IUST), in 2005. Now she is an M.Sc. student of Artificial Intelligence at Shiraz University.
Her research interests include pattern recognition, classification and signal processing.

Dr. Reza Boostani was born in 1973, he has got his B.Sc. in Electronic from Shiraz university in 1996, and got his M.sc. and Ph.D. in Biomedical Engineering from Tehran Polytechnic University in 1999 and 2004 respectively. He was a head of speaker verification project in ITRC company from 2001 to 2002. He has spent his research period in Graz University of technology in BCI field from 2002 to 2003. Since 2004, he is a faculty member of computer science and Eng. department of Shiraz University. His current research interests include biomedical signal processing, statistical pattern recognition and machine learning.