PERGAMON

Computers in Biology and Medicine 28 (1998) 275–287

Computers in Biology and Medicine

Visual classification of medical data using MLP mapping

Emin Çağatay Güler[a], Bülent Sankur[b,*], Yasemin P. Kahya[b], Sarunas Raudys[c]

[a]Biomedical Engineering Institute, Boğaziçi University, Bebek, 80815, Istanbul, Turkey
[b]Department of Electronic Engineering, Boğaziçi University, Bebek, 80815, Istanbul, Turkey
[c]Institute of Mathematics and Informatics, Akademijos 4, 2600, Vilnius, Lithuania

Received 10 October 1997; received in revised form 13 March 1998

Abstract

In this work we discuss the design of a novel non-linear mapping method for visual classification based on multilayer perceptrons (MLP) and assigned class target values. In training the perceptron, one or more target output values for each class in a 2-dimensional space are used. In other words, class membership information is interpreted visually as closeness to target values in a 2D feature space. This mapping is obtained by training the multilayer perceptron (MLP) using class membership information, input data and judiciously chosen target values. Weights are estimated in such a way that each training feature of the corresponding class is forced to be mapped onto the corresponding 2-dimensional target value. © 1998 Elsevier Science Ltd. All rights reserved.

Keywords: Mapping; Visual displays; Multilayer perceptrons; Respiratory sound patterns

1. Introduction

In the interdisciplinary field of biomedical engineering, medical doctors may often work in cooperation with informatic engineers. One of the aims of such a cooperation is to design “wizards” of data analysis and classification for physicians to be used in the diagnosis and evaluation of the pathologies.

Diagnostic tools have biological signals as inputs, and utilize typically lower level signal processing tools such as spectral analysis, transforms, parameter estimation, discriminant

* Corresponding author.

0010-4825/98/$19.00
PII: S0010-4825(98)00010-9

analysis, etc. In addition they can make explicit use of knowledge bases, heuristic rules, expert interactions to classify biological data [1–3]. Among these diagnostic tools, there is often a need for a classification technique that appeals to the intuition of the physician, that could potentially incorporate his personal rules and knowledge. Such interactions are instrumental in building up confidence and plausibility with these diagnostic aids. More explicitly, interactive classification techniques are needed that allow a physician to interpret the results easily and efficiently. Otherwise, without such interactions, simple classification percentages that the present pattern recognition methods yield, do not avail the physician of the intuitive grasp of the problem.

The interaction of the physician becomes possible, if, for example, the data for classification is presented visually. To this effect mapping techniques provide a useful tool for the visualization of data. It has been proven [4] that interactive classifier design based on visual displays gives the user greater insight and confidence in the classification results. It is also well known that multivariate data projections can avoid the curse of dimensionality, enable better visualization of the underlying structure of the data, e.g., put into evidence its clustering tendency, especially for exploratory data analysis tasks. These techniques can be used as alternatives to formal classification methods in many fields but especially in biomedical engineering. By formal methods we simply mean those classification methods where the discriminant surfaces are obtained computationally without any expert interaction.

In this paper, the performance of the proposed mapping and visual classification method will be illustrated with an application on respiratory sounds.

2. Mapping techniques

Mapping techniques constitute the most frequently used methods for visualization of data, for visual assessment of potential structures and class dependencies, and for merging human judgement with formal data processing techniques. A mapping operation consists of a transformation of d-dimensional vectors onto a plane. The fidelity of a mapping method is defined as the amount of peculiarities retained after the dimensionality reduction. The mapping algorithm may be deficient in preserving all the information in d-dimensional data. Therefore the choice of the method becomes crucial in specific applications. Each mapping method is associated with a criterion, which is the measure of fidelity, to be optimized.

Mapping methods can be classified as linear and non-linear. In linear mapping techniques, linear transformations are used in order to map the d-dimensional pattern space into 2-dimensional space. On the other hand, a linear transformation that enables the transition from the multidimensional space to 2-dimensional space does not exist for non-linear methods.

Some of the mapping methods require for their operation that the user enter several parameters, prelabel the classes, cluster or classify the data first, it is then clear that the choice of the mapping that fits the data is not an easy matter. A summary of some features of the mapping techniques is given in Table 1 in order to clarify the differences between methods rather than to establish a priority among them.

The most widely used linear and non-linear mappings are briefly summarized below. A detailed survey of mapping techniques can be found in [4, 5].

Table 1
Comparative features of mapping techniques

Mapping method	Type	Best for	Prelabeling	Number of classes	User supplied parameters
Total principal component (PCM)	linear	Gaussian distribution (does not necessarily reveal the intrinsic structure of the data)	no	2	–
Class-conditional principal component	linear	Gaussian distribution (does not necessarily reveal the intrinsic structure of the data)	yes	2	–
Standardized class-conditional principal component	linear	Gaussian distribution (does not necessarily reveal the intrinsic structure of the data)	yes	2	–
Fisher direction	linear	Gaussian distribution	yes	2	–
Optimal discriminant plane	linear	Gaussian distribution	yes	2	–
Declustering	linear	Gaussian distribution	yes	2	–
Extended declustering (EDM)	linear	Gaussian distribution	yes	2	spread coefficient b
Least squares (LSM)	linear	Gaussian distribution (if the number of classes is not too high the map displays the clusters in a form resembling their original shape)	yes	≥2	–
Projection pursuit	linear	any distribution (*can be used as a test of non-normality*)	no	≥2	projection index p, cut-off radius R
Sammon's	non-linear	any distribution (*useful when the number of projected points is small compared to the dimension of the original space*)	no	2	number of iterations, parameter p
Triangulation	non-linear	any distribution	no	2	reference point
Distance from 2-means (DF2M)	non-linear	Gaussian distribution	yes	2	–
k-NN	non-linear	any distribution (just a tool for analyzing the performance of a *k-NN* classifier)	yes	≥2	k
Multilayer perceptron (MLP)	non-linear	any distribution	yes	≥2	targets on the plane, number of hidden layer nodes, desired error, maximum number of epochs, learning rate

2.1. Linear mappings

Given a set of d-dimensional feature vectors, $\{\boldsymbol{x}_n\}_{n\ =\ 1,\ldots,N}$, the linear map consists of the operation

$$\boldsymbol{y}_n = \boldsymbol{A}\boldsymbol{x}_n + \boldsymbol{b} \tag{1}$$

where $\boldsymbol{A} = [\boldsymbol{a}_1{}',\boldsymbol{a}_2{}']'$ is a 2-by-d transformation matrix, "′" denotes the transpose, and $\boldsymbol{b}$ is a two-component vector added for the generalization of the expression. The criterion to obtain $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ specifies the type of the linear mapping. One advantage of linear mappings is that straight lines in the projection plane correspond to hyperplanes in the higher dimensional space, so that piecewise linear boundaries drawn by the user can be reflected back into decision hypersurfaces in the original d-dimensional data space.

Well known and widely used linear mapping techniques are briefly summarized below.

(i) *The principal component mapping* is a technique where the principal axes of the sample covariance matrix in the original space is used. Principal axes are defined as the eigenvectors of the covariance matrix corresponding to its largest eigenvalues. Three versions of mappings exist depending upon which covariance matrix is being considered. If the total covariance matrix of the original d-dimensional data without a prespecified labeling on the classes is used, the resultant mapping is called the total principal component mapping. On the other hand, in the standardized class conditional mapping and in the class conditional principal component mapping, two classes exactly are required, and therefore two class conditional covariance matrices are needed to solve the eigenproblem. The common point in principal component mappings is that the minimum of the squared error criterion preserves the maximum information in the projection data.

(ii) *The generalized declustering mapping* is another linear mapping method for which Fisher direction, optimal discriminant plane, declustering mapping and extended declustering mappings represent the four different versions in this group. In all these mappings the Fisher discriminant is used to enhance the separability between pairs of classes. This method can only handle two classes in the feature space.

(iii) *The least squares mapping* is another method that requires prelabeling of the classes in the original feature space. One can specify more than two classes a priori. If, for example, there are C classes in the feature space, C class centers are specified on the plane, and transformation parameters are estimated under the constraint that the images of the feature vectors on the plane form minimum variance classes around the preselected class centers. This minimum variance criterion tends to create Gaussian-like configurations of sample points in two dimensions. While this subspace derived from the original space best discriminates Gaussian classes, in extreme cases it may fail.

(iv) *The projection pursuit mapping* which is first proposed by Friedman and Tukey [6] assumes the projection "produces a dense cluster of points while maintaining the overall spread of the data". This method does not make use of prelabeling of the data.

2.2. Non-linear mappings

In non-linear mapping methods, the coordinates of $\boldsymbol{y}_n$ are not linearly related to the coordinates of the original points $\boldsymbol{x}_n$ in a d-dimensional space, i.e., a simple analytical expression to tie the points in the plane and the corresponding points in the original data space is not available. The transformation that one searches for is a mapping of the d-dimensional data onto a plane, which for the sample $\boldsymbol{x}_n$, can be written as:

$$[z_{n1},z_{n2}]' = f([x_{n1},x_{n2},\ldots,x_{nd}]'), \quad n = 1,2,\ldots,N \tag{2}$$

in such a way that each class is transformed into one or more clusters. Non-linear mapping methods can be divided into four groups.

(i) *Sammon's non-linear mapping* projects the data space onto a plane by preserving the distances between original feature vectors in two dimensions. In other words, the sample points are projected on the plane in such a way that the distances between them are proportional to the distances between respective original points. The performance of Sammon's mapping depends upon the sample size and the dimensionality of the data. Its efficiency decreases with increases in dimensionality and sample population. Therefore Sammon's mapping is useful especially when the population is small compared to the dimension of the database.

(ii) *The triangulation mapping* also tries to preserve distances between samples as in Sammon's mapping. The most important difference, however, is that the triangulation mapping exactly preserves the distances of a subset of data, while it disregards completely the distances of the remaining sample set.

(iii) *The distance from two means mapping* is based on the idea of the Mahalanobis distances of a feature vector to the means of the two classes. This method can handle only two data classes.

(iv) *The k-nearest neighbor mapping* is just a tool for analysing and improving the performance of a k-nearest neighbor classifier.

The most important advantage of mapping methods in decision problems is the facility of graphical tools or ability of human designers to identify patterns and to create robust separating hypersurfaces. One disadvantage common to all mapping methods in decision making is that each mapping method has its own specific projection, which in turn tends to insert subjective bias to the results. Another aspect of mapping methods is that the interaction of human users can turn into a disadvantage due to their subjective biases. These disadvantages can be alleviated, however, by re-examining the data with the aid of more than one method, in other words one can cross-validate the classification results. Recall that, because of their inherent differences, different mapping methods produce distinct displays.

3. A new mapping method: multilayer perceptron mapping

Our mapping method consists in training an MLP (multilayer perceptron) in order to estimate and implement the best non-linear mapping from d-dimensional data to 2-dimensional targets. In training this perceptron, one may opt to assign more than one target value for each class as illustrated in Fig. 1. Target values guide the original data to locations where they will be transformed in the plane, in other words, the target values act as attractors for their class data. The use of MLPs to emulate a non-linear function is not new. However the novelty of our approach is that class membership information is used in training this mapping, which improves the separability between the classes.

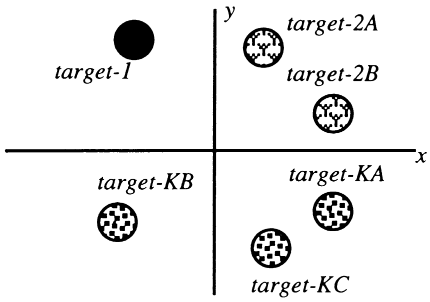

Fig. 1. Target values that can be chosen in implementing our MLP based mapping method. Note that, one may choose more than one target for each of the K classes. Target values only for classes 1, 2 and K are shown respectively, with 1, 2 and 3 target values.

We illustrated the differences between MLPs employed for mapping and classification purposes, respectively, in Fig. 2. The simplest MLP classifier consists of an input layer and two layers of processing neurons as shown in Fig. 2b. Each neuron at the output layer is assigned to one class. On the other hand, an MLP used for data mapping has only two neurons at the output (Fig. 2a). Notice that in the classification problem the desired output values for the K output neurons, $O_1, \ldots, O_K$, are:

	O_1	O_2	O_3	...	O_K
for class-1	1	0	0	...	0
for class-2	0	1	0	...	0
	...				
for class-K	0	0	0	...	1

On the other hand, in the data mapping case, the two outputs, z_1, z_2, for a K class example problem can be denoted as follows:

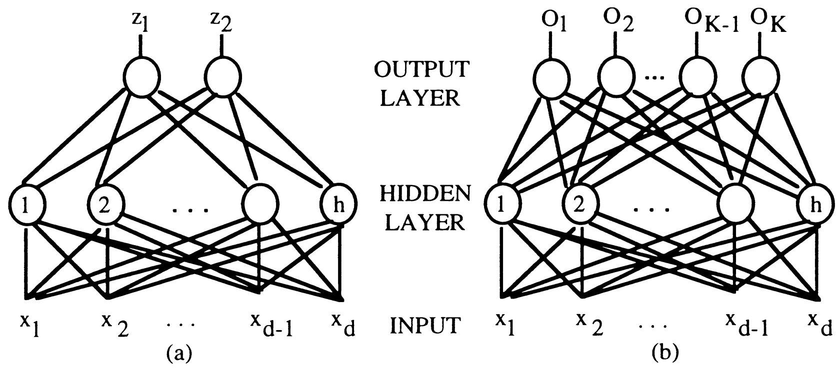

Fig. 2. (a) MLP for data mapping with two output nodes and h hidden layer nodes, (b) MLP for classification with K output nodes each for one class.

	z_1	z_2
for class-1 (two targets)	$T_{11}(1)$	$T_{21}(1)$
	$T_{12}(1)$	$T_{22}(1)$
for class-2 (one target)	$T_{11}(2)$	$T_{21}(2)$
	...	
for class-K (t targets)	$T_{11}(K)$	$T_{21}(K)$
	...	
	$T_{1t}(K)$	$T_{2t}(K)$

Here, the target values are indicated by $T_{a,b}(c)$ where $a = 1, 2$ denotes the two coordinates on the plane, $b = 1, \ldots, t$ indexes the multitude of target values for each class, and finally $c = 1, \ldots, K$, is the class membership. Notice that in the above example, the number of target values in the k classes is taken respectively as $[1, 2, \ldots, t]$ with a total of T target values. In training this mapping algorithm, the target values, $T_{a,b}(c)$, are accepted as the desired cluster centers which are the images of d-dimensional data points on a plane. For example, for a three-class problem with single target assignments per class, the target values could be, respectively, $(T_{11}(1), T_{21}(1)) = (1, 0.5)$, $(T_{11}(2), T_{21}(2)) = (-1, 0.5)$, $(T_{11}(3), T_{21}(3)) = (0, 0.5)$, for a subjectively pleasant plot. The user can greatly influence the mapping procedure and interactively search for the intrinsic structure in data by manipulating the target values. Another convenient way to initialize the target values in the least squares mapping one can use:

$$T_j(c) = [\cos(\alpha_j), \sin(\alpha_j)], \quad \alpha_j = \frac{2\pi(j-1)}{c}, \quad j = 1, \ldots, c. \tag{3}$$

Then the user can interactively reassign objects from one class to another, merge, split or delete classes, or identify new target values by taking into account the neighborhood relationships between the classes.

The execution of the non-linear transformation of the d-dimensional data by the MLP is given as

$$z_{nj} = w_{0j} + \sum_{l=1}^{h}[w_{lj} + f(w_{01} + \sum_{i=1}^{d} x_{ni} w_{li})], \quad j = 1, 2, \quad n = 1, \ldots, N \tag{4}$$

where z_{nj} is one of the two coordinates after mapping of the nth, $n = 1, \ldots, N$, d-dimensional data point, x_{ni}, h is the number of nodes in the hidden layer, N is the total number of data points, "w" denotes weights of the perceptron, and $f(\cdots)$ is a non-linear function. Standard perceptron training algorithms can be used to train this structure. Notice that the MLP will have K output nodes for a K-class recognition problem, while for the mapping problem the number of output nodes is always 2. On the other hand these output values will have as target values a total of T values.

As in the pattern analysis problems, the advantage of the MLP mapping is that non-linear relationships in the data are obtained during the training. On the other hand, it shares the drawbacks presented by other neural networks such as converging to a local minimum, overtraining, etc. The multilayer perceptron mapping method is non-linear, requires prelabeling of the classes, and can operate on the multi classes in the original feature space.

From the medical diagnosis standpoint, the advantages of the proposed method are that the targets can be selected to reflect the doctor's intuition and in cooperation with them. For example targets can be chosen to reflect their notion of distance between pathologies, or multiple targets can be selected to account for alternate symptoms of the same pathology. In fact prototype patients themselves can be selected as targets, whereby the doctor can search for other patients that map close to the reference patient.

4. Experimental results: a medical diagnostic case

The proposed mapping method has been applied to the classification of respiratory sounds into three pathology classes. Respiratory sounds are known to provide useful diagnostic information for various pathologies and anomalies of lungs and airways. Automatic classification of respiratory sounds is significant in that it provides a computer-aided tool to auscultation and increases its potential diagnostic value [1].

The lung sounds are highly non-stationary processes as the sound generating mechanisms change during the course of inspiration and expiration. These sounds show also intrapatient variability due to flow rate and even posture, as well as interpatient variability. It has therefore been found useful to partition the record of respiratory cycles into phases and segments.

Recorded signals in a respiration cycle, typically lasting 2–3 s, were divided into a fixed number of segments as illustrated in Fig. 3. In this figure the flow signal is also superimposed in order to enable the identification of the so-called respiratory phases. The number of phases is six reflecting the early, mid and late stages of the inspiration and expiration half-cycles. Since there are six phases, then each phase contains ten consecutive segments. A separate mapping was applied for each phase feature set which implies that a separate MLP was trained for each phase. The reason why a separate mapper was used in each phase was to combat the non-stationarity of the respiratory sounds. The resulting intermediate maps were then re-projected by a second stage mapper to derive a single image of the patient sound record. Notice that if the data had been stationary, a single phase, hence a single MLP mapper would have sufficed.

The number of segments was chosen to be 60 for a whole respiration cycle corresponding each typically to 30–40 ms intervals, or to 150–200 samples given the 5 kHz sampling rate. The length of these segments is a compromise between being short enough to capture a stationary

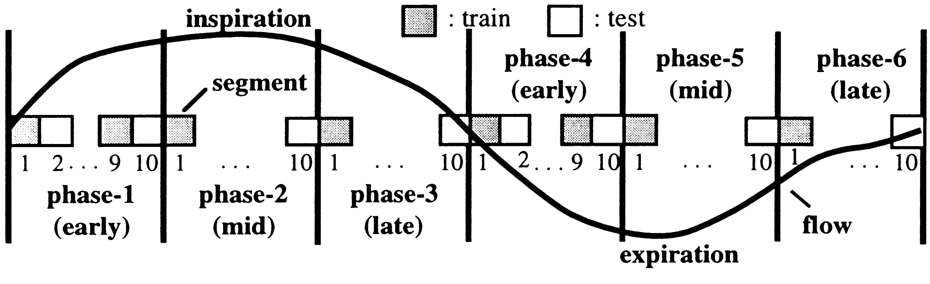

Fig. 3. Division of a respiratory cycle into segments and phases, and the choice of training and test segments from each phase. These operations take place on the respiratory sound signal synchronized with the above flow signal.

interval on one hand, and long enough to yield statistical significance. Each segment was characterized by six cepstral coefficients forming its feature set. These segments in each respiration cycle were further grouped into six phases, namely, early, mid, late inspiration and expiration phases.

Measurement records from 18 chronic obstructive patients (class-1), 19 restrictive lung disease patients (class-2) and 20 healthy subjects (class-3) were used (total of 57 patients) in the experiments.

The block diagram of the mapping that we used for this particular application is shown in Fig. 4. As shown in Fig. 4, at the first stage mapping, each phase MLP receives the cepstral vectors ($d = 6$) of the segments in the corresponding phase portion. The output, i.e. the image, for the nth input sample, x_n^{ϕ}, in phase $f = 1, \ldots, 6$ is z_n^{ϕ}, whose components are given as in Eq. (4). Recall that a separate MLP is trained for each respiratory phase since their dynamic and sound generating characteristics differ significantly. Data is run over all training segments for all the labelled patients, though the patient indices have been omitted in equations for simplicity. The outputs of the phase MLPs are then combined into a new vector with 12 components (recall that each z is 2-dimensional):

$$\zeta_n = [\mathbf{z}_n^1, \mathbf{z}_n^2, \mathbf{z}_n^3, \mathbf{z}_n^4, \mathbf{z}_n^5, \mathbf{z}_n^6].$$

In order to obtain the image, $\boldsymbol{Z}_{kn}$, of a whole respiratory cycle of the kth, $k = 1, \ldots, 57$, patients, a second stage MLP (Fig. 4) is used to map z_n into two dimensions using again an MLP as in Eq. (4) using the same target values as in the first stage. At this stage each patient is represented by five points, each corresponding to one of the five segments in the phases. In the final map these segment images are averaged so that the patient is shown only by one 2D vector:

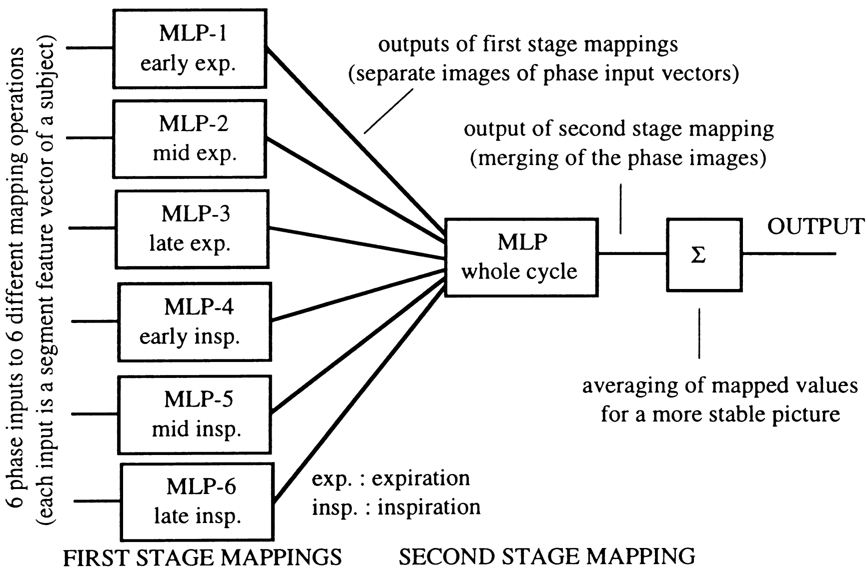

Fig. 4. Block diagram for mapping the respiratory sound data. For the first stage MLP mapping, the data vector size is $d = 6$, and the number of hidden layer nodes are $h = 7$ while for the second stage $d = 12$, $h = 7$.

$$\boldsymbol{\chi}_{\mathbf{k}} = \frac{1}{5}\sum_{n=1}^{5}\mathbf{Z}_{\mathbf{kn}}, \quad k = 1, 2, \ldots, 57 \text{ (patients)}.$$

This two-stage mapping and the final averaging over segments helps to control the scatter due to temporal evolution of the respiratory waveforms. In our experiments with this data, we have defined only one target value, $(T_{11}(c), T_{21}(c))$, $c = 1, 2, 3$, for each class, though as was mentioned in Section 3, more than one target could have been defined for each class. The values of the target pairs were chosen as $(T_{11}(1), T_{21}(1)) = (1, -1)$ *for class-1* (obstructive), $(T_{11}(2), T_{21}(2)) = (-1, -1)$ *for class-2* (restrictive), and$(T_{11}(3), T_{21}(3)) = (0, 0.732)$ for *class-3* (healthy). These target values were kept identical for the 1st and 2nd stage mappings. After several preliminary mapping experiments, $h = 7$ hidden units were chosen for each MLP in both stages (see Fig. 2a).

In order to test the interaction of human observers with the graphical display of the data, we asked four observers to draw piecewise linear decision boundaries on the images obtained from the training data (Fig. 5a). As illustrated in Fig. 3, the training and test data were obtained from alternate segments of the patients' lung sound records. The decision boundaries on the training data images were then simply ported on the test data images as illustrated in Fig. 5b. The decision boundaries specified by Observers 1, 2 and 3 resulted in five, four and four errors in the test image, respectively. Most of the errors were common for all observers. It was

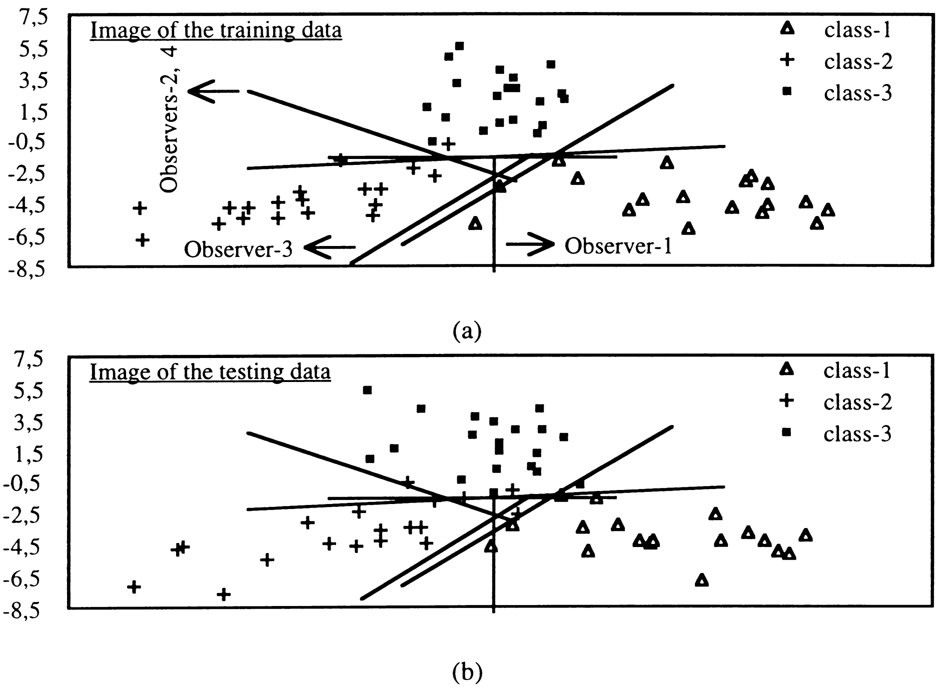

Fig. 5. (a) Mapping images of respiratory sound training data, and decision boundaries specified by three independent observers, (b) mapping images obtained from the test data with the class boundaries obtained in (a).

interesting to note that the decision boundaries drawn by the second and fourth observers were almost the same. The average correct classification performance determined by four observers on the test image is 92.54%. This result is consistent with our previous classification experiments. In those experiments the correct classification performance was 89.47% on the same data set with the cooperation of MLP classifiers, and furthermore the three misclassified subjects was found to be common to both visual and pattern recognition methods.

We also mapped the training data with different techniques, but using the same procedure explained above. In other words the strategy illustrated in Fig. 4 was followed, except that mapping was executed with techniques different than MLP. These techniques were (1) the principal component mapping (PCM), (2) the least squares mapping (LSM), (3) the extended declustering mapping, (EDM) (4) the distance from two means mapping (DF2M). The EDM and DF2M require two classes in the data while the LSM and PCM are not so constrained. Thus, we merged class-2 and class-3 into one class and class-1 remained the same while implementing EDM and DF2M techniques. In EDM, the spread coefficient [4] was chosen as 3, and in DF2M, class-conditional covariance matrices were assumed unequal. Images obtained from these four mappings are given in Fig. 6.

When Figs. 5 and 6 are compared visually, one can notice that (a) none of the four mappings could outperform our method, (b) LSM resulted in a better display of the data compared to PCM, EDM and DF2M mappings. These may be due to the significant overlap between the classes in the original data space. PCM does not utilize class information on the data, and the squared error criterion which will be minimized does not guarantee the preservation of the intrinsic structure of the data. In EDM the separability between the selected pair of classes is tried to be enhanced, while the separability between the classes merged into one group does not contribute to the problem directly. In DF2M, points in the plane are the distances of the original pattern vectors from two class means, and the dimensionality reduction is done as if the data would be classified according to the nearest mean classification rule. These properties of EDM and DF2M require good separation of the classes merged into one group.

We noticed that the EDM and DF2M result in better images if the database really consists of two classes only. On the other hand, our proposed mapping and LSM do not require two classes in the data in contrast to EDM and DF2M. Their common feature is the prespecification of class centers on the plane. Both methods try to minimize the scatter of the mapped versions of data points around the preselected 2D targets. However, the advantage of our method may be the non-linearities used within the nodes of the MLP. The two-layer perceptron is able to create non-linear decision boundaries in the multivariate space. The plane presented by LSM is in fact the subspace of the original space which discriminates the Gaussian clusters best. If the classes are non-Gaussian the LSM may fail. The training procedure in the MLP makes no assumptions on the shape of the class distributions but it concentrates on the errors occurring where distributions overlap. This may be another reason which makes our method more robust compared to the techniques considered in this study and for this particular data.

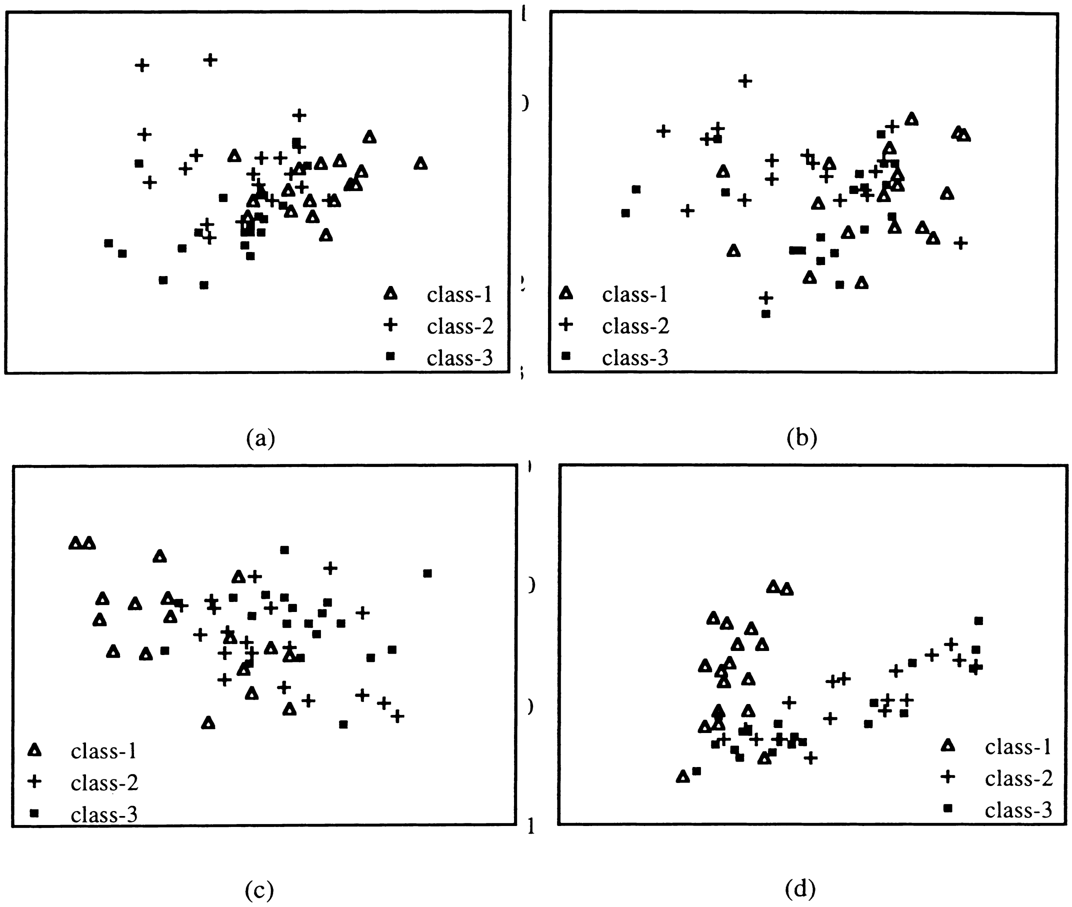

Fig. 6. Images of the respiratory data obtained using (a) the least squares mapping, (b) the principal component mapping, (c) the extended declustering mapping, (d) distance from two means mapping.

5. Conclusions

In this paper, we proposed a new non-linear feature mapping method intended for medical diagnostics. Our method makes use of MLPs with prespecified class target values in two dimensions for the design purpose of interactive classifiers on visual displays. A second innovation in the proposed approach was a temporal segmentation and later mapping fusion scheme. With this non-sequential treatment of the data, that is its partitioning into phases, one could easily deal with evolutionary data as is the case with respiratory cycles.

Preliminary experiments to visually classify the parameter maps of the respiratory sound signals were promising implying that our method can be a useful tool as an assistive device for medical doctors. The interaction of our mapping with human users seemed quite satisfactory in this limited set of experiments. Furthermore, the classification results obtained via the visual plots of the scatter diagrams showed consistency with the previous classification experiments based on statistical pattern recognition methods.

The advantages of the visual classification method as opposed to other formal methods are: (1) the targets can be selected interactively with experts, e.g., physicians, (2) prototype patients can be selected as targets, (3) one can map data with respect to groups of classes, (4) using the mapping display the physician/expert himself determines the decision boundaries for classification purposes.

The proposed method outperformed, at least in subjective classification experiments, other well-established mapping methods, such as the principal component mapping, the least squares mapping, the extended declustering mapping and the distance from two means mapping. The closest competitor to our scheme among the methods in the literature, was the least squares mapping.

References

[1] B. Sankur, Y.P. Kahya, E. Güler, Ç.T. Engin, Comparison of AR based algorithms for respiratory sounds classification, Comput. Biol. Med. 23 (4) (1994) 67–76.
[2] M. Mezzalama, P. Prinetto, B. Morra, Experiments in automatic classification of laryngeal pathology, Med. Biol. Eng. Comput. 21 (1983) 603–611.
[3] D. Cabello, S. Baerro, J.M. Salceda, R. Ruiz, J. Mira, Fuzzy *k-NN* classifiers for ventricular arhythmia detection, Int. J. Biomed Comput. 27 (1991) 77–93.
[4] W. Siedlecki, K. Siedlecka, J. Sklansky, An overview of mapping techniques for exploratory pattern analysis, Pattern Recognition 21 (5) (1988) 411–429.
[5] W. Siedlecki, K. Siedlecka, J. Sklansky, Experiments on mapping techniques for exploratory pattern analysis, Pattern Recognition 21 (5) (1988) 431–438.
[6] J.H. Friedman, J.W. Tukey, A projection pursuit algorithm for exploratory data analysis, IEEE Trans. Comput. 27 (1974) 881–890.

Emin Cagatay Guler received his B.S. degree in Electronics and Communication Engineering from Istanbul Technical University, Istanbul, in 1989 and his M.S. and Ph. D. degrees in Biomedical Engineering from Boğaziçi University, Istanbul in 1992 and in 1998, respectively. His research interests are adaptive signal processing, spectral estimation, pattern recognition, wavelet transform theory and their applications in biomedical engineering.

Bülent Sankur received his Ph.D. degree in Electrical Systems from Rensselaer Polytechnic Institute and has been at Boğaziçi University in the Department of Electric and Electronic Engineering since 1976. He has held visiting positions at the University of Ottawa in Canada, at the Technical University of Delft in Holland and at the Istanbul Technical University. His research interests are in the areas of biomedical signal processing, image and video compression and multimedia systems. He was instrumental in establishing a signal and image processing laboratory at Boğaziçi University.

Yasemin P. Kahya received the B.S. degrees in Electrical Engineering and in Physics from Boğaziçi University in 1980, M.S. degree in Engineering and Applied Sciences from Yale University in 1981 and Ph.D. degree in Biomedical Engineering from Boğaziçi University in 1987. She has been teaching ever since in the Department of Electric and Electronic Engineering at Boğaziçi University. Her current research interests are in the area of designing and building of electronic instrumentation for physiological variables along with signal processing applications in biomedical engineering.

Sarunas Raudys received his Ph.D. degree in Technical Sciences from Institute of Mathematics and Cybernetics of Academy of Sciences of Lithuania in 1979. He has been with the Vilnius Technical University, has been a consultant to industry and has held visiting positions at Michigan State University, Baylor University, Delft University of Technology, Energy Research Center of Netherlands and University of Pierre et Marie Curie. His research interests are in the areas of statistical pattern recognition, artificial neural networks and data analysis.