ECG Analysis Using Nonlinear PCA Neural Networks for Ischemia Detection

Telemachos Stamkopoulos, *Student Member, IEEE*, Konstantinos Diamantaras, *Member, IEEE*, Nicos Maglaveras, *Member, IEEE*, and Michael Strintzis, *Senior Member, IEEE*

***Abstract*— The detection of ischemic cardiac beats from a patient's electrocardiogram (ECG) signal is based on the characteristics of a specific part of the beat called the ST segment. The correct classification of the beats relies heavily on the efficient and accurate extraction of the ST segment features. In the present paper, an algorithm is developed for this feature extraction based on nonlinear principal component analysis (NLPCA). NLPCA is a relatively recently proposed method for nonlinear feature extraction that is usually implemented by a multilayer neural network. It has been observed to have better performance, compared with linear principal component analysis (PCA), in complex problems where the relationships between the variables are not linear. In this paper, the NLPCA techniques are used to classify each segment into one of two classes: normal and abnormal (ST+, ST−, or artifact). During the algorithm training phase, only normal patterns are used, and for classification purposes, we use only two nonlinear features for each ST segment. The distribution of these features is modeled using a radial basis function network (RBFN). Test results using the European ST-T database show that using only two nonlinear components and a training set of 1000 normal samples from each file produce a correct classification rate of approximately 80% for the normal beats and higher than 90% for the ischemic beats.**

***Index Terms*— Biomedical signal processing, ischemia detection, neural networks, principal component analysis, radial basis function.**

I. Introduction

OVER the past decades, a great deal of research has been conducted in the field of biomedical signal processing [1]. In everyday clinical practice, a number of biomedical signals are recorded and used for patient monitoring or diagnostic purposes. The electrocardiogram (ECG) plays a key role in patient monitoring and diagnosis. In the European Union alone, it is estimated that 0.3 ECG's per citizen per year are recorded. The wide usefulness of the ECG, and the ease of recording it in a noninvasive manner, has resulted in concentrating considerable research effort on ECG processing techniques [1], [2]. These techniques deal mainly with ECG pattern recognition [1], [3], [4], parameter extraction, spectro-temporal techniques late potential characterization [5], arrhythmia detection, [6] and noise removal [7].

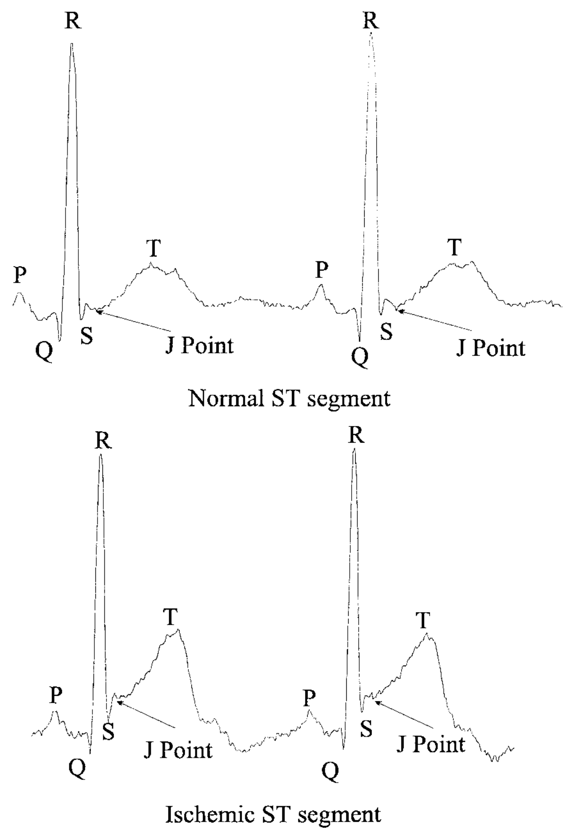

Fig. 1. Normal and ischemic ECG patterns. In the normal case, we observe the constituent ECG waves and the J point. In the ischemic ECG, we observe the ST elevation (this can be depression as well), and we observe in the second beat that the J point is not easily discernible. The data are coming from file e0103 of the European ST-T database, and the lead is V4.

The ECG consists of three basic waves: the P, QRS, and T (Fig. 1). These waves correspond to the far field induced by specific electrical phenomena on the cardiac surface, namely, the atrial depolarization (P wave), the ventricular depolarization (QRS complex), and the ventricular repolarization (T wave). We should note here that the ECG does not look the same in all the leads of the standard 12-lead system used in clinical practice. The ECG polarity and the shape of the ECG constituent waves may change depending on the lead that is used [2]. Numerous techniques have been developed to recognize and analyze these waves, ranging from digital filtering techniques to neural network (NN) and spectro-temporal-based techniques [1], [5], [7]–[17].

Manuscript received July 24, 1997; revised March 13, 1998. This work was supported in part by the Project EPET II—339/IHIS from the Greek General Secretariat of Research and Technology and by the CEC Health Telematics Project I4C-HC1024. The associate editor coordinating the review of this paper and approving it for publication was Prof. Yu-Hen Hu.

T. Stamkopoulos and N. Maglaveras are with The Medical School, Laboratory of Medical Informatics, Aristotelian University, Thessaloniki, Greece (e-mail: nicmag@med.auth.gr; nicmag@vergina.eng.auth.gr).

K. Diamantaras and M. Strintzis are with the Information Processing Chair, Department of Electrical Engineering, Aristotelian University, Thessaloniki, Greece.

Publisher Item Identifier S 1053-587X(98)07795-2.

Ischemic heart disease is one of the most common fatal diseases in the industrialized world [2]. In the United States, it is estimated that 1 million people die due to ischemic/coronary heart disease annually. The key in treating ischemia is its timely detection. Since ECG is the most commonly recorded signal in the patient monitoring and examination process, it becomes important to be able to reliably detect ischemia from ECG analysis. Detection of ischemia can be achieved by analyzing the ST segment of the ECG (Fig. 1). Ischemia is caused by decreased blood flow to parts of the myocardium, due to vessel occlusion or muscle injury [18]. This causes the depolarization of the resting membrane potential of the ischemic region with respect to the resting membrane potential of the normal region. This potential difference causes the flow of an injury current that is manifested in the ECG by an elevated or depressed ST segment (Fig. 1). While in most cases it is easy to discern the ST depression, in other cases the ST depression may not be evident, due, for example, to the relative position of the infarct and the recording point. In addition, ST depression may be influenced by body position, as is often the case with leads III or aVF [19]. Other problems contributing to poor detection and incorrect classification of the ST segment in the ECG include the following: slow baseline drift, noise, sloped ST changes, patient-dependent abnormal ST depression levels, and varying ST-T patterns in the ECG of the same patient. A number of methods have been proposed in the literature for ST detection based on digital filtering, time analysis of the first derivative of the signal and spectrotemporal, wavelet-based and syntactic methods [9], [20]–[23]. These methods tend to measure specific parameters (such as degree of depression, ST-T duration etc.) in ways critically dependent on the correct detection of the J point on the ECG, which is the inflection point following the S wave. In many cases, where the ST segment is sloped or is influenced by noise, it is impossible to reliably identify this point (Fig. 1). In such cases, the above approaches do not produce reliable results.

NN's have been used in the past as pattern and statistical classifiers [24], [25] in many application areas including medicine [26]. For example, NN's were used for QRS/PVC classification [12], [13], arrhythmic events classification, or for detection of atrial fibrillation [14]. NN-based ST segment analysis has been used for

1) automated detection of the J point and the onset of the T wave using adaptive resonance theory [27];
2) ischemia episode detection using adaptive backpropagation NN [28];
3) the classification of ST-T segments.

The latter is achieved in [29] with a classical backpropagation NN using inputs of measured ST-T data such as ST slope, ST-J amplitude, and positive and negative amplitudes of the T wave with emphasis in data coming from myocardial infarction patients. Other possible areas of NN application in ECG analysis and interpretation are pattern recognition and classification following principal component analysis (PCA) techniques [30]–[32] or nonlinear mapping techniques [33].

Biomedical signal processing techniques are usually evaluated using standard annotated databases, which are available worldwide as common references. The main database used for ischemia detection is the European ST-T database [34]. This database includes two channels from Holters corresponding to 90 patients with ischemic heart disease. It includes numerous ischemic episodes of all types, and thus, it is very useful in evaluating ischemia detection algorithms [35].

In this paper, a new method based on NLPCA implemented by NN [36] is employed for ST segment feature extraction, and RBFN is subsequently used for the classification of ischemic ECG's. This method is shown to be quite reliable in the classification of normal and ischemic beats. The structure of the paper is as follows. First, an overview of the NLPCA method is given. Subsequently, the preprocessing of the ECG signals is described. Then, the classification scheme based on RBFN for the ischemic and normal beat detection is discussed, and the results of testing the method on 34 files from the European ST-T database are presented. Final conclusions are drawn in the discussion section.

II. Nonlinear Principal Component Analysis

The purpose of principal component analysis (PCA) is to identify linear correlations between random variables aiming at data dimensionality reduction. The distribution of the variables is "explained" by a few linear features called *principal components or factors*. The mapping from the data space to the feature space is referred to as *coding* and the reverse mapping as *decoding*. In classical PCA, both coding and decoding mappings are assumed linear.

PCA is a purely second-order method, which uses the data covariance matrix $\mathbf{C}_x$ in order to determine the optimal projection subspace. Although PCA has found applications in pattern recognition [37], image processing [38], and various modern approaches in signal modeling, spectral estimation, and array processing [39], there are cases where the second-order statistics used by PCA are not enough to efficiently represent the problem. As an illustrative example [36], consider the problem of two random variables $x_1 = \cos(\phi)$, $x_2 = \sin(\phi)$ produced by a single random angle $\phi \in [0, 2\pi)$. The nonlinear coding function $h(x_1, x_2) = \cos^{-1}(x_1)$ extracts[1] the hidden feature variable ϕ, which is enough to perfectly reconstruct the observations x_1, x_2, using the nonlinear decoding functions $\cos$ and $\sin$. The approximation of x_1, x_2 using a single linear principal component will fail since it will try to approximate a 2-D circle using a straight line.

In such cases, it is more appropriate to assume that the hidden factors are nonlinear functions of the observed variables. Furthermore, the reconstruction of the variables from the factors may also be a nonlinear mapping. In general, we assume that the n-dimensional observation vector $\mathbf{x} = [x_1, \ldots, x_n]^T$ is generated by an underlying feature vector $\boldsymbol{\varphi} = [\phi_1, \ldots, \phi_m]^T$ $(m \leq n)$ via n nonlinear continuous functions from $\mathcal{R}^m$ to $\mathcal{R}^n$, $x_1 = f_1(\boldsymbol{\varphi}), \ldots, x_n = f_n(\boldsymbol{\varphi})$. The coding function $\mathbf{h}$ from $\mathcal{R}^n$ to $\mathcal{R}^m$ and the decoding function $\mathbf{g}$ from $\mathcal{R}^m$ to $\mathcal{R}^n$ are members of some classes $\mathcal{F}_c$

[1]This is not the only coding function that extracts the hidden factor ϕ; for example, consider the functions $h(x_1, x_2) = \sin^{-1}(x_2)$ or $h(x_1, x_2) = \tan^{-1}(x_2/x_1)$, etc.

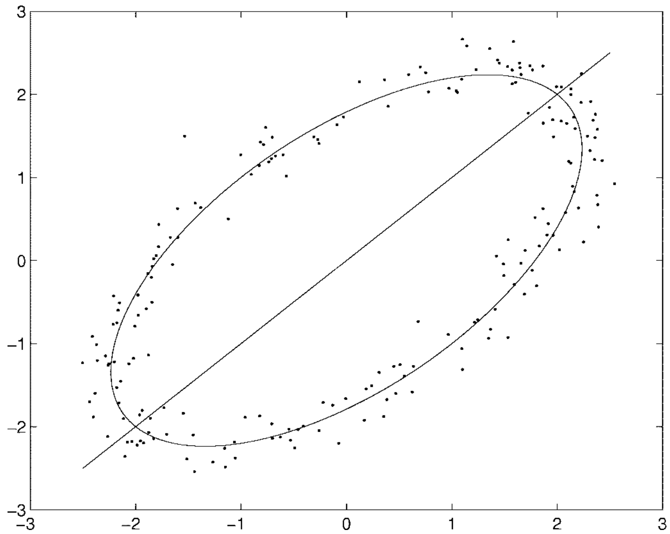

Fig. 2. Set of ellipsoid data demonstrates the difference between nonlinear PCA and standard PCA. The data (dots) are much better represented by the monoparametric nonlinear principal curve (ellipse) than by the linear principal component curve (straight line).

and $\mathcal{F}_d$ of nonlinear functions. The target of the nonlinear PCA (NLPCA) method is the minimization of the nonlinear reconstruction mean squared error (MSE)

$$J = E\|\mathbf{x} - \mathbf{g}(\mathbf{h}(\mathbf{x}))\|^2 \tag{1}$$

by an optimal choice of $\mathbf{g} \in \mathcal{F}_d$ and $\mathbf{h} \in \mathcal{F}_c$. Clearly, the solution to the nonlinear PCA problem depends on both the choice of the sets $\mathcal{F}_c$ and $\mathcal{F}_d$ and the distribution of $\mathbf{x}$. Ordinary (linear) PCA is now a special case for $\mathcal{F}_c$, $\mathcal{F}_d$ being the set of linear mappings.

The unique recovery of the hidden parameters is impossible in general because there are infinite solutions to the NLPCA minimization problem. Indeed, if a pair of functions $\mathbf{h}_1()$, $\mathbf{g}_1()$ achieves the minimum error $J_{\min} = E\|\mathbf{x} - \mathbf{g}_1(\mathbf{h}_1(\mathbf{x}))\|^2$, then so does any pair $\mathbf{h}_1(q^{-1}())$, $\mathbf{g}_1(q())$ for any invertible function $q()$. Nevertheless, the following sets are unique and can be considered as problem inherent [36]:

1) the set $\Im = \{I(\varphi)$, all $\varphi \in R^m\}$ of contours $I(\varphi) = \{\mathbf{x} : \mathbf{h}(\mathbf{x}) = \varphi\}$ for the function $\mathbf{h}$;
2) m-parametric surface $C = \{\mathbf{g}(\varphi)$, all $\varphi \in R^m\}$ generated by $\mathbf{g}$.

C is called the m-parametric *nonlinear principal component surface* of $\mathbf{x}$ (Fig. 2).

The nonlinear PCA has been applied to various complex such as nonlinear dynamical problems appearing in chemical engineering [40], [41] and pattern recognition problems [42], [43].

A. Autoassociative Neural Networks

Consider a two-layer neural network[2] that has a single linear output unit and a nonlinear hidden layer incorporating the sigmoid nonlinear activation function $f(x) = 1/(1 + e^{-x})$.

[2] We follow the convention where the input layer counts as the zeroth layer.

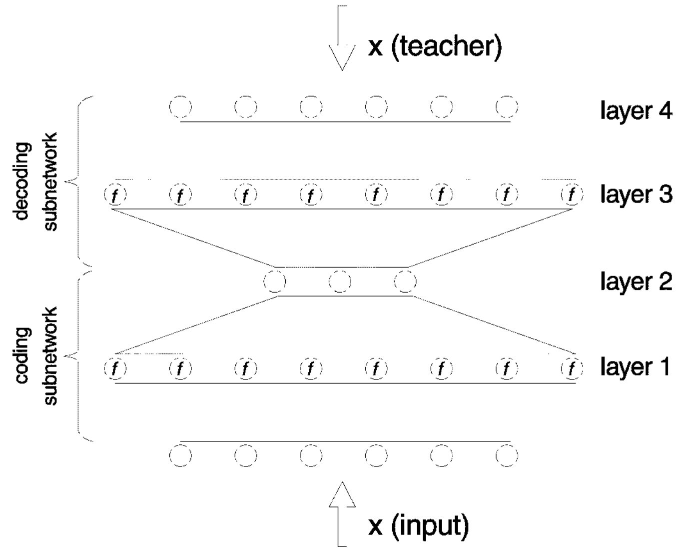

Fig. 3. Auto-associative nonlinear network performing nonlinear PCA. In particular, layers 1 and 3 are the nonlinearlayers, where each node operates using the sigmoidal function f. In our case, we use 80 nodes for these two layers. Layer 2 is the principal component layer, and in our case, we use two principal components. Finally, the input and output layers consist of 20 nodes.

Such a network implements the input–output function

$$\sigma(\mathbf{x}) = \sum_{i=1}^{N} \bar{w}_i f\left(\underline{\mathbf{w}}_i^T \mathbf{x} + \underline{\theta}_i\right) + \bar{\theta} \tag{2}$$

where $\bar{w}_i, \bar{\theta}$ are the upper layer weights and thresholds, and similarly, $\underline{\mathbf{w}}_i, \underline{\theta}_i$ are the lower layer weights and thresholds. It turns out [44]–[46] that the functions of the form (2) can represent any nonlinear continuous bounded function from $\mathcal{R}^n$ to $\mathcal{R}$ with any desired degree of accuracy, provided that the number of hidden units N can be arbitrarily large. Consequently, m units of type (2) can approximate any continuous function from $\mathcal{R}^n$ to $\mathcal{R}^m$ for any dimensions n and m, provided that the hidden layer size can be arbitrarily large. These units form a two-layer feedforward neural network with linear output layer and sigmoid hidden layer.

Suppose now that the classes $\mathcal{F}_c, \mathcal{F}_d$ of the NLPCA coding and decoding functions are the continuous functions from $\mathcal{R}^n$ to $\mathcal{R}^m$ and from $\mathcal{R}^m$ to $\mathcal{R}^n$, respectively. These functions can be implemented by two-layer neural networks, as described above. The total nonlinear PCA network will then be a cascade of two subnetworks, each one consisting of two layers, which correspond to the coding and decoding functions, respectively (see Fig. 3). The input layer of the network has n units equal to the dimensionality of the observation data. The second layer has m linear units, and as it is the output of the coding function, it contains the nonlinear features ϕ. This second layer is also the input to the second subnetwork that computes the decoding function. The fourth layer is the output layer and contains n linear units whose activations form a reconstruction of the input vector. Layers 1 and 3 are nonlinear (sigmoid) and do not have necessarily the same number of units because they are the hidden layers of the coding and decoding subnetworks that implement different functions.

The NLPCA neural model described above was originally proposed by Kramer [47] for the coding and compression of signals appearing in chemical processes. The network learning mode is autoassociative, i.e., the target vector corresponding to the input vector $\mathbf{x}_k$ is $\mathbf{x}_k$ itself.

Let $a_i(l)$ be the activation of unit i in layer $l, w_{ij}(l)$ be the synaptic strength of the connection between unit i in layer l and unit j in layer $l-1$, and call $\theta_i(l)$ the bias for unit i in layer l. With this notation, we can write

$$a_i(l) = \begin{cases} f(u_i(l)), & \text{if } l = 1,3 \\ u_i(l), & \text{if } l = 2,4 \end{cases}$$
$$u_i(l) = \sum_{i=1}^{N_{l-1}} w_{ij}(l)a_j(l-1) + \theta_i(l). \tag{3}$$

The number of units in layer l is denoted by N_l. We also define the activation of the zeroth layer to be the input $a(0) = x$.

Our goal is to minimize the output MSE. Since the target of the network is the same as the input, our MSE over the M training patterns $\mathbf{x}_1, \ldots, \mathbf{x}_M$, becomes

$$J_{\text{NET}} = \frac{1}{2}\sum_{k=1}^{M}\sum_{i=1}^{n}(a_{i,k}(4) - x_{i,k})^2. \tag{4}$$

Typically, the backpropagation algorithm [48] is used to minimize J_{NET}.

III. ECG Signal Preprocessing

In this paper, the main goal of ECG preprocessing is to prepare a description of the ST segment suitable for input to the feature extractor without loss of information. This is accomplished here by computing the differences of ischemic ST segment template from the normal (reference) template. Let $\tilde{y}_{kn} = \{y_{k1}, y_{k2}, \ldots, y_{kg}\}$ be the sequence of samples of the ST segment. The normal template $\tilde{n}_n = \{\tilde{n}_1, \tilde{n}_2, \ldots, \tilde{n}_g\}$ is constructed for each ECG as the average of the ten first normal ST segments $\tilde{y}_{kn}$. If $\{y_i\}$ is the sequence of the ST segment samples, the difference of the two sequences $x_i = y_i - \tilde{n}_i,\ i = 1, 2, \ldots, g$ defines the feature extractor input process.

The ST segment is assumed to begin 60 ms after the R peak in normal sinus rhythm case. The R peak is detected using the Pan and Tompkins algorithm [3]. In the case of tachycardia (RR-interval < 600 ms), the beginning of the ST segment is taken at 40 ms after the R peak. The ST segment for each heartbeat has a predefined length of 160 ms (this means that the end point is 220 ms after R peak in the normal case and 200 ms otherwise). These values are in general agreement with the recommendation of the European ST-T database and with the observations in [28], [30], and [34].

In order to minimize the probability of false detection of ST depression and in order to eliminate low-frequency noise, the isoelectric level must be correctly identified. Our method is based on the assumption that the isoelectric level of the signal lies in the area approximately 80 ms left of the R peak, where the first derivative becomes equal to zero for at least 10 ms or in the flattest 20-ms segment. Let $y_1, y_2, \ldots, y_n$ be the ST segment samples of a beat. More specifically, let $y'_1, y'_2, \ldots, y'_{n-1}$ be their first differences and y_r the sample where the R peak occurs. The isoelectric level samples y_b are then defined if either

$$|y'_{r-j-\text{int}(0.08f)}| = 0, \quad j = 1, 2, \ldots, 0.01f$$
$$\text{or} \tag{5}$$
$$|y'_{r-j-\text{int}(0.08f)}| \le |y'_{r-i-\text{int}(0.08f)}|, \quad i, j = 1, 2, \ldots, 0.02f$$

is satisfied, where f is the sampling frequency, which in our case is 250 Hz [34]. After the isoelectric level is found, it is easy to align the current beat with the previous corrected one $(y^p_{b_p})$ by using the declination of the line connecting the isoelectric levels of the two beats. If $\gamma = \frac{y_b - y^p_{b_p}}{n_b}$, where n_b is the number of samples between the two baseline points, the current beat is corrected with respect to the isoelectric level by multiplying its samples by γ

$$y_i \rightarrow \gamma y_i. \tag{6}$$

After this procedure, the final part of the ST segment consists of $g = 40$ samples for a sampling frequency of 250 Hz. Finally, the g points initially composing each ST segment are reduced in number to $N_i = 20$ by replacing the values of every $g/N_i = 2$ consecutive points by their average.

IV. Classification Procedure

The classification method is based on radial basis functions networks [49]–[51]. This network approximates a data set distribution using a linear combination of Green's functions

$$F(x) = \sum_{i=1}^{N} w_i G(\|\vec{x} - \vec{x}_i\|) \tag{7}$$

where $G(\cdot)$ the Green's function, and $\vec{x}_i$ is its center.

In our case, the Green's functions are defined to be multivariate Gaussian functions characterized by a mean vector $\vec{x}_j$ and common variance σ_j^2. Therefore

$$G(\|\vec{x} - \vec{x}_j\|) = e^{-\frac{\|\vec{x} - \vec{x}_j\|^2}{2\sigma_j^2}} \tag{8}$$

and

$$F(\vec{x}) = \sum_{i=1}^{N} w_i e^{-\frac{\|\vec{x} - \vec{x}_j\|^2}{2\sigma_j^2}} \tag{9}$$

which consists of a linear superposition of multivariate Gaussian basis functions (probability bells) with centers $\vec{x}_j$ and widths σ_j^2.

The learning process is realized by a radial-basis function network (RBFN) [51]. This network responds well to the resulting NLPCA distribution of the ECG transformed beats, which form numerous local clusters on the principal component feature space (plane in our case since we use two principal components). This network consists of three layers. The input layer is made up of source nodes. The inputs here are the coordinates of the states of the NLPCA hidden layer. The number and the activation function of the nodes of the second layer is determined by a self-organizing rule and depends on the number of clusters existing in the principal component feature space. The output layer supplies the response of the network to the activation patterns applied to the input layer.

The activation function of nodes is a Gaussian basis function. If the data set consists of N vectors $\mathbf{D} = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N\}$, then the initial number of nodes is two. The first one has a Gaussian function centered on the center of data set determined by the vector $\vec{x}_{\text{av}} = \frac{\sum_{i=1}^{N} \vec{x}_i}{N}$, and the second is centered on the point determined by the vector $\vec{x}_p$, which maximizes the function

$$P(\vec{x}) = \int_S Q(\vec{x})\, dS \quad \text{with} \quad Q(\vec{x}) = \begin{cases} 1, & \text{if } \vec{x} \in \mathbf{D} \\ 0, & \text{if } \vec{x} \notin \mathbf{D}. \end{cases} \tag{10}$$

The standard deviation of Gaussian functions is assumed to be constant and equal to

$$\sigma = \gamma \frac{\sum_{i=1}^{N} \prod_{j=1}^{m} \left(x_i^{(j)} - \bar{x}^{(j)} \right)}{\prod_{j=1}^{m} \sqrt{\sum_{i=1}^{N} \left(x_i^{(j)} - \bar{x}^{(j)} \right)^2}} \tag{11}$$

where

- m number of dimension (number of nodes in the hidden layer of NLPCA NN);
- $\bar{x}^{(j)}$ mean value for dimension j;
- γ constant depending on the total number of Gaussian functions.

Construction of the second layer of the RBFN utilizes a self-organizing algorithm with the following steps.

1) For each point of the data set $\mathbf{D}$, calculate the output of the Gaussian functions for all hidden nodes

$$y_i = e^{-\frac{\|\vec{x}_j - \vec{c}_i\|}{\sigma^2}} \text{ where } i < n \text{ (number of nodes).}$$

2) Find the winner

$$y_w = \max(y_i), \quad i < n.$$

3) If $y_w > \alpha T$, then change the position of the center of the winner using

$$\vec{c}_w^{(k+1)} = \vec{c}_w^{(k)} + r \cdot \left(\vec{x}_j - \vec{c}_w^{(k)} \right)$$

where α is an arbitrary constant (in our case equal to 2.0), r is the convergence rate of the learning algorithm, and T a threshold whose value, in our case, is $e^{-1.5}$.

4) If $y_w < T/\alpha$, then construct a new hidden node with center at the point $\vec{x}_j$ and standard deviation equal to σ and increase the number of total number n of hidden nodes by 1.

The output layer consists of only one node. A delta rule algorithm is used in order to define the weights of the linear combination of Gaussian functions whose summation produces the output node value. The teacher is the distribution of data set points on the feature space that comprises of two principal components. The delta rule learning rate is constant during this training stage.

After the construction of the RBFN, a threshold is found for the output in order to correctly classify 80% of the normal beats since our training set is composed only of normal beats. This is actually a discrimination border between normal and abnormal patterns. The latter include all patterns with abnormal ST+, ST− and artifacts. This threshold need not be the same for all files. It is chosen so that it can correctly classify 80% of the normal beats of the training set of each specific file.

TABLE I
CLASSIFICATION RESULTS FOR DIFFERENT FILES OF THE EUROPEAN ST-T DATABASE. PERCENTAGES OF CORRECT CLASSIFICATION FOR NORMAL AND ABNORMAL PATTERNS

Correct classification

File	**Normals**	**Abnormals**	*File*	**Normals**	**Abnormals**
e0103	80.00%	87.01%	*e0129*	79.28%	89.75%
e0104	78.52%	84.82%	*e0139*	79.06%	81.44%
e0105	79.35%	62.05%	*e0147*	79.09%	80.14%
e0106	79.35%	71.85%	*e0148*	79.64%	58.70%
e0111	79.17%	60.56%	*e0151*	79.57%	94.42%
e0112	79.39%	77.08%	*e0154*	78.60%	100.00%
e0113	79.58%	99.34%	*e0159*	79.76%	53.21%
e0114	79.68%	99.59%	*e0162*	79.71%	10.92%
e0116	79.20%	94.44%	*e0163*	78.99%	0.00%
e0107	78.70%	90.06%	*e0166*	79.63%	78.11%
e0108	79.69%	91.27%	*e0170*	79.48%	100.00%
e0110	78.92%	100.00%	*e0202*	79.26%	81.17%
e0118	79.96%	70.98%	*e0203*	78.73%	53.26%
e0119	79.96%	61.25%	*e0204*	79.69%	83.44%
e0121	78.76%	95.45%	*e0205*	79.20%	53.46%
e0122	78.76%	100.00%	*e0206*	79.74%	59.57%
e0127	79.39%	93.15%	*e0207*	79.35%	100.00%
Total	**79.32%**	**75.19%**			

V. RESULTS

We ran classification experiments on 34 files of the European ST-T database. Each file consists of more than 4000 patterns that are either normal or abnormal of one kind (ST+ or ST− but not both). For the training set, we used only normal beats (approximately 25% of the normal consecutive beats encountered in each file); this is a novel approach for ischemic beat detection since in all algorithms previously discussed in the literature, the training set was comprised of normal, ischemic beats, and artifacts [28], [31], [32], [33]. In this way, we consider two classes: normals and abnormals, the latter including both ischemic beats and artifacts. Thus, for each file, a binary classification problem is to be solved. The classification approach used relies on definition on the x-y feature plane (where the x and y axis correspond to the two principal components used) of closed regions containing 80% of the normal beats. Every beat represented by a point on the outside of the normal regions is classified as abnormal. This approach is file dependent and is followed for every file tested.

Each classification experiment consists of three steps. In the first, we preprocess the signal as explained in the ECG preprocessing section. In the second, we perform nonlinear feature extraction based on NLPCA, where each ST segment is mapped on the principal component feature space. The third step is final classification using RBFN [51] and the definition of the threshold ensuring correct classification of 80% of normal beats.

Table I shows the classification results for the files of the European ST-T database, where the algorithm was tested, as well as indicators of the overall performance of the algorithm. The table shows the sensitivity indices for normal and abnormal pattern classification. As we can observe, in the normal pattern case, the sensitivity index is approximately 80%. For

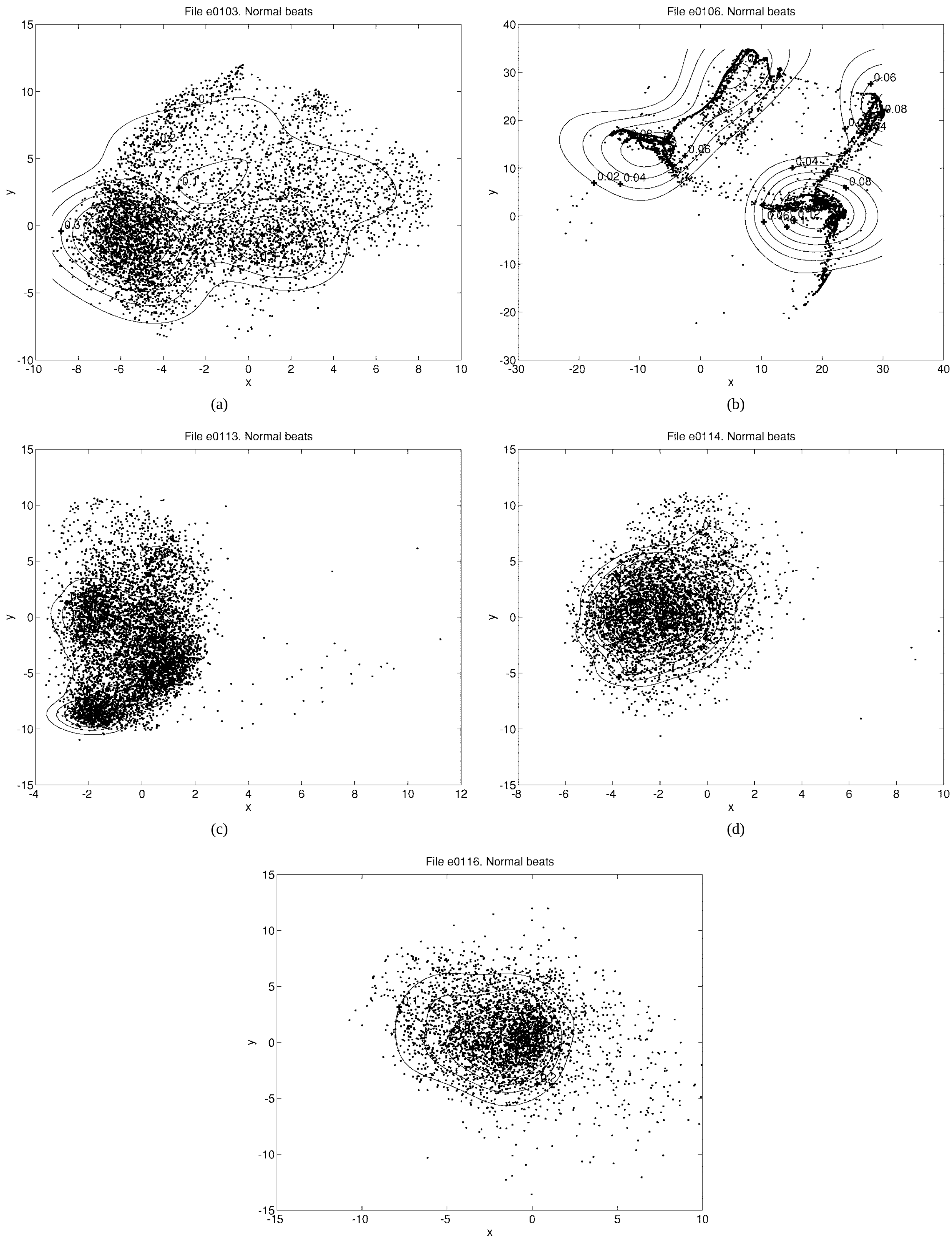

Fig. 4. Five representative cases of normal beat classification. We observe that the mapping and clustering of data varies widely from file to file; however, the classification contours are adjusted to account for the correct classification of at least 80% of the normal beats.

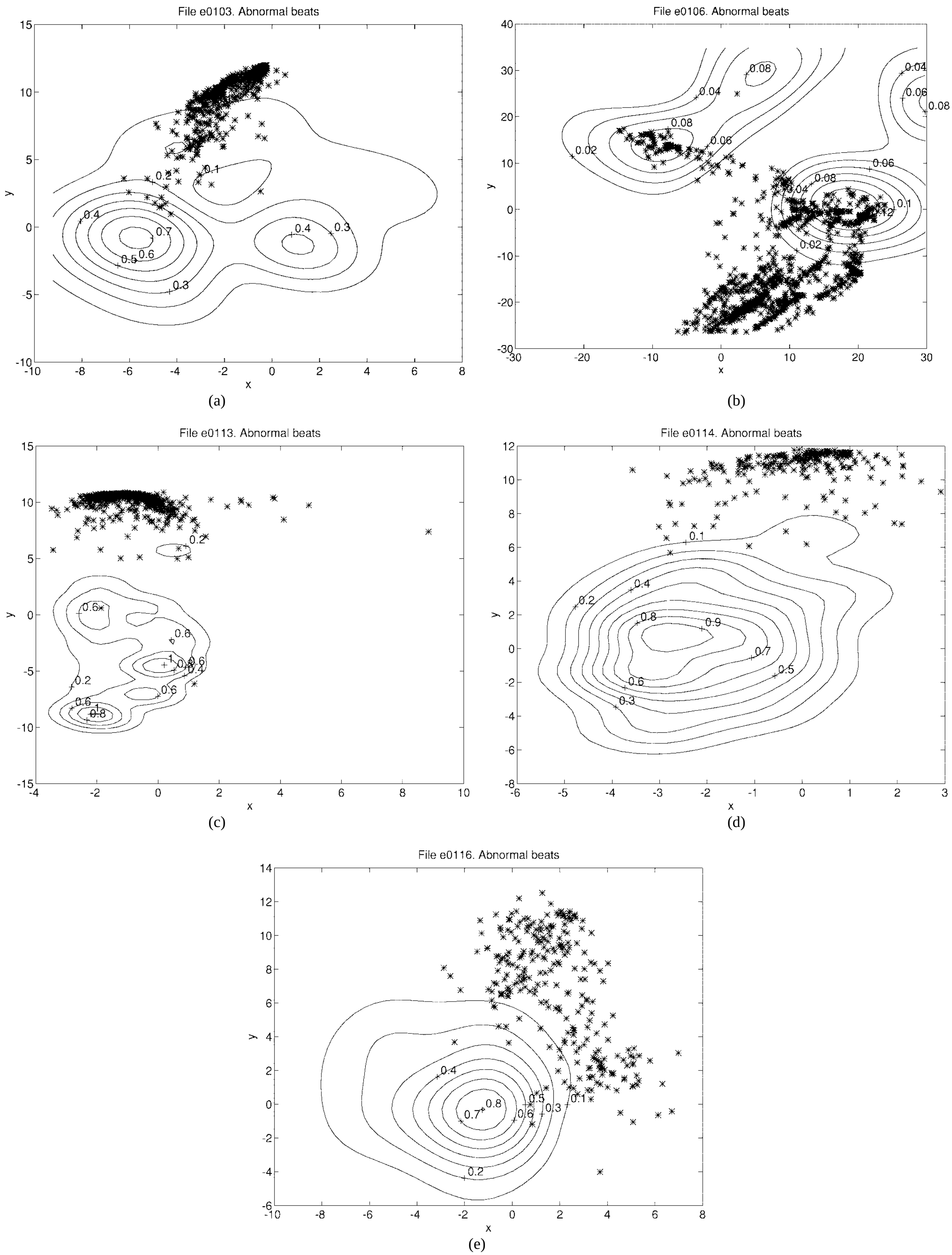

Fig. 5. Same five representative cases of Fig. 4 for abnormal beat classification. We observe that even though the mapping and clustering of data varies widely from file to file, when the classification contours are adjusted to account for the correct classification of the 80% of the abnormal beats, the sensitivity of abnormal beat classification reaches the 99% mark in the cases of the e0113 and e0114 files.

the abnormal pattern recognition, the sensitivity index is in 13 files higher than 90%, in ten files below the 70% mark, and in the remaining 11 files between 70 and 90%. The overall classification index is 79.32% for normal beats and 75.19% for abnormal beats. We observed that the sensitivity index of the algorithm is relatively low only when either more than 80% or fewer than 1% of the beats of the record are abnormal (ischemic). It should be noted that on the average, approximately 5 to 20% of the beats in a Holter are ischemic, as indeed is the case in the rest of the files examined.

These results are obtained using beat-by-beat analysis. By contrast, in ischemia episode detection, the whole sequence of beats is checked [34]. In all algorithms in the literature, the performance of the algorithms detecting ischemia episodes is far better than those used for ischemic beat detection [28], [31], [32], [35].

Fig. 4 shows representative classification spaces for normal beats for five of the files where the algorithm was tested. As we can observe, normal regions containing 80% of the normal beats are usually found in a unique cluster. We note that even in the case of file e0106, where the normal points are broken into two distinct clusters, the RBFN classification method gives good results.

Fig. 5 shows the location of the abnormal points with respect to the classification contours. As we can observe, with the exception of file e0106, we obtain very good results for abnormal beat classification.

Finally, Fig. 6 shows the three-dimensional plot of the probability density function (PDF) of the classification surfaces used in two files with the normal points shown under the PDF surfaces.

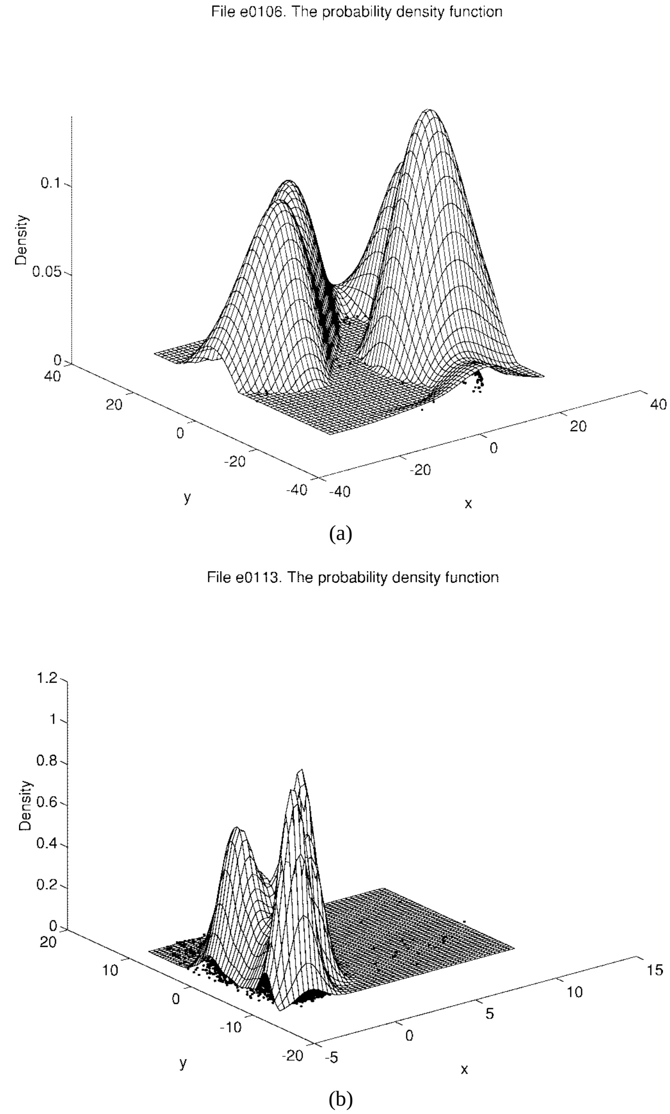

Fig. 6. PDF's used for the classification in files e0106 and e0113. We observe the points lying underneath the PDF's, which are subsequently mapped on the x-y plane, as seen in Figs. 4 and 5.

VI. DISCUSSION

In our work, we studied the representation power of NCPA for the description of cardiac beats. The representation power was manifested by high rates of successful classification in the discrimination problem between ischemic and normal beats. The method is designed to work on a file-to-file basis, meaning that training is done for each patient separately. The fact that we use only two principal components renders the training of the resulting network a fairly easy task. We also tested the algorithm using one, three, and four principal components. When only one component was used, the results were considerably inferior. When three or four components were used, the training time and the computational complexity increased without significant improvement of the results compared with those obtained using two principal components. Following NLPCA-based transformation of the ECG beats and their representation in the principal component feature space, numerous local clusters were found, which were finally classified using RBFN [51].

The training set used in this algorithm consists only of normal beats. This is a novel approach in the sense that all other NN-based ischemia classification algorithms proposed in the literature use patterns from both normal and abnormal beats [12], [28], [29], [30]. In this approach, we defined a normal region contained by a contour in the two-dimensioanl space for the normal beats. Our intention was to find a threshold that would produce equally good classification for normal and abnormal beats. We showed that with this approach, we can keep the sensitivity of the normal beat detection within 80%, whereas the abnormal beat sensitivity can be very high (more than 90% in thirteen files). The ischemic beat sensitivity drops to small values only in the extremely rare case of files where the ischemic beats are either far fewer (less than 1%) or more than 80% than the normal beats (e0162 and e0163).

The performance indices are analogous to the ones reported by other techniques, where information from both channels of the European ST-T database is used [4], [28], [34], [35]. In these cases, linear PCA and NN [31], [32], or adaptive backpropagation NN [28], or techniques based on digital filtering and heuristic processes are used [4], [35] and yield sensitivity indices for ischemic beat detection between 72 and 78%. In fact, PCA combined with NN yield an ischemic episode sensitivity of 73% when no artifacts but only ST+ and ST− patterns are used in the training set [31]. In our case, the training set includes only normal beats that are classified by the proposed algorithm with a classification percentage for normal beats of 80%. Moreover, the total abnormal classification percentage with our method (75.19%) refers to ischemic beat detection rather than ischemia episode detection. This comparison gives a measure of the promising results produced by NLPCA, even though in [4], [30]–[35] the artifacts are included in the normal class rather than the abnormal one.

In conclusion, a new classification and feature extraction algorithm was presented. The algorithm was tested in application to the detection of ischemic beats in ECG Holter recordings, which is one of the most important biomedical signal processing problems. The method apparently exhibits superior performance compared with other methods using PCA/NN for ischemic beat/episode detection [30]–[31].

REFERENCES

[1] A. Cohen, *Biomedical Signal Processing*. Boca Raton, FL: CRC, 1988.
[2] P. Coumel and O. Garfein, *Electrocardiography: Past and Future, Ann. NY Acad. Sci.*, vol. 601, 1990.
[3] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, pp. 230–236, 1985.
[4] A. Taddei, G. Constantino, R. Silipo, M. Emdin, and C. Marchesi, "A system for the detection of ischemic episodes in ambulatory ECG," *Comput. Cardiol.* New York: IEEE Comput. Soc. Press, 1995, pp. 705–708.
[5] O. Meste, H. Rix, P. Caminal, and N. Thakor, "Ventricular late potentials characterization in time-frequency domain by means of a wavelet transform," *IEEE Trans. Biomed. Eng.*, vol. 41, pp. 625–634, 1994.
[6] N. V. Thakor and Z. Yi-Sheng, "Applications of adaptive filtering to ECG analysis: Noise cancellation and arrhythmia detection," *IEEE Trans. Biomed. Eng.*, vol. 38, pp. 785–794, 1991.
[7] W. Philips, "Adaptive noise removal from biomedical signals using warped polynomials," *IEEE Trans. Biomed. Eng.*, vol. 43, pp. 480–482, 1996.
[8] W. Rey, J. D. Laird, and P. G. Hugenholtz, "P-wave detection by digital computer," *Comput. Biomed. Res.*, vol. 4, pp. 509–522, 1971.
[9] G. M. Friesen *et al.*, "A comparison of the noise sensitivity of nine QRS detection algorithms," *IEEE Trans. Biomed. Eng.*, vol. 37, pp. 85–98, 1990.
[10] E. Skordalakis, "Recognition of the shape of the ST segment in ECG waveforms," *IEEE Trans. Biomed. Eng.*, vol. BME-33, pp. 972–974, 1986.
[11] P. Laguna, R. Jane, and P. Caminal, "Automatic detection of wave boundaries in multilead ECG signals: Validation with the CSE database," *Comput. Biomed. Res.*, vol. 27, pp. 45–60, 1994.
[12] M. Strintzis, X. Magnisalis, G. Stalidis, and N. Maglaveras, "Use of neural networks for electrocardiogram (ECG) feature extraction recognition and classification," *Neural Network World J.*, vols. 3–4, pp. 313–327, 1992.
[13] H. Chow, G. B. Moody, and R. G. Mark, "Detection of ventricular ectopic beats using neural networks," *Comput. Cardiol.*, pp. 659–662, 1992.
[14] S. G. Artis, R. G. Mark, and G. B. Moody, "Detection of atrial fibrillation using artificial neural networks," *Comput. Cardiol.*, pp. 173–176, 1991.
[15] R. Silipo, M. Gori, A. Taddei, M. Varanini, and C. Marchesi, "Classification of arrhythmic events in ambulatory electrocardiogram, using artificial neural networks," *Comput. Biomed. Res.*, vol. 28, pp. 305–318, 1995.
[16] P. Kotsas, C. Pappas, M. Strintzis, and N. Maglaveras, "Nonstationary ECG analysis using Wigner-Ville transform and wavelets," *Comput. Cardiol.*, pp. 499–502, 1993.
[17] D. Morlet, J. P. Couderc, P. Touboul, and P. Rubel, "Wavelet analysis of high-resolution ECG's in postinfarction patients: Role of the basic wavelet and of the analyzed beat," *Int. J. Biomed. Comput.*, vol. 39, pp. 311–325, 1995.
[18] L. S. Geddes and W. E. Cascio, "Effects of acute ischemia on cardiac electrophysiology," in *The Heart and Cardiovascular System*, H. A. Fozzard *et al.*, Eds. New York: Raven, 1991, vol. 2, pp. 2021–2054.
[19] N. Goldschlager and M. J. Goldman, *Principles of Clinical Electrocardiography*, East Norwalk, CT: Prentice-Hall, 1989.
[20] A. Gallino *et al.*, "A computer system for analysis of ST segment changes on 24 hour Holter monitor tapes: Comparison with other available systems," *J. Amer. Coll. Cardiol.*, vol. 4, pp. 245–252, 1984.
[21] S. J. Weisner, W. J. Tompkins, and B. M. Tompkins, "A compact, microprocessor-based ECG ST-segment analyzer for the operator room," *IEEE Trans. Biomed. Eng.*, vol. BME-29, pp. 642–649, 1982.
[22] P. Hsia *et al.*, "An automated system for ST segment and arrhythmia analysis in exercise radionuclide ventriculagraphy," *IEEE Trans. Biomed. Eng.*, vol. BME-33, pp. 585–593, 1986.
[23] C. Li, C. Zheng, and C. Tai, "Detection of ECG characteristic points using wavelet transforms," *IEEE Trans. Biomed. Eng.*, vol. 42, pp. 21–28, 1995.
[24] D. R. Hush and B. G. Horne, "Progress in supervised neural networks. What's new since Lippmann?," *IEEE Signal Processing Mag.*, pp. 1–39, 1993.
[25] J. A. Freeman and D. M. Skapura, *Neural Networks: Algorithms, Applications and Programming Techniques*. Reading, MA: Addison-Wesley, 1991.
[26] A. S. Miller, B. H. Blott, and T. K. Hames, "Review of neural network applications in medical imaging and signal processing," *Med. Biol. Eng. Comput.*, vol. 30, pp. 449–464, 1992.
[27] Y. Suzuki and K. Ono, "Personal computer system for ECG ST-segment recognition based on neural networks," *Med. Biol. Eng Comput.*, vol. 30, pp. 2–8, 1992.
[28] N. Maglaveras, T. Stamkopoulos, C. Pappas, and M. Strintzis, "Use of neural networks in detection of ischemic episodes from ECG leads," in *Neural Networks for Signal Processing IV*, J. Vlontzos, J. N. Hwang, and E. Wilson, Eds. New York: IEEE, 1994, pp. 518–524.
[29] L. Ebenbbrandt, B. Devine, and P. W. MacFarlane, "Neural networks for classification of ECG ST-T segments," *J. Electrocardiol.*, vol. 25, pp. 167–173, 1992.
[30] P. Laguna, G. B. Moody, and R. G. Mark, "Analysis of the cardiac repolarization period using the KL transform: Applications on the ST-T database," *Comput. Cardiol.*, pp. 233–236, 1994.
[31] R. Silipo, P. Laguna, C. Marchesi, and R. G. Mark, "ST-T segment change recognition using artificial neural networks and principal component analysis," *Comput. Cardiol.*, pp. 213–216, 1995.
[32] F. Jager, R. G. Mark, G. B. Moody, and S. Divjak, "Analysis of transient ST segment changes during ambulatory monitoring using the Karhunen-Loeve transform," *Comput. Cardiol.*, pp. 691–694, 1992.
[33] N. Maglaveras, T. Stamkopoulos, C. Pappas, and M. Strintzis, "ECG processing techniques based on neural networks and bidirectional associative memories," *J. Med. Eng. Technol.*, vol. 22, no. 3, pp. 106–111, 1998.
[34] A. Taddei *et al.*, "The European ST-T database: Standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography," *Europ. Heart J.*, vol. 13, pp. 1164–1172, 1992.
[35] F. Jager, G. B. Moody, A. Taddei, and R. G. Mark, "Performance measures for algorithms to detect transient ischemic ST segment changes," *Comput. Cardiol.*, pp. 369–372, 1991.
[36] K. I. Diamantaras and S. Y. Kung, *Principal Component Neural Networks: Theory and Applications*. New York: Wiley, 1996.
[37] A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[38] K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[39] W. Therrien, *Discrete Random Signals and Statistical Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1992.
[40] R. Rico-Martinez, K. Krischer, and I. G. Kevrekidis, "Discrete vs. continuous-time nonlinear signal processing of Cu electrodissolution data," *Chem. Eng. Commun.*, vol. 118, pp. 25–48, 1992.
[41] I. G. Kevrekidis, R. Rico-Martinez, R. E. Eske, R. M. Farber, and A. S. Lapedes, "Global bifurcations in Rayleigh-Benard convection. Experiments, empirical maps and numerical bifurcation analysis," *Phys. D*, vol. 71, pp. 342–362, 1994.
[42] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski, "SEXNET: A neural network identifies sex from human faces," in *Advances in Neural Information Processing Systems III*, R. P. Lippmann *et al.*, Eds. San Mateo, CA: Morgan Kaufmann, 1991, pp. 572–577.
[43] G. W. Cottrell and J. Metcalfe, "EMPATH: Face, emotion, and gender recognition using holons," in *Advances in Neural Information Processing Systems III*, in R. P. Lippmann *et al.*, Eds. San Mateo, CA: Morgan Kaufmann, 1991, pp. 564–571.
[44] A. R. Barron, "Approximation and estimation bounds for artificial neural networks," Tech. Rep. 59, Dept. Stat., Univ. Illinois at Urbana-Champaign, Feb. 1991.

[45] A. Cybenko, "Approximation by superpositions of a sigmoidal function," *Math. Contr., Signals Syst.*, vol. 2, pp. 303–314, 1989.
[46] A. Hornik, M. Strinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359–366, 1989.
[47] A. Kramer, "Nonlinear principal component analysis using autoassociative neural networks," *J. Amer. Inst. Chem. Eng. (AIChE)*, vol. 37, pp. 233–243, Feb. 1991.
[48] E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing (PDP): Exploration in the Microstructure of Cognition*, D. E. Rumelhart, J. L. McClelland, and the PDP Research Group, Eds. Cambridge, MA; MIT Press, 1986, pp. 318–362, ch. 8, vol. 1.
[49] T. Poggio and F. Girosi, "Networks for approximation and learning," *Proc. IEEE*, vol. 78, pp. 1481–1497, 1990.
[50] J. Moody and C. Darken, "Fast learning in networks of locally-tuned processing units," *Neural Comput.*, vol. 1, pp. 281–294, 1989.
[51] S. Haykin, *Neural Networks: A Comprehensive Foundation*. New York: Macmillan, 1994.

Telemachos Stamkopoulos (S'97) was born in 1968 in Grevena, Greece. He received the B.Sc. degree in physics in 1991 and the M.Sc. degree in radioelectronics in 1996, both from the Aristotelian University of Thessaloniki, Thessaloniki, Greece. He is currently a Doctorate candidate in medical informatics at the Aristotelian University of Thessaloniki.

Currently, he is an Orthodox Priest of the Holy Archdiocese of Thessaloniki, serving with the churches of Saint Nikolaos and Dimitrios. He is also a Guest Researcher with the Laboratory of Medical Informatics, Aristotelian University of Thessaloniki. His current research interests include neural networks, nonlinear systems, fuzzy control, and biomedical signal processing.

Konstantinos Diamantaras (M'93) was born in Athens, Greece, in 1965. He received his Diploma in electrical engineering from the National Technical University of Athens, Greece, in 1987, and his Ph.D. degree also in electrical engineering from Princeton University, Princeton, NJ, in 1992.

In 1992 he joined Siemens Corp. Research, Princeton, as a post-doctoral researcher. Since 1995 he is with the Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Greece, working as a research scientist involved in various national and international projects. His current research interests include neural networks, pattern recognition, image and signal processing, computer vision and computer architecture. He is the author of the book "Principal Component Neural Networks: Theory and applications," coauthored with S. Y. Kung, and published by Wiley Interscience.

Dr. Diamantaras is a member of INNS, NYAS, and the Technical Chamber of Greece.

Nicos Maglaveras (S'80–M'87) received the Bachelor degree in electrical engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece in 1982 and the M.Sc. and Ph.D. degrees from Northwestern University, Evanston, IL, in 1985, and 1988, respectively, in electrical engineering with emphasis in biomedical engineering.

He is currently an Assistant Professor with the Laboratory of Medical Informatics, Aristotelian University of Thessaloniki. His current research interests are in nonlinear biological systems simulation, cardiac electrophysiology, medical expert systems, ECG analysis, physiological mapping techniques, parallel processing, medical imaging, medical informatics, telematics, and neural networks. He has also developed graduate and undergraduate courses in the areas of medical informatics, computer architecture and programming, biomedical signal processing, and biological systems simulation. He has over 70 publications in refereed international journals and conference proceedings in the above areas. He has served as a reviewer in the CEC AIM technical reviews and in a number of international journals. He has participated in national research projects such as the Health Telematics, Leonardo, TMR, and ESPRIT programs of the CEC.

Dr. Maglaveras is a member of the Greek Technical Chamber, the New York Academy of Sciences, the CEN/TC251-WG5, and Eta Kappa Nu.

Michael Strintzis (S'68–M'70–SM'80) received the Diploma in electrical engineering from the National Technical University of Athens, Athens, Greece in 1967 and the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, in 1969 and 1970, respectively.

He then joined the Electrical Engineering Department, the University of Pittsburgh, Pittsburgh, PA, where he served as Assistant Professor from 1970 to 1976 and Associate Professor from 1976 to 1980. Since 1980, he has been Professor of Electrical and Computer Engineering, University of Thessaloniki, Thessaloniki, Greece. His current research interests include two- and three-dimensional image coding, image processing and biomedical signal and image processing.

Dr. Strintzis was awarded one of the Centennial Medals of the IEEE in 1984.