Chapter 2
Wavelet-based Machine Learning Techniques for ECG Signal Analysis

Roshan Joy Martis, Chandan Chakraborty and Ajoy Kumar Ray

Abstract Machine learning of ECG is a core component in any of the ECG-based healthcare informatics system. Since the ECG is a nonlinear signal, the subtle changes in its amplitude and duration are not well manifested in time and frequency domains. Therefore, in this chapter, we introduce a machine-learning approach to screen arrhythmia from normal sinus rhythm from the ECG. The methodology consists of R-point detection using the Pan-Tompkins algorithm, discrete wavelet transform (DWT) decomposition, sub-band principal component analysis (PCA), statistical validation of features, and subsequent pattern classification. The *k*-fold cross validation is used in order to reduce the bias in choosing training and testing sets for classification. The average accuracy of classification is used as a benchmark for comparison. Different classifiers used are Gaussian mixture model (GMM), error back propagation neural network (EBPNN), and support vector machine (SVM). The DWT basis functions used are Daubechies-4, Daubechies-6, Daubechies-8, Symlet-2, Symlet-4, Symlet-6, Symlet-8, Coiflet-2, and Coiflet-5. An attempt is made to exploit the energy compaction in the wavelet sub-bands to yield higher classification accuracy. Results indicate that the Symlet-2 wavelet basis function provides the highest accuracy in classification. Among the classifiers, SVM yields the highest classification accuracy, whereas EBPNN yields a higher accuracy than GMM. The use of other time frequency representations using different time frequency kernels as a future direction is also observed. The developed machine-learning approach can be used in a web-based telemedicine system, which can be used in remote monitoring of patients in many healthcare informatics systems.

R. J. Martis (✉) · C. Chakraborty
School of Medical Science and Technology, IIT, Kharagpur, India
e-mail: roshaniitsmst@gmail.com

A. K. Ray
Department of Electronics and Electrical Communication Engineering,
IIT, Kharagpur, India

S. Dua et al. (eds.), *Machine Learning in Healthcare Informatics*,
Intelligent Systems Reference Library 56, DOI: 10.1007/978-3-642-40017-9_2,

Keywords Arrhythmia · Normal sinus rhythm · DWT · SVM · Neural network · GMM

2.1 Introduction

In the modern world cardiovascular disease (CVD) is one of the most common causes of death, and is responsible for approximately 30 % of deaths worldwide, and nearly 40 % of deaths in high-income, developed countries [1, 2]. Even though the CVD rates are declining in high-income countries, the rates are increasing in every other part of the world [1].

Generally, the sino-atrial (SA) node acts as the pacemaker of the heart, and the primary source of electrical impulse. Cardiac arrhythmia (also known as dysrhythmia) represents a heterogeneous group of conditions in which there is abnormal electrical activity in the heart. During arrhythmia, other impulse sources may dominate the sinus node and act as independent sources of impulses. Arrhythmia is one kind of CVD, which if left untreated may lead to life-threatening medical emergencies that can result in cardiac arrest, hemodynamic collapse, and sudden death. Abnormalities of both impulse formation and impulse conduction can result in cardiac arrhythmias [3]. The heartbeat interval may be regular or irregular, and may be too fast or too slow. Early intervention with appropriate therapy is recommended in many arrhythmias; if left untreated, such arrhythmias may lead more serious complications. Arrhythmias like ventricular fibrillations and ventricular flutter are imminently life-threatening.

Increasing incidence of cardiovascular disease and death has drawn attention worldwide to the research and development of methods for mass screening to provide prognostic healthcare. One of the greatest challenges for both developed and under-developed countries is the delivery of high-quality cardiac care to the entire population. The lack of sufficiently qualified cardiac experts may, however, limit individual attention for patients and force healthcare professionals to cater to critical conditions and patients requiring immediate attention. The development of automated tools to detect cardiac arrhythmias with considerable accuracy is challenging. Widespread applications of such tools by qualified nurses or paramedics trained to handle the equipment can greatly strengthen the screening programs and aid in providing mass cardiac care with scarce resources.

Electrocardiography (ECG) is a noninvasive test for recording the electric activity of the heart over time and can be captured by surface electrodes. ECG is the simplest and most specific diagnostic test for many heart abnormalities, including arrhythmia, and is especially essential in screening for heart problems. The ECG pattern obtained from a normal subject is known as a normal sinus rhythm. The assessment of alternations in the heart rhythm using an ECG is commonly used to diagnose and assess the risk of any given arrhythmia. Different computational tools and algorithms are being developed for the analysis of the

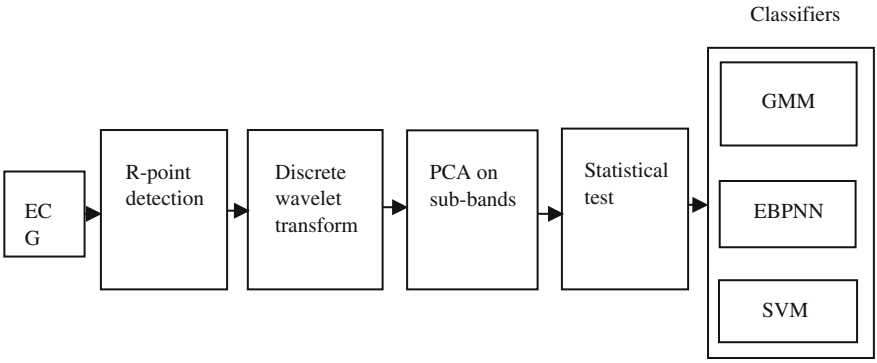

Fig. 2.1 Machine-learning approach of ECG classification into normal sinus rhythm and arrhythmia

ECG signal, and its automated diagnosis. In this chapter, the authors have made an attempt to use machine-based classification of ECG signals to sort normal sinus rhythm and arrhythmia signals into their respective classes.

Many methods for the detection of QRS complex (or the R-point) in the ECG have been proposed [4–6]. The Pan-Tompkins algorithm is commonly used because of its computational simplicity. The wavelet-based method proposed by [5], later extended by [6], can also be used for R-point detection in the ECG. The Pan-Tompkins algorithm has been used in the analysis in this chapter because of its simplicity and higher detection rate.

Few approaches for the classification of arrhythmia beats have been described in the literature [7, 8]. Most of these approaches use principal component analysis (PCA) in the time domain signal [9]. Recently, [10] gave an account of the use of PCA in DWT sub bands. Here, DWT sub-band features are compressed using PCA. Since DWT provides compact supported basis space for the signal, the PCA should provide higher compression than time domain counterparts.

2.2 Materials

In the proposed work, the open source data available at www.physionet.org from MIT BIH arrhythmia and the MIT BIH normal sinus rhythm database is used. The database is explained as follows.

2.2.1 MIT- BIH Normal Sinus Rhythm Database

The MIT-BIH normal sinus rhythm database consists of 18 long term ECG recordings of subjects referred to the Arrhythmia Monitoring Laboratory at Boston's Beth Israel Deaconess Medical Center. Subjects included in this database were found to have had no significant arrhythmias; they included five men, aged 26−45 and thirteen women, aged 20−50. The ECG data was digitized at 128 Hz.

2.2.2 MIT BIH Arrhythmia Database

The MIT BIH arrhythmia database consists of 48 half-hour excerpts of two channel ambulatory ECG data obtained from 47 subjects studied by the BIH arrhythmia laboratory between 1975 and 1979. Twenty-three recordings were randomly taken from a set of 4,000 24 h ambulatory ECG data collected from a mixed population including both inpatients (approximately 60 %) and outpatients (approximately 40 %) at the medical center. The remaining 25 recordings were selected from the same set to include less common but clinically significant arrhythmias. The ECG recordings were sampled at 360 Hz per channel with an 11-bit resolution over the 10 mV range.

2.3 Methodology

Figure 2.1 depicts the machine learning approach of the proposed ECG classification system. The proposed methodology consists of an automated detection of the R-point using the Pan-Tompkins algorithm, wavelet sub-band decomposition using multiple DWT basis functions, principal component analysis (PCA) on DWT sub-bands, statistical significance tests using independent sample *t*-tests, and automated classification using three classifiers, Gaussian mixture model (GMM), error back propagation neural network (EBPNN), and support vector machine (SVM) classifiers.

Prior to R-point detection, some pre-processing is necessary to remove noise and artifacts that the signal may contain. Also, the two classes of signals (arrhythmia and normal sinus rhythm) are sampled at different rates. Therefore, re-sampling is also required.

2.3.1 Preprocessing

Since the signals considered for analysis are sampled at different rates, it is necessary to choose a common sampling rate and re-sample the signals. We have chosen 250 Hz as the common sampling rate, and both signals are re-sampled using standard re-sampling techniques [11]. Also, the signals chosen are from an open source database, and might contain noise, artifacts, and power line interference. It is, therefore, necessary to preprocess the signal. Some basic filters [12] have been used here for noise and artifact filtering.

2.3.2 R-point Detection

The R-wave in the QRS complex of ECG has a high amplitude and an easily detectable peak. The R-point is, therefore, chosen as a characteristic point for registration. A number of algorithms are being reported in the literature for the detection of R-point. The Pan-Tompkins algorithm (1985) is a popular approach for QRS detection, which is computationally simple and, hence, takes less time to run on a computer. In addition to this method, there is a method using the quadratic spline-based discrete wavelet transform [6] that detects the beats accurately, but this method is computationally exhaustive. We have chosen the Pan-Tompkins method due to its computational simplicity and ease in implementation. An extended version of the Pan-Tompkins algorithm consists of the following steps.

1. Compute the first derivative of ECG, and find its absolute value.
2. Smooth this signal by passing through a moving average filter as follows.

$$y(n) = \frac{1}{4}\{x(n) + 2x(n-1) + x(n-2)\}, \tag{2.1}$$

where$x(n)$and $y(n)$ represent the input and output of the smoothing filter.
3. Compute the derivative of the smoothened signal and its absolute value.
4. Smooth the signal obtained from step 3 using the filter in Eq. (2.1).
5. Sum the signal obtained from steps 2 and 4.
6. Threshold the signal obtained from step 5, and obtain square pulses.
7. Compensate for the group delay due to the involved filters by advancing in time.

The derivative gives the slope information, whereas smoothing removes high-frequency noise. The above operations are multistage filtering methods with a non-linear operation in between, which yields the R-point.

2.3.3 DWT Computation

Though Fourier analysis [12] is a traditional tool for the analysis of global frequencies present in the signal, it lacks in temporal resolution due to the increased frequency resolution. Some frequency resolution can be exchanged to get better time resolution. This exchange is performed by defining short duration waves called mother wavelet functions so that the given signal for analysis is projected on this basis function. In traditional Fourier transform, the data is projected on sinusoidal basis functions which extend the span of time domain, i.e., $-\infty$ to $+\infty$. The wavelet basis function [13] is parameterized by the translation 'b' and dilation 'a,' such basis function is given by,

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right). \tag{2.2}$$

Equation (2.2) provides a basis for wavelet transformation. The ECG signals are decomposed for translation and dilation in order to get a multi-resolution representation. This is the case of continuous wavelet transform. This transform is made discrete using a dyadic grid scale in order to get a discrete wavelet transform (DWT) [14]. Such DWT at scale 2^{-m} and time location n is given by

$$\psi_{m,n}(t) = 2^{\frac{m}{2}} \cdot \psi(2^{-\frac{m}{2}} \cdot t - n). \tag{2.3}$$

The dyadic grid sampled DWT are generally orthonormal. Using the basis function in Eq. (2.3), the DWT can be expressed as the inner product between the ECG signal $x(t)$and the basis function as

$$T_{m,n} = \int_{-\infty}^{\infty} x(t)\psi_{m,n}(t)dt \tag{2.4}$$

$T_{m,n}$ is the wavelet coefficient at scale (or dilation) m and location (or translation) n, and it provides the detail (fine information) present in the signal.

The dyadic grid sampled orthonormal discrete wavelets are associated with scaling functions and their dilation equations. The scaling function is associated with signal smoothing and has the same form as the wavelet. It is given by,

$$\emptyset_{m,n}(t) = 2^{-m/2}\emptyset(2^{-m} \cdot t - n), \tag{2.5}$$

where $\emptyset_{m,n}(t)$ has the property $\int_{-\infty}^{\infty} \emptyset_{0,0}(t)dt = 1$.

Often $\emptyset_{0,0}(t)$ is referred to as the father scaling function or father wavelet. The scaling function is orthogonal to the translations of itself, but not to dilations of itself. The smoothing of the signal (or the coarse details or the envelope of the signal) is obtained by convolving the scaling function with the signal, and the obtained samples are called approximation coefficients and are defined as

$$S_{m,n} = \int_{-\infty}^{\infty} x(t)\emptyset_{m,n}(t)dt. \tag{2.6}$$

A continuous approximation of the signal can be obtained at scale m using following equation,

$$x_m(t) = \sum_{n=-\infty}^{\infty} S_{m,n}\emptyset_{m,n}(t), \tag{2.7}$$

where $x_m(t)$ is a smooth, scaling function-dependent version of the signal at scale m. Using both approximation and wavelet (detail) coefficients, the signal can be expressed as follows

$$x(t) = \sum_{n=-\infty}^{\infty} S_{m_0,n}\emptyset_{m_0,n}(t) + \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} T_{m,n}\psi_{m,n}(t). \tag{2.8}$$

From Eq. (2.8), we can see that the original continuous signal is expressed as a combination of an approximation of itself at arbitrary index, m_0 added to a succession of signal details from scales m_0 to negative infinity. The signal detail at scale m is given by,

$$d_m(t) = \sum_{n=-\infty}^{\infty} T_{m,n}\psi_{m,n}(t). \tag{2.9}$$

From Eqs. (2.7) and (2.9), we can write

$$x(t) = x_{m_0}(t) + \sum_{m=-\infty}^{\infty} d_m(t). \tag{2.10}$$

From Eq. (2.10), it easily follows that

$$x_{m-1}(t) = x_m(t) + d_m(t). \tag{2.11}$$

From Eq. (2.11), we can see that if we add the signal detail at an arbitrary scale to the signal approximation at the same scale, we get the signal approximation at an increased resolution. Hence, wavelet transformation provides multi-resolution analysis (MRA) capability.

In this work, different basis functions are used. They are Daubechies-4, Daubechies-6, Daubechies-8, Symlet-2, Symlet-4, Symlet-6, Symlet-8, Coiflet-2 and Coiflet-5. All the considered wavelet families are orthogonal.

The frequency components in each of the sub-bands are shown in Fig. 2.2. Since the sampling frequency of the signal under study is 250 Hz, the maximum frequency contained by the signal will be 125 Hz. Therefore, in the first level, approximation will consist of 0–62.5 Hz frequencies, whereas first level detail consists of 62.5–125 Hz frequencies.

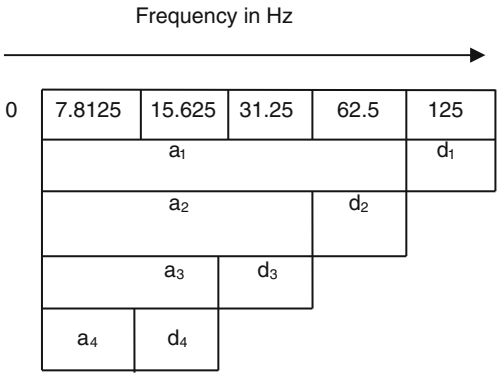

Fig. 2.2 Wavelet decomposition: Distribution of frequencies in various sub-bands

2.3.4 Sub-band Principal Component Analysis

There will be a large number of DWT coefficients in every sub-band of the ECG. If all these coefficients are considered, they will create a large computational burden on the classifier. Therefore, it is wise to represent these coefficients by fewer components. In this study, we have used PCA [15] to reduce the number of features in each of the sub-bands of interest. We identified four sub-bands based on the frequency present in the signal. The four sub-bands are 2nd-level detail, 3rd-level detail, 4th-level detail, and 4th-level approximation. Each of these sub-band wavelet coefficients is subjected to PCA, and the components are chosen such that they will contain 98 % or more of the total energy present in that sub-band.

Mathematically, PCA projects the data from the original coordinate system to a new coordinate system in which the first coordinate corresponds to the direction of maximum variance, and successive coordinates correspond to the directions in decreasing order of variance. Some directions contribute less variability, and those directions need not be preserved in our representation. In the new coordinate system, the axes are called principal components (PCs). A bound of 98 % containment of total variability of segmented ECG is used as a threshold on the total variance in all the considered PCs. PCA consists of following steps.

Compute data covariance matrix as

$$C = \sum_{i=1}^{N} (x_i - \overline{x})(x_i - \overline{x})^T, \tag{2.12}$$

where x_i represents the ith pattern $\overline{x}$ represents the pattern mean vector, and N is the number of patterns.

Compute the matrix V of Eigen vectors and diagonal matrix of Eigen values D as

$$V^{-1}CV = D. \tag{2.13}$$

The Eigen vectors in V are sorted in descending order of Eigen values in D, and the data is projected on these Eigen vector directions by taking the inner product between the data matrix and the sorted Eigen vector matrix.

2.3.5 Statistical Test

The DWT features in compact supported basis space provide sparser representation for ECG in sub-bands. When PCA is applied on sub-bands, it should provide higher compression, and the method is more meaningful. Therefore, it is expected for the principal components of DWT features to provide better statistical significance than time domain principal components. Both time domain features and DWT features are compared against the two classes of signals for equality of class group means using independent sample t test [16].

2.3.6 Classification

The significant DWT features obtained from statistical tests are used for subsequent pattern classification. We have used three classifiers, Gaussian mixture model (GMM), error back propagation neural network (EBPNN), and support vector machine (SVM).

2.3.6.1 Gaussian Mixture Model

We have a two-class pattern classification of ECG into normal sinus rhythm and arrhythmia classes. The GMM assumes that the features are normally distributed, and each class is characterized by its mean (μ_k) and covariance matrix (Σ_k). Since we have applied an orthogonal transformation in compact supported basis space, the features are likely to be uncorrelated. The off-diagonal elements in the covariance matrix are approximately zero. The probability density function of GMM for every sample belonging to a given class k, is given by

$$P(x_n|\omega_k) = \frac{1}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} \exp\left\{-\frac{1}{2}(x_n - \bar{x}_k)^T |\Sigma_k|^{-1}(x_n - \bar{x}_k)\right\}, \tag{2.14}$$

where

$$\bar{x}_k = \frac{1}{|X_k|} \sum_{x_n \in \omega_k} x_n \tag{2.15}$$

and

$$\Sigma_k = \frac{1}{|X_k|}\sum (x_n - \bar{x}_k)(x_n - \bar{x}_k)^T = diag\left(\sigma_{ii}^2\right), 1 \le i \le d \tag{2.16}$$

The corresponding posterior probabilities are given by Bayes's rule as

$$P(\omega_k|x_i) = \frac{P(x_i|\omega_k)}{\sum_{k=1}^{2} P(\omega_k)P(x_i|\omega_k)} \tag{2.17}$$

Initially, the mean and covariance matrices are assigned with some random values. The values are updated using an expectation maximization (EM) algorithm and a maximum likelihood estimation method. The re-estimation formulae are as follows.

$$\widehat{\mu}_j = \frac{\sum_{i=1}^{N} x_i \cdot P\left(\omega_j|x_i\right)}{\sum_{i=1}^{N} P\left(\omega_j|x_i\right)} \tag{2.18}$$

$$\widehat{\sigma}_j^2 = \frac{\sum_{i=1}^{N} (x_i - \widehat{\mu}_j)^2 P\left(\omega_j|x_i\right)}{\sum_{i=1}^{N} P\left(\omega_j|x_i\right)} \tag{2.19}$$

$$p\left(\widehat{\omega}_j\right) = \frac{1}{N}\sum_{i=1}^{N} P\left(\omega_j|x_i\right) \tag{2.20}$$

An initial model having parameters (μ_k, Σ_k) is assumed from the data. The EM algorithm has two steps: an E step and an M step. During the E step, the class conditional density is computed according to Eq. 2.14 and the posterior density is also computed according to Eq. 2.17. During the M step, the model parameters are re-estimated according to Eqs. 2.18−2.20. The process is continued until the new model remains almost identical to the previous model. At this point, the algorithm is said to be converged. The GMM optimizes the following objective function,

$$J = \prod_{n}\sum_{k} p(\omega_k)p(x_n|\omega_k). \tag{2.21}$$

GMM minimizes the product over all the patterns, the total class conditional density weighted with the respective prior probability.

2.3.6.2 Error Back Propagation Neural Network

An error back propagation neural network [17] is used in our study. The neural network is trained on the training set of the data such that the weights get updated recursively with respect to the patterns. This is also an optimization problem where following objective function is minimized.

$$J(\omega) = \frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{c}\left\{y_k(x_n,\omega) - t_k^n\right\}^2, \tag{2.22}$$

where $y_k(x_n, \omega)$ is the network response for the kth class neuron in the output layer and t_k^n is the target for kth class of nth observation feature vector.

The gradient descent method is used in the analysis to adapt the network weights. We have used adaptive serial learning from the data using minimum mean square error criterion. Once the network is trained, the test signal is fed to the neural network and the data is classified to one of the two predefined classes.

2.3.6.3 Support Vector Machine

SVM [18] is a single layer, highly nonlinear network which optimizes the class separation boundary such that the distance from the features falling in a given class to the hyperplane gets simultaneously maximized. SVM is a supervised classifier that has generalization ability [19] in the sense that it can classify an unseen pattern correctly. If $(x_i, y_i), i = 1 : N$ is the data set, x_i is the ith pattern point, and y_i is the corresponding class label, then let $c+$ and $c-$ be the centroids for two classes in binary classification problem. The classifier output will be

$$y_i = sgn((x - c) \cdot w) = sgn((x.c+) - (x.c-) + b) \tag{2.23}$$

where

$$b = \frac{1}{2}(||c\ -||^2 - ||c\ +||^2). \tag{2.24}$$

The optimal hyperplane separating the two classes and satisfying condition given in Eq. 2.23 is

$$minimize_{w,b}\frac{1}{2}\|w\|^2 \tag{2.25}$$

such that

$$y_i((w.x_i) + b) \geq 1, i = 1, \ldots N. \tag{2.26}$$

The Lagrangian dual of Eq. 2.25 is a quadratic programming problem used to find the optimal hyperplane separating the two classes.

2.3.7 k-fold Cross Validation

k-fold cross validation [20] is used for $k = 3$. Here, the total number of samples are sub-sampled into three (k) sets; one set is used for testing, whereas the other two sets are used to train the classifier. The process is repeated two more times such

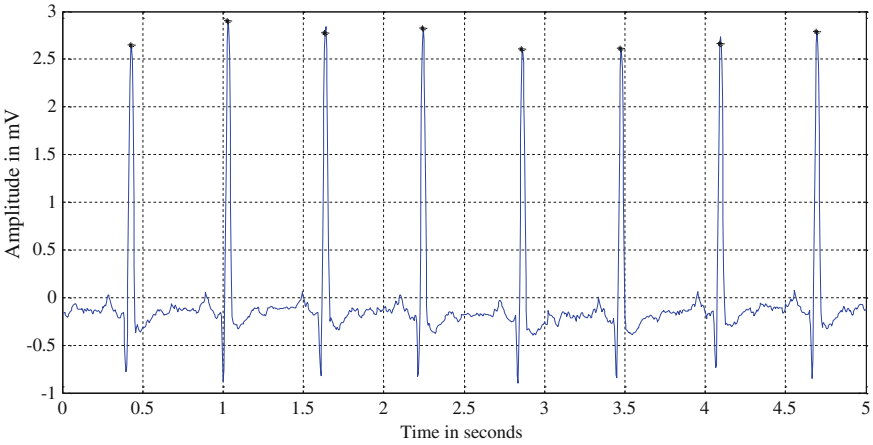

Fig. 2.3 R-point detection in normal sinus rhythm signal

that every sub-partition is used as a testing set and the rest are used for classifier training. The three accuracies are averaged to estimate the average classifier performance. Using *k*-fold cross validation, the bias in choosing the samples from the population can be overcome.

2.4 Results and Discussion

In order to apply our proposed methodology, a two-class ECG classification problem has been formulated based on the MIT BIH arrhythmia and MIT BIH normal sinus rhythm datasets (described in Sect. 2.2). The Pan-Tompkins algorithm is used to detect the R-point because of its simplicity and accuracy. The detection of the R-point is shown in Fig. 2.3, where the detected R-point is marked with a black asterisk. It can be seen from Fig. 2.3 that the Pan-Tompkins method detects the R-point with good precision. In fact, the Pan-Tompkins algorithm is a multistage filtering (differentiation, smoothing, etc.) and a nonlinear element (rectification) between the linear operations in the algorithmic steps.

Once the R-point is detected, a window (or one segment) of 200 samples is extracted by choosing 99 points on the left of the R-point, and 100 points on the right of the R-point and used for further classification. The power spectral density (psd) is computed using an autoregressive method and is plotted for a normal sinus rhythm and arrhythmia signal in Fig. 2.4. The objective of computing psd. is to identify the frequencies of interest so that they can descriminate the two kinds of beats (normal sinus rhythm and arrhythmia) distinctly. We can observe from Fig. 2.4 that frequencies in the range of $0-50$ Hz can be used for that purpose. Hence, by referring to Fig. 2.2 and the graph in Fig. 2.4, it is observed that the subbands of interest are detail 2, detail 3, detail 4, and approximation 4.

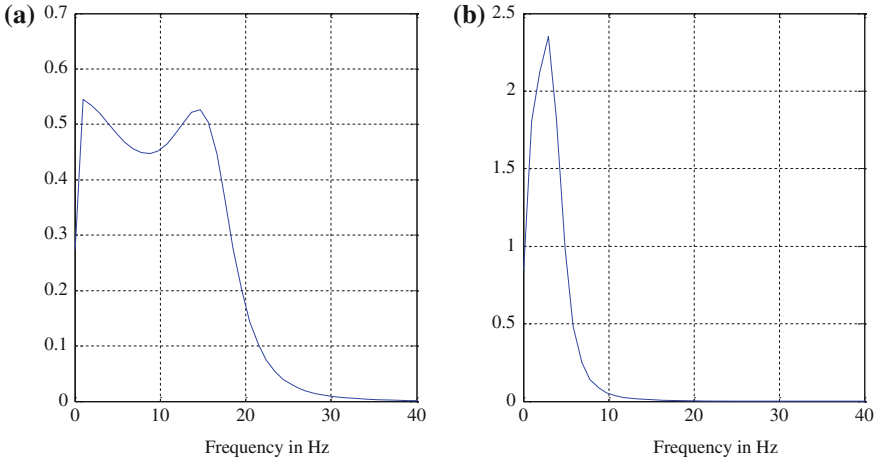

Fig. 2.4 The power spectrum of **a** Normal sinus rhythm, **b** Arrhythmia signal

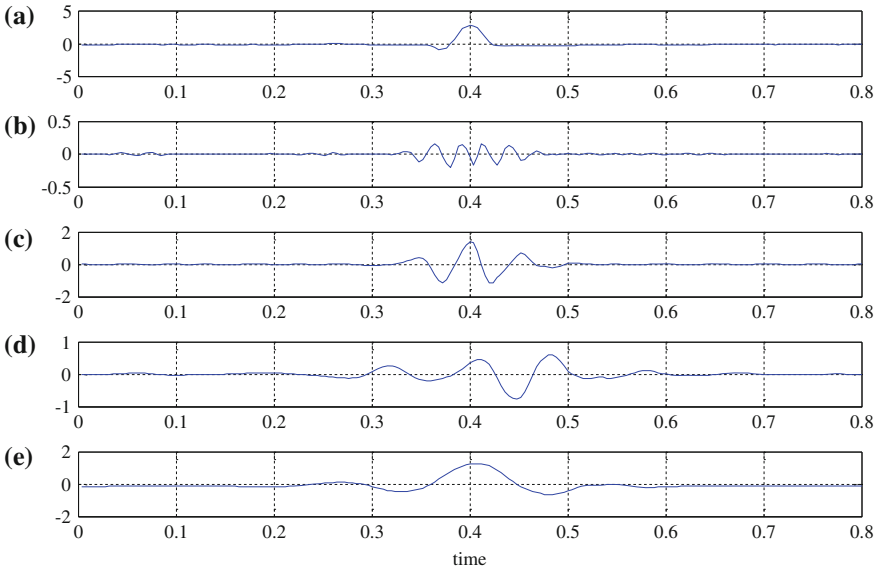

Fig. 2.5 DWT decomposition of normal sinus rhythm signal **a** Original signal, **b** Detail-2, **c** Detail-3, **d** Detail-4, **e** Approximation-4 signals

The DWT using Daubechies-4 wavelet is shown for normal sinus rhythm signal in Fig. 2.5. We can see that all the sub-bands of interest contain some signal component that can be used for performing classification. Figure 2.6 shows the DWT computed using the Daubechies-4 wavelet for an arrhythmia signal. We can see that the DWT decompositions of the two signals look different. If these

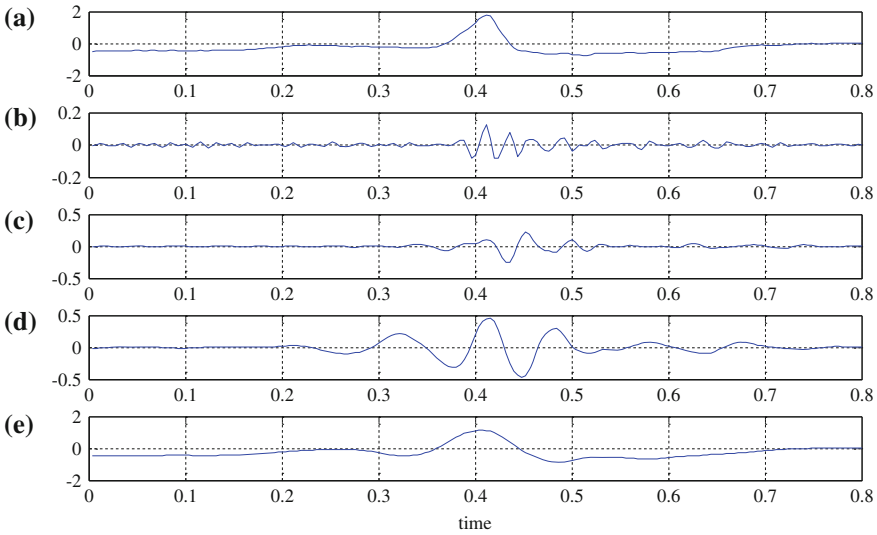

Fig. 2.6 DWT decomposition of arrhythmia signal **a** Original signal, **b** Detail-2, **c** Detail-3, **d** Detail-4, **e** Approximation-4 signals

coefficients are compressed and represented by fewer components, they can be used as features for subsequent classification. The reason for compression is that using fewer components reduces the computational burden on the classifier.

PCA is applied on each sub-band of interest and different wavelet basis functions are used. We use the Daubechies-4, Daubechies-6, Daubechies-8, Symlet-2, Symlet-4, Symlet-6, Symlet-8, Coiflet-2, and Coiflet-5 wavelet basis functions. PCA is an orthogonal transformation which maps the data into the directions of maximum variability. Since DWT is a compact supported basis function, having sparse representation, PCA on it should provide higher compression. The number of principal components is chosen so that the components contain 98 % variability of the respective sub-band. For different basis functions, the number of principal components chosen from each of the sub-bands and the total variability of the data contained is shown in Tables 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9.

The Eigen value profile for the Daubechies-4 wavelet is shown in Fig. 2.7.

After PCA, the compressed components are subjected to a statistical significance test. Based on the p-value provided by the statistical test, the significance of components is decided, and the significant features are used for further classification. The statistical significance test is performed for every basis function and the results are tabulated in Table 2.10. It is observed from Table 2.10 that the DWT domain principal components are more significant than time domain components on the basis of the statistical test.

The GMM classification for the Daubechies-6 wavelet basis function features is shown in Fig. 2.8. We can see that the log likelihood in the graph increases and becomes steady after convergence of the algorithm.

Table 2.1 DWT decomposition using Daubechies-4

Sub-band	Number of PCs	Energy in %
Detail 2	5	98.1770
Detail 3	3	99.2262
Detail 4	2	98.5930
Approximation 4	3	98.8920

Table 2.2 DWT decomposition using Daubechies-6

Sub-band	Number of PCs	Energy in %
Detail 2	3	98.1957
Detail 3	2	99.1075
Detail 4	2	99.4899
Approximation 4	2	98.3697

Table 2.3 DWT decomposition using Daubechies-8

Sub-band	Number of PCs	Energy in %
Detail 2	7	98.1937
Detail 3	2	99.4728
Detail 4	2	99.0444
Approximation 4	2	98.5753

Table 2.4 DWT decomposition using Symlet-2

Sub-band	Number of PCs	Energy in %
Detail 2	2	98.2329
Detail 3	3	99.4611
Detail 4	2	99.3243
Approximation 4	2	98.4532

Table 2.5 DWT decomposition using Symlet-4

Sub-band	Number of PCs	Energy in %
Detail 2	10	98.0661
Detail 3	2	99.0872
Detail 4	2	99.6493
Approximation 4	2	98.6254

The classification by EBPNN is shown in Fig. 2.9, which shows that the total mean-squared error reduces with epochs. We can observe that the EBPNN algorithm converges in 19 epochs for Daubechies-6 wavelet family features. Figure 2.10 shows SVM classification with linear kernel. Since the data is linearly separable, we have used only linear kernel SVM.

Table 2.6 DWT decomposition using Sym-6

Sub-band	Number of PCs	Energy in %
Detail 2	3	98.1292
Detail 3	2	98.8807
Detail 4	2	99.5307
Approximation 4	2	98.1622

Table 2.7 DWT decomposition using Sym-8

Sub-band	Number of PCs	Energy in %
Detail 2	10	96.6749
Detail 3	2	98.8332
Detail 4	2	99.6216
Approximation 4	3	98.6306

Table 2.8 DWT decomposition using Coiflet-2

Sub-band	Number of PCs	Energy in %
Detail 2	10	96.9230
Detail 3	2	99.1172
Detail 4	2	96.6110
Approximation 4	3	98.6443

Table 2.9 DWT decomposition using Coiflet-5

Sub-band	Number of PCs	Energy in %
Detail 2	10	96.0415
Detail 3	2	98.2607
Detail 4	2	99.6023
Approximation 4	2	98.5899

Table 2.11 shows classification accuracies of various schemes using different wavelet basis functions. It can be observed from Table 2.11 that EBPNN provides higher accuracy than GMM and SVM leads to the highest accuracy. Amongst various wavelet families, it can be noted that Symlet-2 consistently performs better for all the classifiers and has the highest possible accuracy.

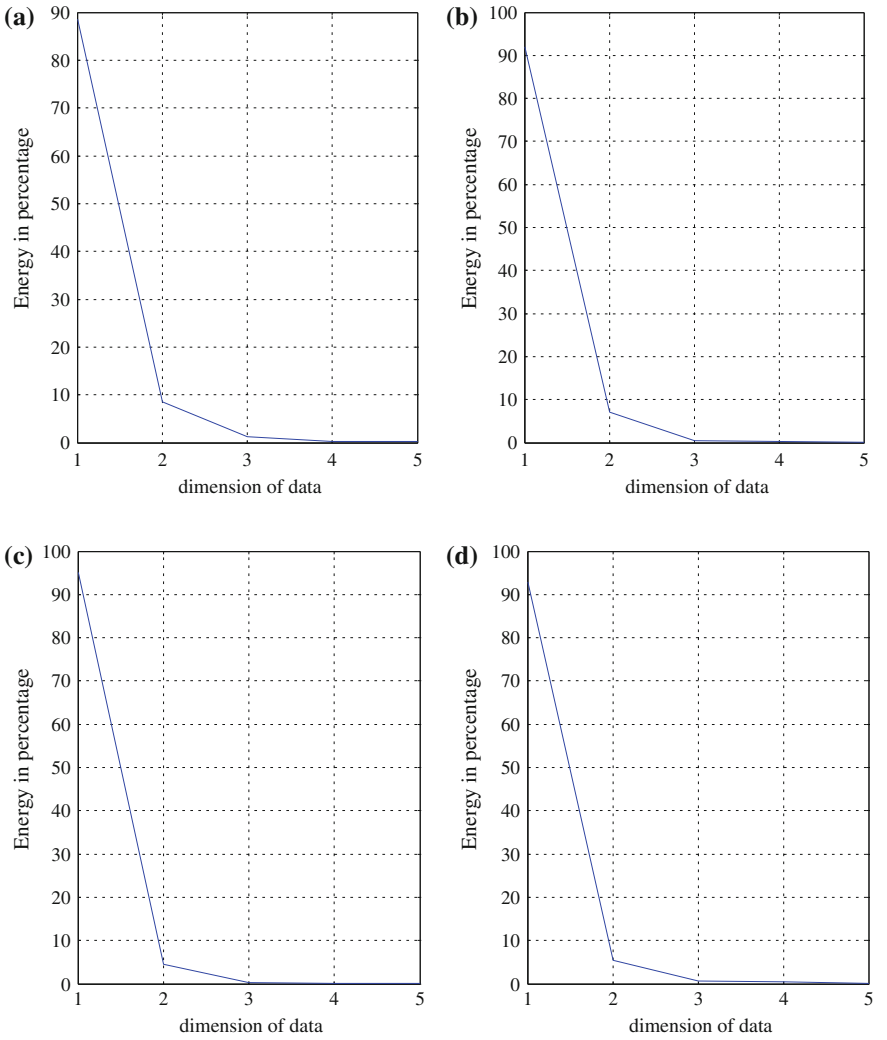

Fig. 2.7 PCA on DWT sub-bands **a** Detail 2, **b** Detail 3, **c** Detail 4, and **d** Approximation 4, decomposition using the db6 wavelet

Table 2.10 Statistical significance test for time domain and Daubechies-6 based DWT sub-band features

Time domain PCs	Statistical significance		DWT domain PCs	Statistical significance	
	t	p		t	p
1	−61.998	0.000	1	−70.9029	0.000
2	−156.68	0.000	2	−52.5739	0.000
3	0.1702	0.8650	3	−3.4508	0.000
4	−1.7762	0.0763	4	−28.7637	0.000
5	0.4157	0.6778	5	37.3751	0.000
6	0.4035	0.6867	6	−27.1491	0.000
7	−0.1840	0.8541	7	22.3390	0.000
8	−0.3165	0.7517	8	10.5710	0.000
9	0.8103	0.4181	9	−93.7320	0.000

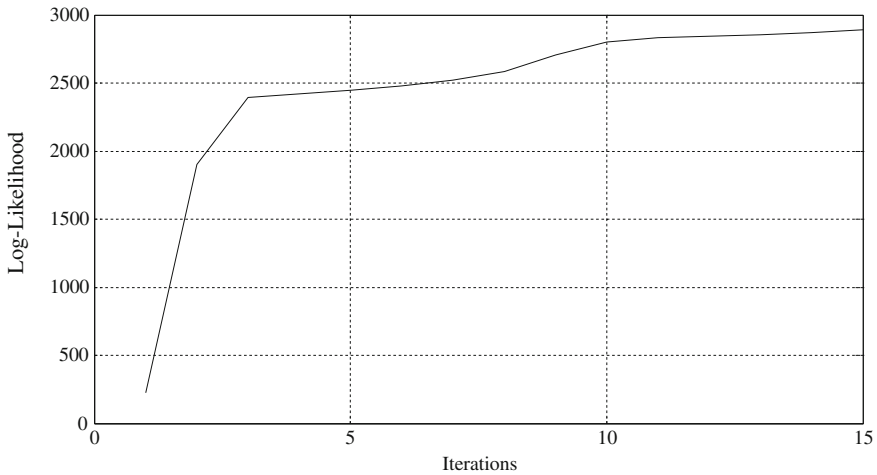

Fig. 2.8 GMM classification, log-likelihood increasing with iterations

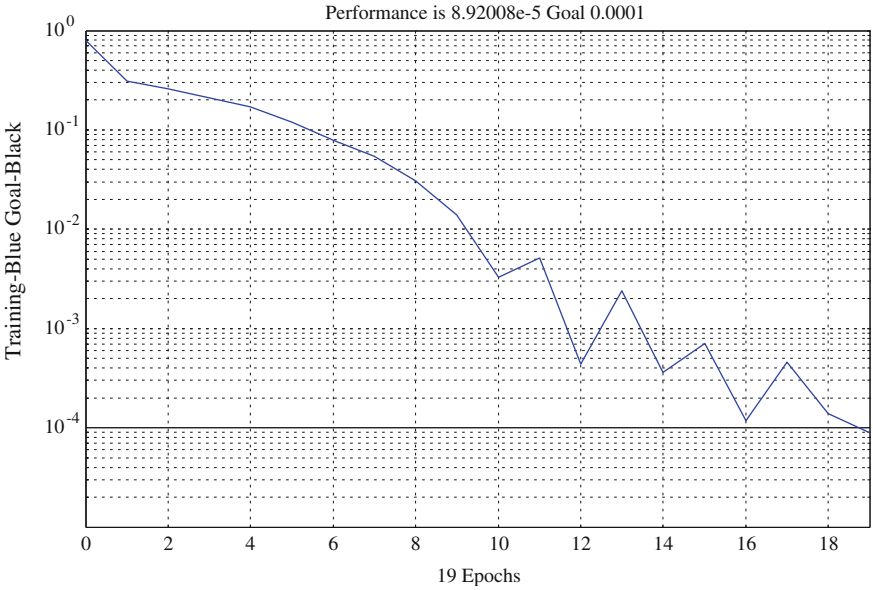

Fig. 2.9 EBPNN classification, network converging in 19 epochs for Daubechies-6 features

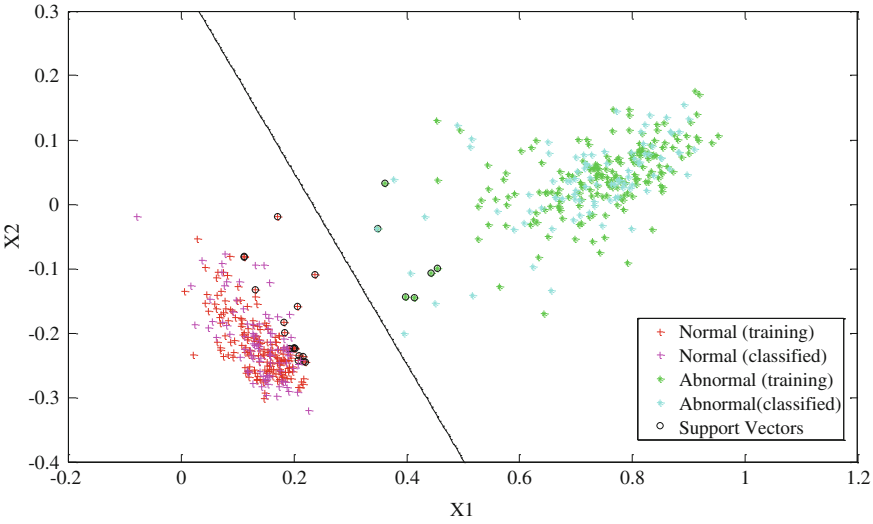

Fig. 2.10 SVM classification using linear kernel for Daubechies-6 wavelet features

Table 2.11 The average classification accuracies of various classifiers using different wavelet basis functions

Wavelet basis function	GMM	EBPNN	SVM
Db4	87.36	93.41	95.60
Db6	88.52	94.23	96.72
Db8	85.27	93.78	95.39
Sym2	89.78	95.17	97.23
Sym4	85.18	93.79	96.81
Sym 6	88.21	94.18	97.14
Sym8	84.39	93.89	96.68
Coif 2	85.91	92.96	95.95
Coif 5	85.16	92.83	95.42

2.5 Conclusion

In this chapter, a systematic approach is developed for screening arrhythmia and normal sinus rhythm from their ECG profiles. We have extracted time frequency features using various basis functions, including Daubechies, Symlet, and Coiflet wavelet families. PCA is applied on time frequency sub-band features and, in this compact supported basis space, higher compression is expected. Based on our experiments, we have determined that different basis functions distribute energy in different sub-bands in a unique way for a given wavelet. Our methodology exploits this energy distribution so that the features are well represented, thus resulting in higher accuracy. These time-frequency features are markers of disease, since these features are able to discriminate the data into two classes. As a future direction, other time-frequency representations can be used to see how the energy compaction is achieved. In addition, various other dimensionality reduction techniques can be used for performance. The machine-learning methodology given in this chapter can be used efficiently in telemedicine systems to identify abnormal events in the ECG signals so that emergency cases can be identified and such patients can be attended for critical care.

References

1. Fauci AS, Braunwald E, Kesper DL, Hauser SL, Longo DL, Jamesonn JL, Loscalzo J (2008) Harrison's principles of internal medicine, 17th edn. Mc-Graw Hill, New York
2. Park K (2005) Park's textbook of preventive and social medicine, 18th edn. Banarsidas Bhanot publishers, India
3. Guyton AC, Hall JE (2006) Textbook of medical physiology, 11th edn. W. B Saunders Co, Philadelphia
4. Pan J, Tompkins WJ (1985) A real time QRS detection algorithm. IEEE Trans Biomed Eng 32(3):230–236
5. Li C, Zheng C, Tai C (1995) Detection of ECG characteristic points using wavelet transforms. IEEE Trans Biomed Eng 42(1):21–29

6. Martinez JP, Almeida R, Olmos S, Rocha AP, Laguna P (2004) A wavelet based ECG delineator: evaluation on standard databases. IEEE Trans Biomed Eng 51(4):570–581
7. Throne RD, Jenkins JM, Winston SA, DiCarlo LA (1991) A comparison of four new time domain techniques for discriminating monomorphic ventricular tachycardia from sinus rhythm using ventricular waveform morphology. IEEE Trans Biomed Eng 38(6):561–570
8. Krasteva V, Jekova I (2007) QRS template matching for recognition of ventricular ectopic beats. Ann Biomed Eng 35(12):2065–2076
9. Martis RJ, Chakraborty C, Ray AK (2009) A two stage mechanism for registration and classification of ECG using gaussian mixture model. Pattern Recogn 42(11):2979–2988
10. Martis RJ, Krishnan MM, Chakraborty C, Pal S, Sarkar D, Mandana KM, et al (2012) Automated screening of arrhythmia using wavelet based machine learning techniques. J Med Syst 36(2):677–688
11. Vaidyanathan PP (2004) Multirate systems and filter banks. Pearson education (Asia) Pte. Ltd, Delhi
12. Oppenheim AO, Schaffer RA (2003) Discrete time signal processing. Mc-Graw Hill edition, New York
13. Addison PS (2005) Wavelet transforms and the ECG: a review. Physiol Meas 26(5): R155–199
14. Strang G, Nguyen T (1996) Wavelets and filter banks. Willesley Cambridge Press, MA
15. Duda R, Hart P, Stork D (2001) Pattern classification, 2nd edn. Wiley, New York
16. Gun AM, Gupta MK., Dasgupta B (2008) Fundamentals of statistics (Vol. I and II), 4th edn. World Press Private Ltd, Kolkata
17. Bishop C (1995) Neural networks for pattern recognition. Oxford University press, New York
18. Christianini N, Taylor JS (2000) An introduction to support vector machines and other kernel based learning methods, Cambridge university press, Cambridge
19. Gunn S (1998) Support vector machines for classification and regression, Technical report, University of Southampton
20. Schneider J (1997) Cross validation. http://www.cs.cmu.edu/∼schneide/tu5/node42.html. Accessed 15 Aug 2010

http://www.springer.com/978-3-642-40016-2

Machine Learning in Healthcare Informatics

Dua, S.; Acharya, U.R.; Dua, P. (Eds.)
2014, XII, 332 p. 119 illus., 50 illus. in color., Hardcover
ISBN: 978-3-642-40016-2