Discriminant analysis of multivariate time series: Application to diagnosis based on ECG signals

Elizabeth Ann Maharaj[a], Andrés M. Alonso[b]

[a] *Department of Econometrics and Business Statistics, Monash University, Australia*
[b] *Department of Statistics, Universidad Carlos III de Madrid, Spain*

Abstract

In analyzing ECG data, the main aim is to differentiate between the signal patterns of healthy subjects and those of individuals with specific heart conditions. We propose an approach for classifying multivariate ECG signals based on discriminant and wavelet analyses. For this purpose we use multiple-scale wavelet variances and wavelet correlations to distinguish between the patterns of multivariate ECG signals based on the variability of the individual components of each ECG signal and on the relationships between every pair of these components. Using the results of other ECG classification studies in the literature as references, we demonstrate that our approach applied to 12-lead ECG signals from a particular database compares favourably. We also demonstrate with real and synthetic ECG data that our approach to classifying multivariate time series out-performs other well-known approaches for classifying multivariate time series.

Keywords:

Discriminant Analysis, Wavelet Variances, Wavelet Correlations, ECG Signals.

Accepted at Computational Statistics and Data Analysis

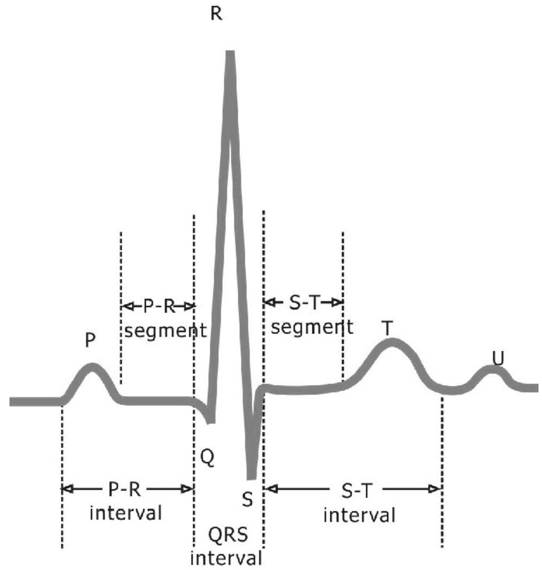
https://doi.org/10.1016/j.csda.2013.09.006

1. Introduction

An electrocardiogram (ECG) is a test that records the electrical activity of the heart. The ECG is recorded between electrodes on the surface of the human body. The standard is the 12-lead ECG which examines the electrical activity of the heart from 12 points of view. Cardiac abnormalities have been conventionally monitored and diagnosed by human observers relying on the detection of the presence of particular features in the output of the ECG. This output is a graph (or sometimes several graphs, representing each of the leads) with time represented on the x-axis and voltage represented on the y-axis. A dedicated ECG machine would usually print onto graph paper which has a background pattern of 1 mm squares, with bold divisions every 5 mm in both vertical and horizontal directions. By definition, a 12-lead ECG will show a short segment of the recording of each of the 12 leads. This is often arranged in a grid of four columns by three rows, the first columns being the limb leads (i,ii and iii), the second column the augmented limb leads (aVR, aVL and aVF) and the last two columns being the chest leads (V1-V6). Each column will usually record the same moment in time for the three leads and then the recording will switch to the next column, which will record the heart beats after that point. It is possible for the heart rhythm to change between the columns of leads.

A typical waveform from a single heart beat [1] is shown in Figure 1 on Page 3. Any ECG signal imparts two pieces of information. The first is the duration of the electrical wave crossing the heart which in turn works out whether the electrical activity is normal, slow, or irregular , and the second is the amount of electrical activity passing through the heart muscle which determined whether parts of the heart are too large or overworked. Normally, the frequency range of an ECG signal is of 0.05-100 Hz and its dynamic range is of 1 - 10 mV (mV represents milli-volts; a volt is the unit for electric potential (voltage); Hz represents a hertz which is the inverse of a second). The ECG is characterised by five peaks and valleys labeled by the letters P, Q, R, S, T. In some cases another peak, U is also present. The performance of an ECG analysing system

Figure 1: A typical ECG waveform of a single beat



depends mainly on the accurate and reliable detection of the QRS complex, as well as the T- and P-waves. The P-wave represents the activation of the upper chambers of the heart, the atria, while the QRS complex and the T-wave represent the excitation of the lower chambers of the heart, the ventricles. The QRS complex is the most prominent part of the waveform within the ECG signal. Since it reflects the electrical activity within the heart during ventricular contraction, the time of its occurrence as well as its shape provide some information about the current state of the heart.

Interpretation of the ECG allows diagnosis of a wide range of heart conditions which can vary from minor to life threatening. Due to large numbers of people with suspected heart conditions in hospitals all around the world, the need has arisen for automated detection of heart abnormalities to assist diagnoses. Hence, the classification of ECG signals is an important task in biomedical science. Large databases exist for large numbers of patients and controls, Many techniques of automatic detection using ECG signals have been developed, enabling objective quantitative classification, where previously just qualitative diagnostic criteria were applied.

Many authors have proposed methods to classification ECG signals with varying degrees of accuracy. In particular, Al-Naima and Al-Timemy [2], De Chazel and Reilly

[3] , Heden et al. [4] and Bozzola et al. [5] have proposed methods that been applied to 12-lead ECG signals.

In our study of 12-lead ECG signals , we are particularly interested in differentiating between patterns of ECG signals of individuals with the heart condition, myocardial infarction, and of those without this condition. To this end, Al-Naima and Al-Timemy [2] used Discrete Fourier Transform (DFT) coefficients and Discrete Wavelet Transform (DWT) coefficients as the discriminating features with neural network classifiers. Using 12-lead ECG data with a training set of 45 records (26 controls and 19 with myocardial infarction) and a test set of 20 records (12 controls and 8 with myocardial infarction), the sensitivities achieved for the test set were between 80% to 90%, while the specificity was 90%. Note that sensitivity is the percentage of cases correctly classified as being in the class of a particular condition, e.g., myocardial infarction, while specificity is the percentage of cases correctly classified in the control group, i.e., being of the normal class.

De Chazel and Reilly [3] used linear and quadratic discriminants with five features sets, including DWT coefficients, standard cardiology features and time domain features. Using a database of 500 12-lead ECG signals from 345 patients with different cardiac diseases, and from 155 controls, and multiple runs of ten-fold cross-validation, they obtained sensitivities of 69-73%, 78-83% and 25-38% when they classified anterior, inferior or combined myocardial infarction, while the specificity of their best classifier was 90%.

Heden et al. [4] conducted a study using 12-lead ECG records of a group of 1120 individuals with acute myocardial infarction and a control group consisting of 10,352 cases. They used six time domain measurements from each of the 12 leads as inputs in to artificial neural networks. Using a three-fold cross-validation procedure, their method achieved sensitivities of 46.2%-65.9% and specificities of 86.3-95.4%.

Bozzola et al. [5] extracted a set of 8 time domain parameters from each of the 12-lead ECG records and input these into a hybrid neuro-fuzzy system for the classification of myocardial infarction. They used a training set of ECG records of 179 controls and 404 with myocardial infarction, and a test set of ECG records of 60 controls and 135 with myocardial infarction. Their method achieved test set sensitivities of 72% , 80-88% and 52-60% when they classified anterior, inferior or combined myocardial infarction and specificities of 92-93%.

In all of the techniques employed by the above-mentioned authors, the components of the 12-lead ECG signal are treated as if they were independent of each other. In practice, each 12-lead ECG signal can be regarded as a 12-component multivariate time series. An important consideration in the analysis of multivariate time series is the relationship between the individual components of each series. Hence, given the absence of this consideration in ECG classification literature, we are motivated to examine the inclusion of these interrelationship features to assess whether they provide useful information, and thus lead to perhaps more accurate classification results. In this paper, we therefore propose the discriminant analysis of the multivariate time series, namely, the 12-lead ECG signals, based on wavelet features of variances and correlations. In using these features, our goal is to distinguish between the patterns of multivariate ECG signals based on the variability of the individual components of each ECG signal, and the relationships between these components. Taking into account the relationships between every pair of components is a novel approach, and to our knowledge has not been considered before in multiple-lead ECG classification.

Since an ECG signal is of high frequency, time domain analyses of such signals, using summary statistics misses out on useful information in the intricacies of the signal. Hence a frequency representation of the signal would be more useful. This can be achieved by using the Fast Fourier Transform (FFT) technique. However, the limitation of the FFT is that it is unable to provide that information regarding the exact location of the frequency components in time. Wavelet transformation of the time series overcomes this limitation in that the signal is decomposed into coefficients at a number of frequency bands referred to as scales. The coefficients at each scale have a time location and represent the ECG signal in a particular frequency band. Furthermore, an advantage of using wavelet features is that the time series under consideration do not necessarily have to be mean stationary or variance stationary as indeed most ECG data are not. Because we are interested in the variability of each of the 12-lead ECG components we can obtain more useful information from the variances of the the wavelet coefficients at each of the frequency bands than simple obtaining the single variance of each lead in the time domain . Since we are also interested in the relationship between the every pair of leads, we can obtain more useful information from the correlation between every pair of leads at each of the frequency bands, instead of a single correlation coefficient between a pair of leads in the time domain. All of this therefore provides a motivation for the use of these wavelet based features.

We first demonstrate the performance of our approach on synthetic ECG data and we compare it to that of other well-known methods for classifying multivariate time series. We then apply our approach to data from a publicly available ECG database and we show that this approach achieves very good accuracy rates in differentiating between the 12-lead ECG records of healthy individuals and those with myocardial infarction.

In Section 2 we describe the tools required for our proposed method and the method itself, respectively. In Section 3 we evaluate the performance of this approach using synthetic ECG data when compared to other multivariate classification approaches. In Section 4, we discuss the processing of ECG records available in the PTB Diagnostic ECG Database [6] and we apply this method to differentiate between the ECG signals of 52 healthy controls and 148 subjects with myocardial infarction. This ECG data allows us to evaluate the performance of our proposed method and compare it with other multivariate classification approaches. We conclude in Section 5.

2. Methods

2.1. Discriminant Analysis

In practice, a time series is known to belong to one of g groups. The task is to classify the time series into one of these g groups in an optimal manner. Assumptions are made concerning the Gaussian probability density function of the different groups. In linear discriminant analysis, it is assumed that the groups have equal covariance matrices and differ only in their means. While in quadratic discriminant analysis, the covariance matrices of the groups are not assumed to be equal. There are several ways to evaluate the performance of a discriminant analysis procedure. One method is to split the sample into training and hold-out samples and evaluate the error rate associated with the hold-out set which was not used in deriving the classification rule. Another method is to use the hold-out-one technique of cross-validation which is particularly useful if the samples sizes are not very large. This technique holds out the observation to be classified, deriving the classification function from the remaining observations (see [8] for more details). The procedure is repeated of each member of the sample and an overall error rate is determined.

When dealing with univariate time series, one can use features of the time series such as autocorrelations, periodogram coefficients, wavelet features etc., in a standard discriminant analysis. Several authors have proposed discriminant analysis of univariate time series. See for example, [9] and [10]. With multivariate time series, the task is more complex because as well as taking into account the multiple series associated with each object, one has to also take into account the relationships between the components of each multivariate time series. Kakizawa et al. [11] developed approaches for classifying stationary multivariate time series using spectral matrices with Kullback–Leibler (K–L) and Chernoff discrepancy measures. Shumway [12] extended this with the Kullback–Leibler (K–L) discrepancy to locally stationary time series.

2.2. Wavelet Analysis and Wavelet Features

In what follows, using the notation of [13], we give a brief description of wavelet analysis and the associated features of wavelet variances and wavelet correlations.

The Discrete Wavelet Transform (DWT), which is an orthonormal transform, re-expresses a time series of length T in terms of coefficients that are associated with a particular time and with a particular dyadic scale as well as one or more scaling coefficients. The j-th dyadic scale is of the form 2^{j-1} where $j = 1, 2, \ldots, J$, and J is the maximum allowable number of scales.

The number of coefficients at the j-th scale is $T/2^j$, provided $T = 2^J$. In general the wavelet coefficients at scale 2^{j-1} are associated with frequencies in the interval $[1/2^{j+1}$,$1/2^j]$. Large time scales give more low frequency information, while small time scales give more high frequency information about the time series. It is possible to recover the time series $X_t, t = 1, 2, \ldots, T$ from its DWT by synthesis. That is, the multi-resolution analysis (MRA) of a time series is expressed as

$$X_t = \sum_{j=1}^{J} d_j + s_J, \tag{1}$$

where d_j is the wavelet detail (series of inverse wavelet coefficients at scale j) and s_J is the smooth series which is the inverse of the series of scaling coefficients. Hence a time series and its DWT are actually two representations of the same mathematical entity.

The maximum overlap discrete wavelet transform (MODWT) is a modification of the DWT. Under the MODWT, the number of wavelet coefficients created will be the same as the number of observations in the original time series. Because the MODWT decomposition retains all possible times at each time scale, the MODWT has the advantage of retaining the time invariant property of the original time series. The MODWT can be used in a similar manner to the DWT in defining a multi-resolution analysis of a given time series. In contrast to the DWT, the MODWT details and smooths are associated with zero phase filters making it easy to line up features in a MRA with the original time series more meaningfully.

If $\{\tilde{h}_{j,l}, l = 0, 1, \ldots, L_j\}$ is the j-level MODWT wavelet filter of length L_j, associated with scale $\tau_j \equiv 2^{j-1}$ then if $\{X_t\}$ is a discrete parameter stochastic process and

$$W_{j,t} \equiv \sum_{l=0}^{L_j} \tilde{h}_{j,l} X_{t-l} \tag{2}$$

represents the stochastic process by filtering $\{X_t\}$ with the MODWT filter $\{\tilde{h}_{j,l}\}$, and if it exists and is finite, the time independent MODWT wavelet variance at the j-th dyadic scale $\tau_j \equiv 2^{j-1}$ is defined as

$$\nu_X^2(\tau_j) \equiv \mathrm{var}\{W_{X,j,t}\}. \tag{3}$$

It can be shown that

$$\sum_{j=1}^{\infty} \nu_X^2(\tau_j) = \mathrm{var}\{X_t\}, \tag{4}$$

i.e., the wavelet variance decomposes the variance of the stochastic process across scales (see [13], p296-298 for more details).

Given a time series $x_t, t = 1, 2 \ldots, T$, which is a realization of the stochastic process X_t, an unbiased estimator of $\nu_X^2(\tau_j)$ is

$$\widehat{\nu}_X^2(\tau_j) \equiv \frac{1}{M_j} \sum_{t=L_j-1}^{T-1} \widehat{W}_{X,j,t}^2, \tag{5}$$

where $\widehat{W}_{j,t}^2$ are the MODWT coefficients associated with the time series x_t and $M_j = N - L_j + 1$ is the number of wavelet coefficients excluding the boundary coefficients that are affected by the circular assumption of the wavelet filter.

Given two appropriate stochastic processes $\{X_t\}$ and $\{Y_t\}$ with MODWT coefficients $W_{Xj,t}$ and $W_{Yj,t}$, respectively, the wavelet covariance is defined as

$$\nu_{XY}(\tau_j) \equiv \mathrm{cov}\{W_{Xj,t}, W_{Yj,t}\}, \tag{6}$$

and it gives the scale-based decomposition of the covariance between $\{X_t\}$ and $\{Y_t\}$. The wavelet covariance can be standardized to yield the wavelet correlation

$$\rho_{XY}(\tau_j) \equiv \frac{\nu_{XY}(\tau_j)}{\nu_X(\tau_j)\nu_Y(\tau_j)}. \tag{7}$$

For time series x_t and y_t which are realizations of $\{X_t\}$ and $\{Y_t\}$ respectively, replacing the wavelet variances and covariance by their unbiased estimators, we get the estimated wavelet correlation

$$\widehat{\rho}_{XY}(\tau_j) \equiv \frac{\widehat{\nu}_{XY}(\tau_j)}{\widehat{\nu}_X(\tau_j)\widehat{\nu}_Y(\tau_j)}. \tag{8}$$

2.3. Implementation

In what follows, we will apply discriminant analysis to sets of multivariate time series using the wavelet variances and wavelet correlations as the discriminating variables or features. Maharaj and Alonso [10] used wavelet variances as features for the discrimination of univariate time series while D'Urso and Maharaj [14] used wavelet variances and wavelet correlations in fuzzy clustering algorithms to cluster multivariate time series.

Wavelet filters of lengths 2, 4, 6 and 8 of the Daubechies family (db2, db4, db6, db8), of length 8 from the Symmletts family (sym8), and of length 6 from the Coiflets family (cf6) will be used to generate the MODWT coefficients, and hence the MODWT variances and correlations of the signal. These are commonly used filters to decompose discrete time series, hence their choice. For more details on the wavelet filters refer to Chapter 4 of [13]. In order to ensure that the boundary coefficients which have an effect on the estimated scale by scale wavelet variance and correlations are excluded, only a specific number of scales should be used for each filter, depending on the sample size.

For example, if the length of each of multivariate time series under consideration which consists of two components is, say, $T = 64 = 2^J$ where $J = 6$, and we use a filter that enables decomposition up to $J - 2 = 4$ scales, then each component time series would be decomposed into MODWT series at of four scales. Hence, associated with each component MODWT time series would be four wavelet variances, and a wavelet correlation between every scale-based pair of component MODWT time series . Therefore, there would be 12 features, (8 wavelet variances and 4 wavelet correlations) associated with each multivariate time series under consideration. In practice, as the number of components of the multivariate time series and their lengths increase, the number of wavelet variances and correlations and hence the number of feature variables would also increase.

In Sections 3 and 4, we will be implementing the procedures for series lengths, 4096 and 8192, i.e., 2^J, with $J = 12$ and $J = 13$, respectively. In Section 3, we will be considering synthetic ECG data with 3 lead while in Section 4, we will be considering actual ECG data with 12 leads. Table 1 on Page 11 shows the maximum allowable number of scales for each of the filters for series length 2^J (see [13] p. 136 for more details), as well as the number of wavelet variances and wavelet correlations that will be considered in the initial implementation of the procedures in Sections 3 and 4. In what

Table 1: Maximum allowable number of scales and number of wavelet features.

Wavelet filter	db2	db4	db6	db8	sym8	cf6
Scales $T = 2^J$	*J*	*J-2*	*J-3*	*J-3*	*J-3*	*J-3*
Scales $T = 2^{12}$, 3 leads	12	10	9	9	9	9
Wavelet variances	36	30	27	27	27	27
Wavelet correlations	36	30	27	27	27	27
Scales $T = 2^{13}$, 12 leads	13	11	10	10	10	10
Wavelet variances	156	132	120	120	120	120
Wavelet correlations	858	726	660	660	660	660

follows, we will implement a stepwise discrimination procedure to select the features (discriminating variables) that will achieve the minimum hold-out classification error rate. We use the stepwise implementation of [15] that selects the relevant variables (in this case the variables being the wavelet variances and correlations from the various scales) in order to minimize the misclassification error. We use the hold-out-one cross-validation technique to evaluate the performance of our procedure.

Assumptions of multivariate normality are made about the probability distribution of the group feature variables in linear and quadratic discriminant analysis. It can be shown that under the assumption of multivariate normality, the sample linear and quadratic discriminant functions are asymptotically optimal in the presence of homoscedasticity and heteroscedasticity, respectively (see [16]). Serroukh et al. [17] have shown that MODWT wavelet variance estimators are asymptotically normal for linear processes, while Serroukh and Walden [18] have shown that for bivariate linear processes, the MODWT wavelet covariance estimators are asymptotically normal. It follows that the MODWT wavelet correlation estimators are also asymptotically normal. While a situation is which all variables under consideration are shown to exhibit univariate normality may help achieve multivariate normality, it will not guarantee it. Thus, the group feature variables of the combined MODWT wavelet variances and wavelet correlations may not necessarily be asymptotically multivariate normal.

In most real applications of discrimination analysis, the assumption of multivariate normality might not be strictly met. Many authors have conducted studies on the robustness of the discriminant functions and have found that some of them are fairly robust to departures from assumed models with little or no modification (see e.g., [19], [20]). Furthermore since asymptotic normality of the MODWT wavelet variance and of the MODWT wavelet covariance are based on the assumption that the underlying univariate and bivariate processes are linear (see [17] and [18]), this assumption of linearity will not necessarily be met for the ECG data under consideration in Sections 3, and 4. However, we will proceed with using the wavelet feature variables in linear and quadratic discriminant analysis even though the multivariate normality assumption may not be strictly satisfied by these feature variables.

3. Discriminant Analysis of Synthetic ECG data

In this section, we evaluate the performance of the discriminant analysis with the wavelet features using synthetically generated ECG data. We use the generator available at the Open-Source Electrophysiological Toolbox [21] that implements the dynamic model developed by McSharry et al. [22] and extended to multichannel ECG by Sameni et al. [23] and Clifford et al. [24]. The model is a three-dimensional formulation of single dipole of the heart:

$$\begin{aligned}
\dot{\theta} &= \omega \\
\dot{x} &= -\textstyle\sum_i \frac{\alpha_i^x \omega}{(b_i^x)^2} \Delta\theta_i^x \exp\left[\frac{-(\Delta\theta_i^x)^2}{(b_i^x)^2}\right] \\
\dot{y} &= -\textstyle\sum_i \frac{\alpha_i^y \omega}{(b_i^y)^2} \Delta\theta_i^y \exp\left[\frac{-(\Delta\theta_i^y)^2}{(b_i^y)^2}\right] \\
\dot{z} &= -\textstyle\sum_i \frac{\alpha_i^z \omega}{(b_i^z)^2} \Delta\theta_i^z \exp\left[\frac{-(\Delta\theta_i^z)^2}{(b_i^z)^2}\right],
\end{aligned}$$

where $\Delta\theta_i^x = \theta - \theta_i^x \mod 2\pi$, $\Delta\theta_i^y = \theta - \theta_i^y \mod 2\pi$, $\Delta\theta_i^z = \theta - \theta_i^z \mod 2\pi$ and $\omega = 2\pi f$, where f is the beat-to-beat heart rate. The three coordinates of the dipole, x, y and z, are modelled by a summation of Gaussian function with amplitudes α_i^x, α_i^y and α_i^z;

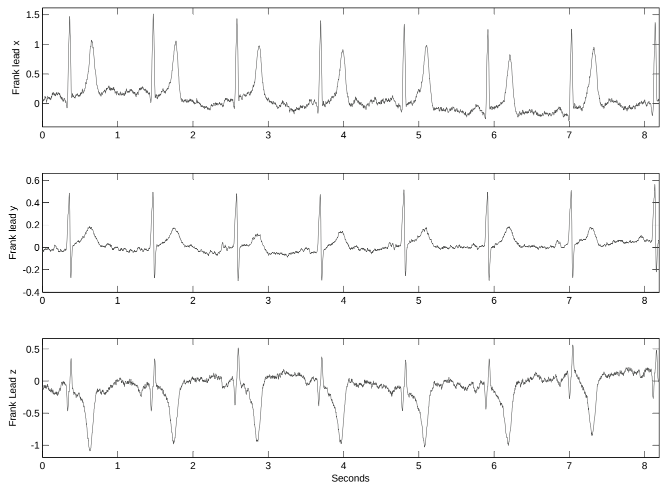


Figure 2: Synthetic ECG signals.

widths b_i^x, b_i^y and b_i^z; located at rotational angles θ_i^x, θ_i^y and θ_i^z, respectively. For details see [23] and [25].

In Table 2 on Page 14 we reproduce the values of the parameters of the above synthetic model used in [23] and [25]. As these authors suggest we add some random values to the parameters in order to generate more realistic ECG patterns. Particularly, in our simulation experiment we have used $\theta_i^x = \theta_i^x + Z_i^x b_i^x$, $\theta_i^y = \theta_i^y + Z_i^y b_i^x$ and $\theta_i^z + Z_i^z b_i^z$, where Z_i^x, Z_i^y and Z_i^z are independent standard Gaussian random variables. Additionally, the generator allows the addition of some coloured noise to the synthetic generated ECG.

Figure 2 on Page 13 shows a three lead ECG generated by this procedure with time in seconds on the horizontal axis. In order to generate two different population using this procedure, we generate a sample of random ECG signals with the parameters in Table 2 on Page 14 and another sample using the same parameters with the exceptions of $a_{10}^x = 0.39 * \lambda$, $a_8^y = 0.08 * \lambda$ and $a_9^z = 0.35 * \lambda$ with $\lambda > 1$.

Figure 3 on Page 15 with time in seconds on the horizontal axis, illustrates the differences in the two generated signals using $\lambda = 1$ and $\lambda \in \{1.25, 1.5, 1.75, 2\}$. The signals generated with $\lambda > 1$ show a hyperacute T wave which is the first manifes-

Table 2: Parameters of the synthetic ECG model.

Index(i)	1	2	3	4	5	6	7	8	9	10	11
α_i^x (mV)	0.03	0.08	-0.13	0.85	1.11	0.75	0.06	0.10	0.17	**0.39**	0.03
b_i^x (rad)	0.09	0.11	0.05	0.04	0.03	0.03	0.04	0.60	0.30	0.18	0.50
θ_i^x (rad)	-1.09	-0.83	-0.19	-0.07	0.00	0.06	0.22	1.20	1.42	1.68	2.90
α_i^y (mV)	0.04	0.02	-0.02	0.32	0.51	-0.32	0.04	**0.08**	0.01		
b_i^y (rad)	0.07	0.07	0.04	0.06	0.04	0.06	0.45	0.30	0.50		
θ_i^y (rad)	-1.10	-0.90	-0.76	-0.11	-0.01	0.07	0.80	1.58	2.90		
α_i^z (mV)	-0.03	-0.14	-0.04	0.05	-0.40	0.46	-0.12	-0.20	**-0.35**	-0.04	
b_i^z (rad)	0.03	0.12	0.04	0.40	0.05	0.05	0.80	0.40	0.20	0.40	
θ_i^z (rad)	-1.10	-0.93	-0.70	-0.40	-0.15	0.10	1.05	1.25	1.55	2.80	

tation of acute myocardial infarction. For each population, we generate 100 ECG of length $T = 2^{12} = 4096$ and we apply the different discriminant procedures with the wavelet features and the Kullback-Leibler discrimination information and the Chernoff information procedures. These measures have been described in detail by Kakizawa et al. [11] and by Shumway and Stoffer [26]. For the wavelet-based procedures we used wavelet filters from the Daubechies family (db2, db4, db6, db8), from the Symmletts family (sym8) and from the Coiflets family (cf6) to generate the MODWT coefficients, and hence the MODWT variances and correlations.

Table 1 on Page 11 shows the maximum allowable number of scales for each of the filters for series of lengths $T = 2^{12}$ as well as the number of features associated with the three leads that are initially input into the discriminant procedure. Using the stepwise implementation of [15], one hundred simulations were carried out each time and an average overall correct classification rate was determined.

Figure 4 on Page 16 and Figure 5 on Page 16 show the average overall classification rates over 100 simulations for the synthetic ECG signals using linear and quadratic discriminant analyses, respectively, with the wavelet variances (var), wavelet variances

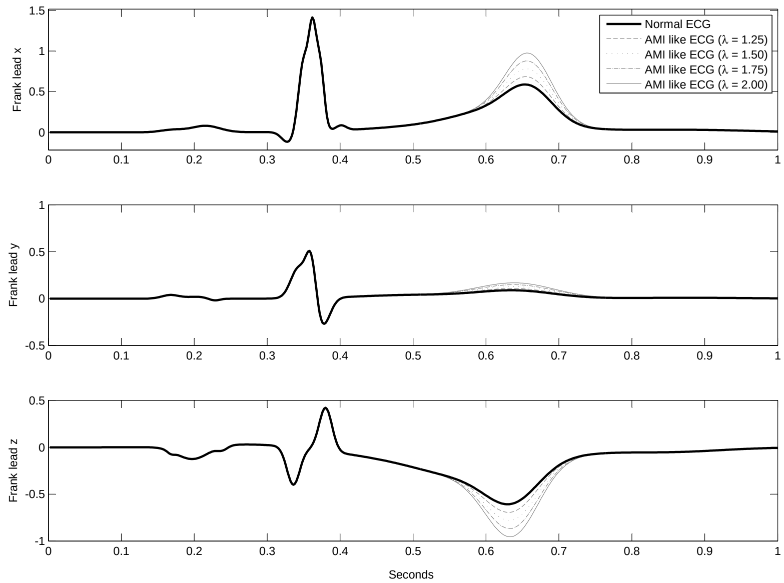


Figure 3: Single beat of a synthetic ECG signals: Normal and Acute Myocardial Infarction.

and wavelet correlations (varcor) and wavelet correlations (corr). We observe that the best results were obtained for λ = 1.25 and λ= 1.5 by using the wavelet variances and correlations together whereas for λ = 1.75 and λ= 2.00, the best results were obtained either when only the wavelet variances were used or when both the wavelet variances and correlations were used. For all values of λ when only wavelet correlations were the input variables, the classification rates were much lower. However, wavelet correlations provide useful information to discriminate since the combined results are sometimes better than results using only wavelet variances. It is clear the there is very little difference in the results between the different wavelet filters and between the linear and quadratic methods.

Figure 6 on page 17 compares the overall classification rates using the Kullback-Leibler (KL) discrimination information and the Chernoff(CH) information measures with the average overall classification rates across the six wavelet filters. Note that on the horizontal axis, L1, L2, L3 refer to results from the linear discriminant using the

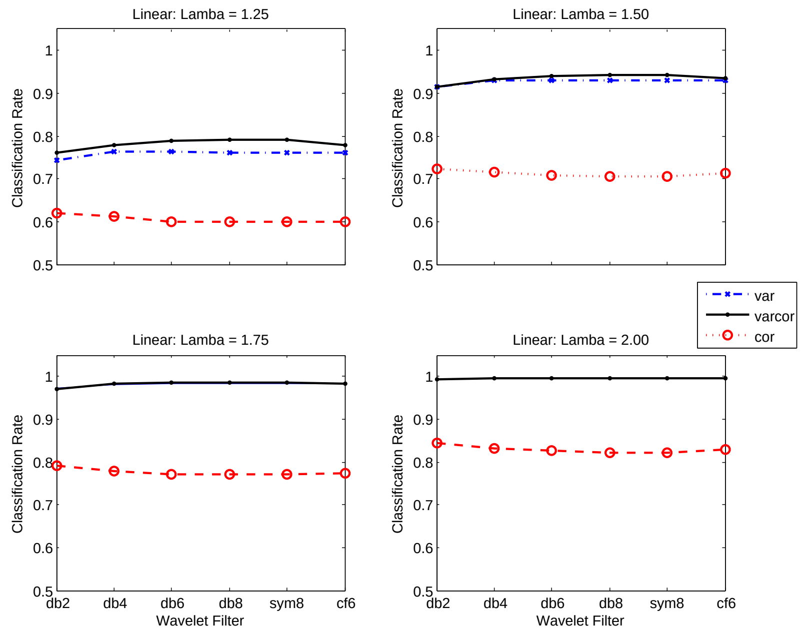


Figure 4: Classification rates for synthetic ECGs using the linear discriminant functions

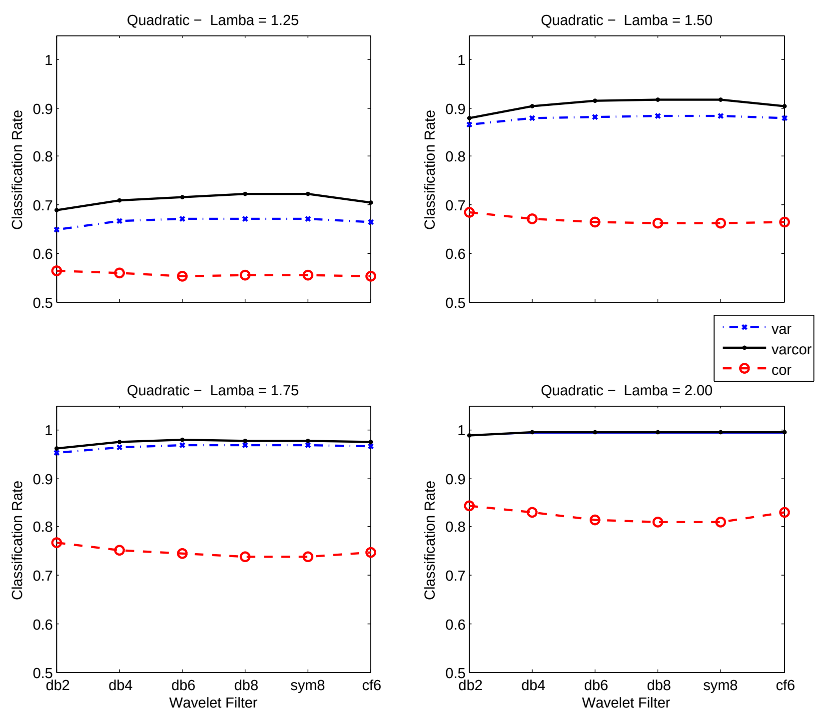
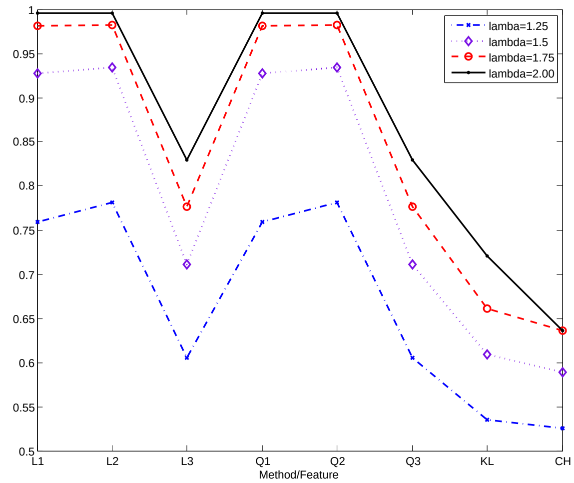


Figure 5: Classification rates for synthetic ECGs using the quadratic discriminant functions

Figure 6: Wavelet-based average classification rates and Kullback-Leibler and Chernoff classification rates for synthetic ECGs



features, wavelet variances, combined wavelet variance and correlations and wavelet correlations, respectively. Likewise the results pertaining to the quadratic discriminant are labeled accordingly. It can be observed that all wavelet based discriminant procedures outperform the Kullback-Leibler information and Chernoff information procedures. For the Kullback-Leibler information and Chernoff information procedures, we considered different values for the bandwidth used to estimate the spectral densities. The considered bandwidths were in the range [0.001, 0.01] that corresponds to from 5 to 41 contiguous fundamental frequencies that are close to the frequency of interest (see [26] p. 197 for more details.

In Figure 7 on Page 18, we present the boxplots of the misclassification rate estimates for $\lambda = 1.25$ and 1.5 which are the most challenging scenarios. We present the results for linear and quadratic methods using the wavelet variances and correlations together. The figure illustrates the stable behavior of the proposed methods across the different wavelet filters. Also, the linear procedures appear to be preferable to the quadratic procedures. The Kullback-Leibler information and Chernoff information procedures are outperformed by wavelet based procedures. As expected, Figures 4 to 6 and Figure 7 show that the performance of all methods improve when λ increases.

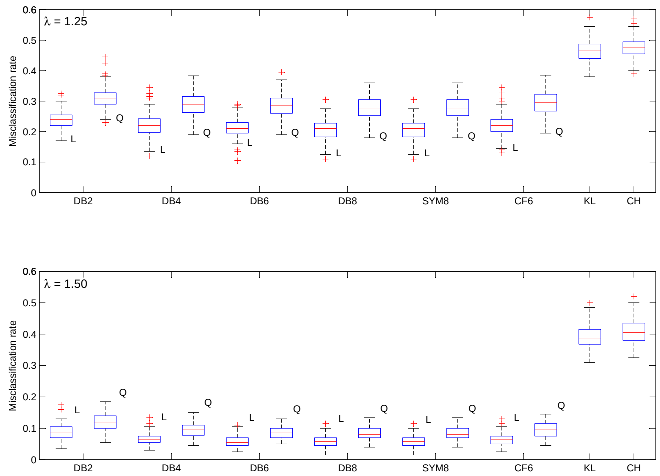


Figure 7: Boxplots with the misclassification rates of the simulation with synthetic ECG signals.

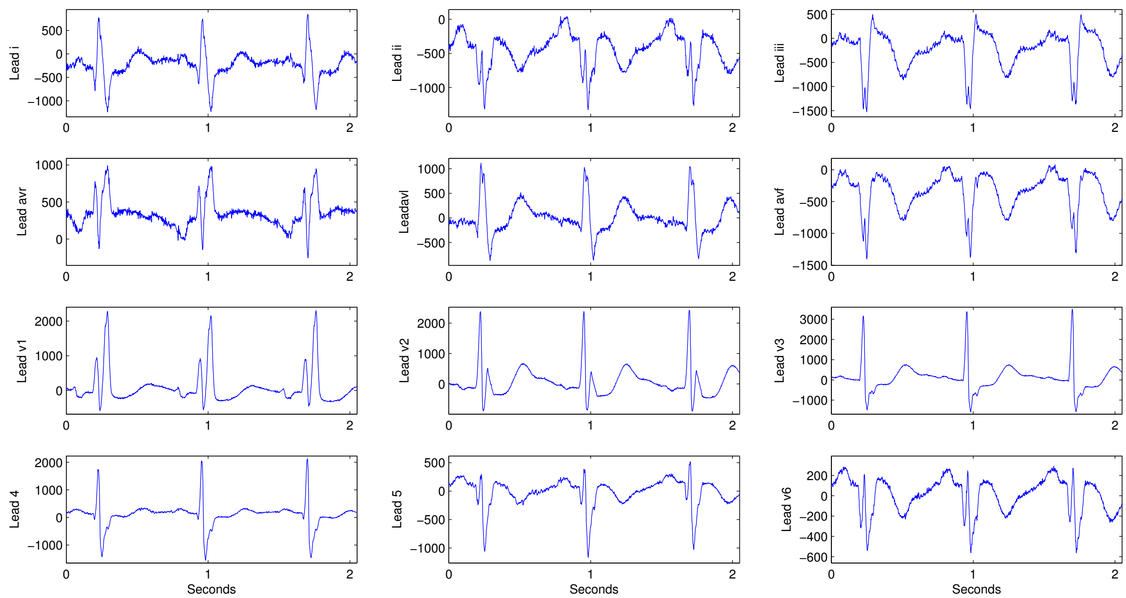
4. Discriminant Analysis of ECG data

The ECG data analyzed comes from the PTB Diagnostic ECG Database [6]. This freely available data is a small subset of the database used by *cardioPATTERN - Telemedical ECG-Evaluation and Follow up* [27] that have a patented procedure for ECG classification.

In this application, we are interested in distinguishing between ECG signals of individuals with myocardial infarction and those of healthy controls. The available dataset consists of 200 records of the conventional 12 leads (i, ii, iii, avr, avl, avf, v1, v2, v3, v4, v5, v6) for 148 patients with myocardial infarction and 52 healthy controls.

Relevant information from the PTB Diagnostic ECG Database [6] and in [7] about how the ECG data in this collection were obtained is as follow: Each signal is digitized at 1000 samples per second (sampling frequency of 1kHz)), with 16 bit resolution over a range of 16.384 mV. The signal bandwidth is 0Hz - 1kHz. and the noise level is < 0.3 μV/$\sqrt{\text{Hz}}$. The anti-aliasing filter used to counteract information loss is an 8-th order 1kHz Bessel Filter of 3dB frequency. Note that mV represents a milli-volt and μV a micro-volt; a volt is the unit for electric potential (voltage); Hz represents a hertz which

Figure 8: ECG of a patient with myocardial infarction over 2 seconds.

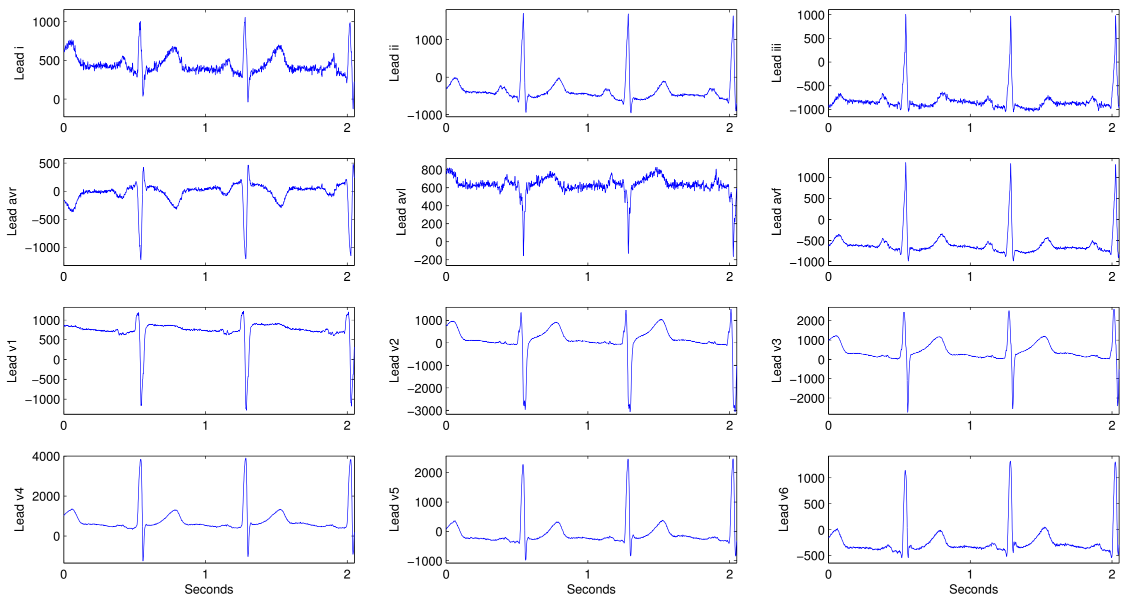


is the inverse of a second, kHz represents a kilohertz and dB represents a decibel which is a logarithmic unit that indicates the ratio of a physical quantity (power of intensity) relative to a specified or implied reference level.

For each record, we read the first $2^{12} + 2^{13}$ observations. We discard the first 2^{12} observations since they could have some exogenous anomalies. The remaining $2^{13} = 8192$ observations correspond to around eleven heart beats. In this case the number of scales $J = 13$ and the maximum allowable number of scales when the different wavelet filters are used is determined from Table 1 on Page 11. Figure 8 on Page 19 and Figure 9 on Page 20 present the 12-lead ECG signals over a two second period (2048 observations) of a subject with myocardial infarction and that of a healthy control, respectively.

In Section 1, we referred to the fact that the shapes of the QRS complex and P- and T-waves as well as the time of occurrence of the QRS complex provide some information of the current state of the heart. Examination of these two second signals reveal that

Figure 9: ECG of a healthy control over 2 seconds.



for all leads the time of occurrence of the QRS complex as well as the P- and T-waves in the signals of a subject with myocardial infarction is about 250 milli-seconds before that in the signals of a healthy control. We now describe the shapes of the QRS complex and the P- and T-waves in the ECG signals of a subject with myocardial infarction compared to those in the ECG signals of a healthy control.

- Lead i: Wider QRS, slightly flatter T-wave.

- Lead ii: More peaked P- and T-waves, inverted QRS compared to the upright QRS in the healthy control signal.

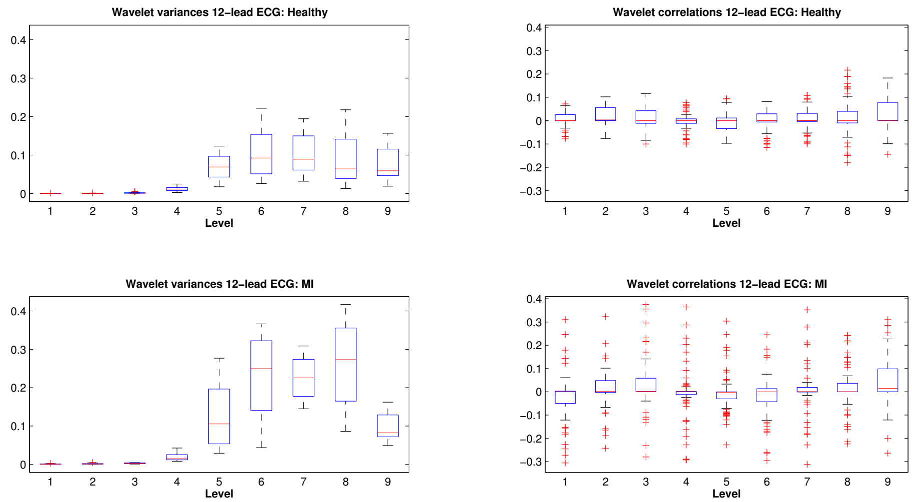
- Lead iii: QRS is noticeably inverted and slightly wider compared to the upright QRS in the healthy control signal; the T-wave is a valley compared to the healthy control signal which has a slight peak.

- Lead avr: Deeper P-wave, QRS is wider with a higher R-wave and a deeper S-wave

- Lead avl: QRS is upright and wider compared to that of the healthy control signal which is inverted; T-wave in more peaked.
- Lead avf: P-wave more peaked, QRS inverted and wider compared to that of the healthy control signal which is upright.
- Lead v1: P-wave distinctly peaked compare to a flattened P-wave in the healthy control signal, QRS wider and upright compared to that of the healthy control signal which is inverted.
- Lead v2: QRS upright compared to that of the healthy control signal which is inverted; T-wave more peaked.
- Lead v3: No apparent difference in shapes in any of the waves compared to those of the healthy control signal.
- Lead v4: No apparent difference in shapes in any of the waves compared to those of the healthy control signal.
- Lead v5: QRS inverted and wider compared to that of the healthy control signal which is upright; T-wave flatter.
- Lead v6: QRS inverted and wider compared to that of the healthy control signal which is upright; T-wave is more peaked.

So, clearly there are differences in the patterns of the ECG signals of subjects with myocardial infarction and healthy controls. Figure 10 on Page 22 shows boxplots of wavelet variances and correlations generated using the sym8 filter of the corresponding 12-lead ECG signals from Figure 8 and Figure 9 at each of 9 scales. We make the following observations: there is a distinct difference in the spread of the wavelet variances of the 12 leads from Scales 5 to 9 between the myocardial infarction subject and the healthy control; there are a larger number of outlying wavelet correlation values for the myocardial infarction subject than for the healthy control. We will now apply our

Figure 10: Boxplots of selected wavelet variances and correlations of the 12-lead ECGs.

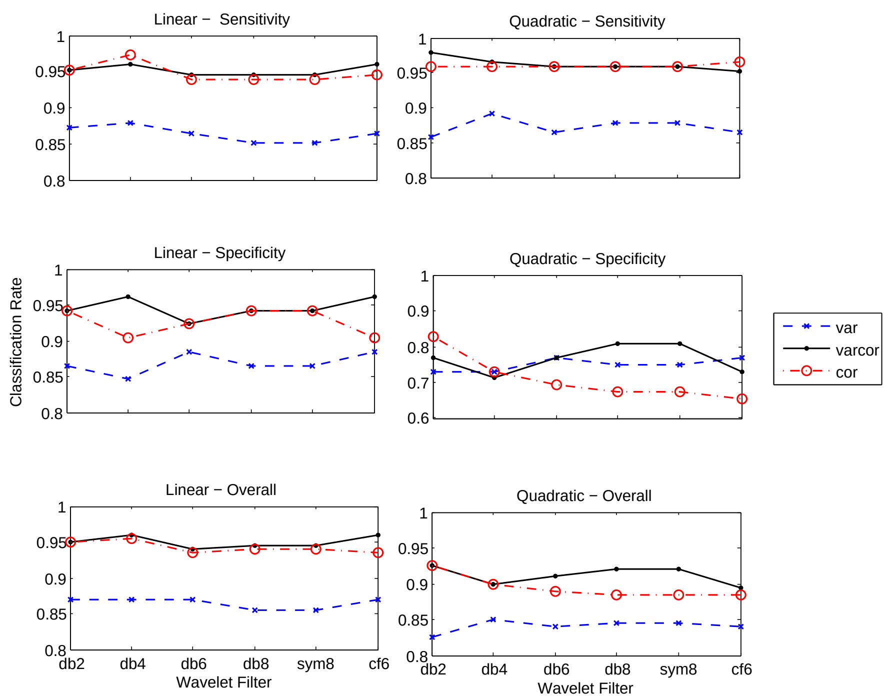


proposed procedure to test how successful it is in differentiating between the ECG patterns of the subjects with myocardial infarction and the healthy controls.

Figure 11 on Page 23 shows the classification rates for patients with myocardial infarction, for healthy controls and the overall classification rates using linear and quadratic discriminant analyses, with the wavelet variances (var), wavelet variances and wavelet correlations (var-cor) and wavelet correlations (cor). The results are shown using each of the six wavelet filters discussed in Section 2.3.

In most cases, the best results were obtained when the wavelet variances and correlations were both the input variables. When only wavelet variances were the input variables, the classification rates were much lower. When only wavelet correlations were the input variables, in some cases the misclassification rates were similar to, or sometimes slightly larger than when both wavelet variance and correlations were input together. It is clear that the wavelet correlations provide useful information about the relationships between the leads of each ECG signal and hence make an important contribution to distinguishing between the patterns of ECG signals of individuals with myocardial infarction, and those of healthy controls. These classification results imply sensitivities of 95-96% (95-98%) and specificities of 92-96% (73-81%) when using the linear (quadratic)

Figure 11: Wavelet-based classification rates for 12-lead ECGs



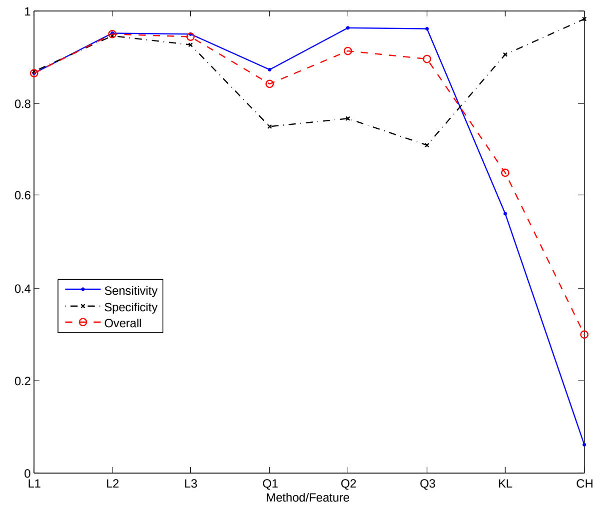
procedure with wavelets variances and correlations together.

Regarding the overall classification rates, the linear procedure generally outperforms the quadratic procedure. Only in few cases did the quadratic procedure outperform the linear procedure when classifying myocardial infarction ECGs. In general, for the linear procedure, the error rates were fairly similar for the different wavelet filters. This was also the case for the quadratic procedure.

Figure 12 on Page 24, compares the classification rates using the Kullback-Leibler (KL) discrimination information and the Chernoff(CH) information measures with the average classification rates across the six wavelet filters. Note that on the horizontal axis, L1, L2, L3 refer to results from the linear discriminant using the features, wavelet variances, combined wavelet variance and correlations and wavelet correlations, respectively. Likewise the results pertaining to the quadratic discriminant are labeled accordingly.

For the Kullback-Leibler information and Chernoff information procedures, we consider different values for the bandwidth used to estimate the spectral densities. The

Figure 12: Wavelet-based average classification rates and Kullback-Leibler and Chernoff classification rates for 12-lead ECGs



considered bandwidths were in the range [0.001, 0.01] that corresponds to from 9 to 81 contiguous fundamental frequencies that are close to the frequency of interest (see [26] p. 197 for more details).

The procedures based on the Kullback-Leibler information and Chernoff information have good specificities (90.4 and 98.1, respectively) but very poor sensitivities (56.1 and 6.1, respectively), thus resulting in the poor overall classification rates. Hence on average, in terms of overall classifications, all wavelet-based discriminant procedures outperform the Kullback-Leibler information and Chernoff information procedures.

It can be observed from Table 1 on Page 11 that in applying discriminant analysis to the 12-lead ECG signals, the number of input features varies between 990 and 1057 depending on the type of wavelet filter used. This could be seen as a limitation to applying our procedure since there are many more discriminating variables than there are cases. However, this has been overcome by implementing stepwise discriminant analysis. Furthermore, this not have a great impact on the time taken to run the program since the classification of a new signal takes, using a wavelet- based procedure; less than one second (using a personal computer with a Intel(R) Core(TM) i7 CPU 920 @ 267GHz).

In order to test how sensitive our proposed procedure is under different conditions, we conducted several experiments. These are described in the Subsections 4.1 to 4.5.

4.1. Preprocessed ECG signals

It should be noted that our results have been obtained by applying the procedures to the ECG signals as they are available on the PTB database, i.e., no additional preprocessing was applied. In the ECG literature, one of the recommended methods for preprocessing signals is to denoise the signals (e.g., refer to [30]) . In order to test how sensitive our analysis is to denoising, we applied the Savitzky-Golay method of filtering to denoise our data. This method involves fitting a p-th order polynomial to a fixed frame of consecutive points. Refer to [31] for more details. We used frame sizes of 11, 21, and 31 points and we fitted 0-th, 3-rd and 5-th order polynomials to each frame size, i.e., D1 = (0,11); D2 = (0,21); D3 = (0,31); D4 = (3,11); D5 = (3,21); D6 = (3,31); D7 = (5,11); D8 = (5,21) and D9 = (5,31). A signal is reduced to being noiseless when the 0-th polynomial is fitted. Figures 13 and 14 on Page 26 show single ECG signals of healthy control and that of an individual with myocardial infarction, respectively, with three denoised versions of each, viz., D1, D4 and D7. It is clear that with these combinations of frame size and polynomial order, the P, Q, R, S and T wave patterns are retained. The parts of the signals between these waves are smoothed out, with the degree of smoothing increasing slightly with the increasing order of the polynomial. The selection of the frame size, does not appear to have much impact on the degree of smoothing.

The first frame in Figures 15 to 17 on Pages 27 to 28 show the average classification rates from across the six wavelet filters and the classification rates using the Kullback-Leibler (KL) and the Chernoff(CH) information measures, with the remaining three frames showing the classification rates for the corresponding denoised versions. Note that as before on the horizontal axis, L1, L2, L3 refer to results from the linear discriminant using the features, wavelet variances, combined wavelet variance and correlations and wavelet correlation, respectively. Likewise the results pertaining to the quadratic

Figure 13: Single Healthy Control ECG and denoised, D1, D4 and D7 ECGs

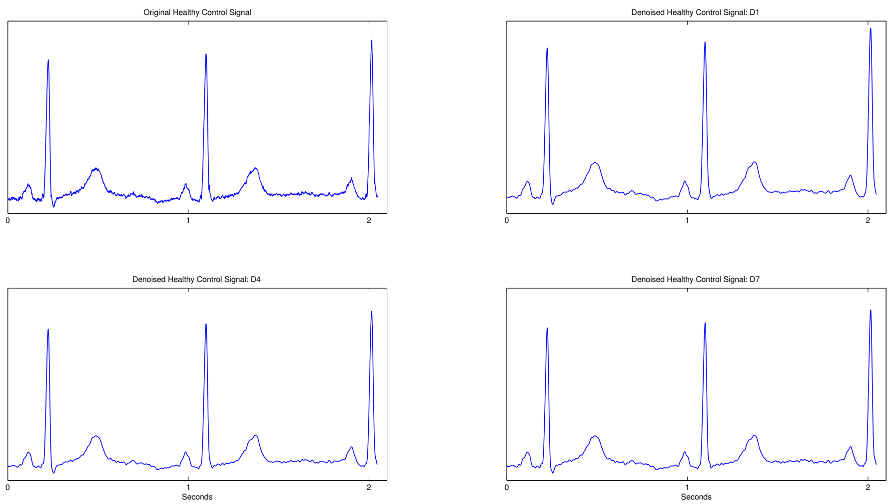


Figure 14: Single Myocardial Infarction ECG and denoised, D1, D4 and D7 ECGs

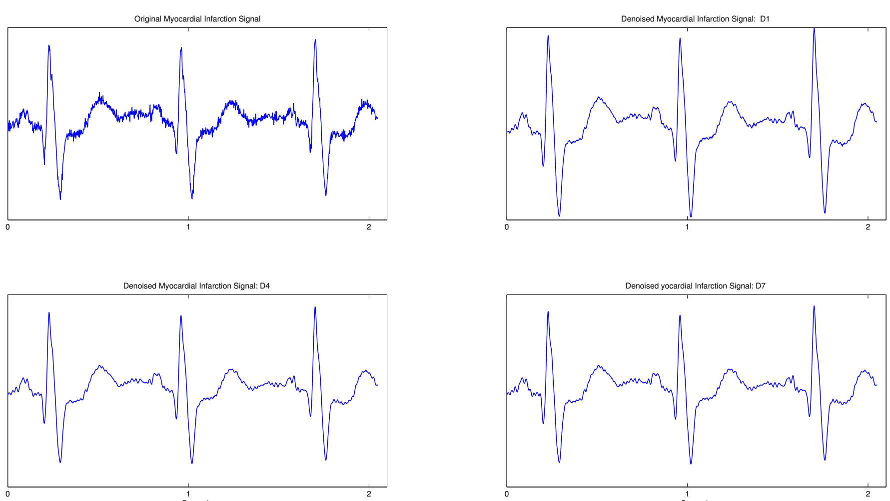


Figure 15: Classification rates for 12-lead originals ECGs and denoised, D1, D2 and D3 ECGs

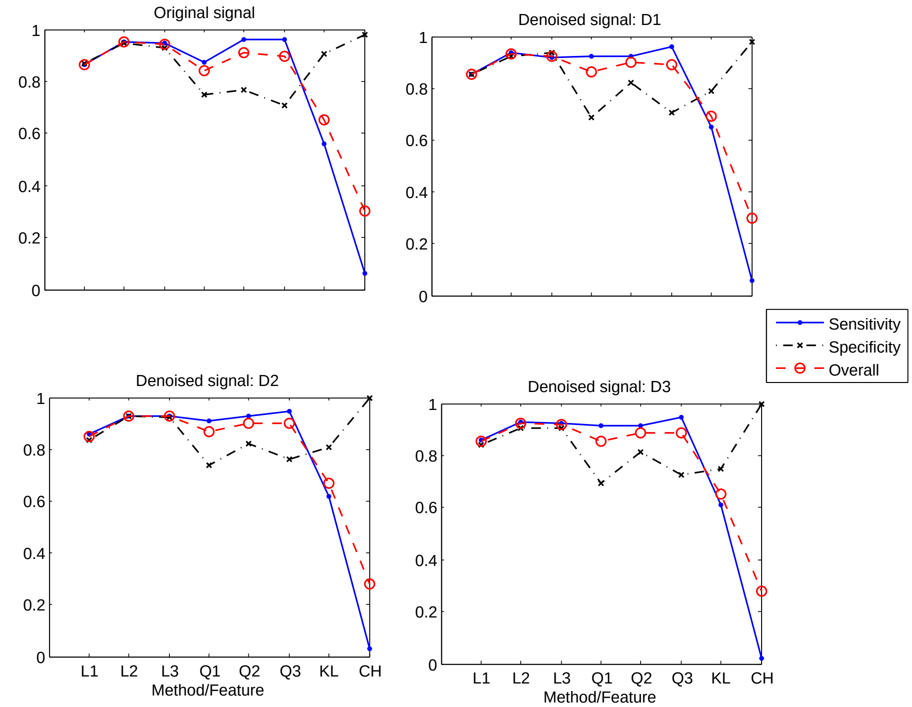


Figure 16: Classification rates for 12-lead originals ECGs and denoised, D4, D5 and D6 ECGs

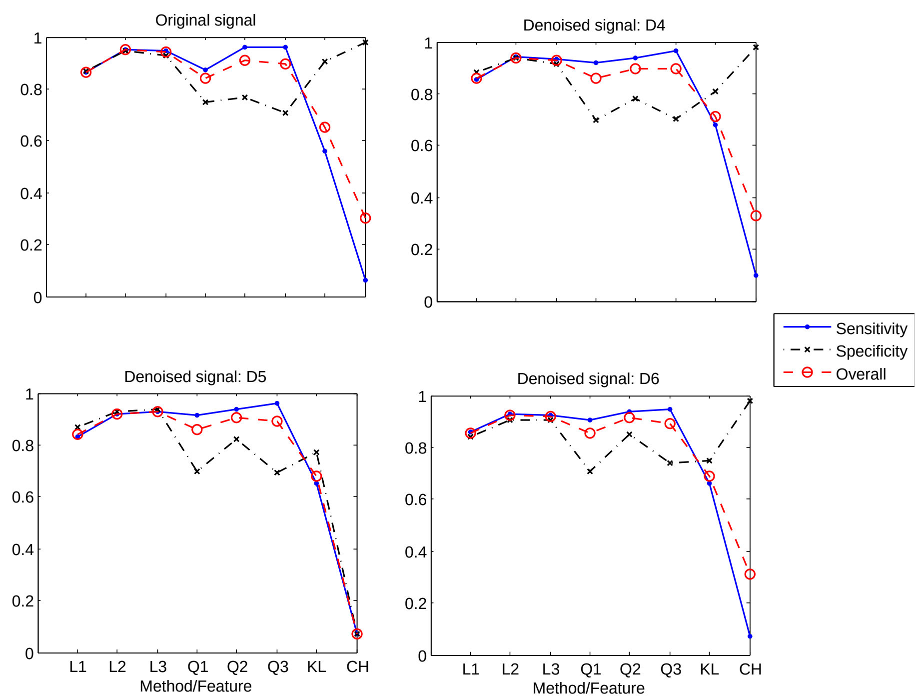
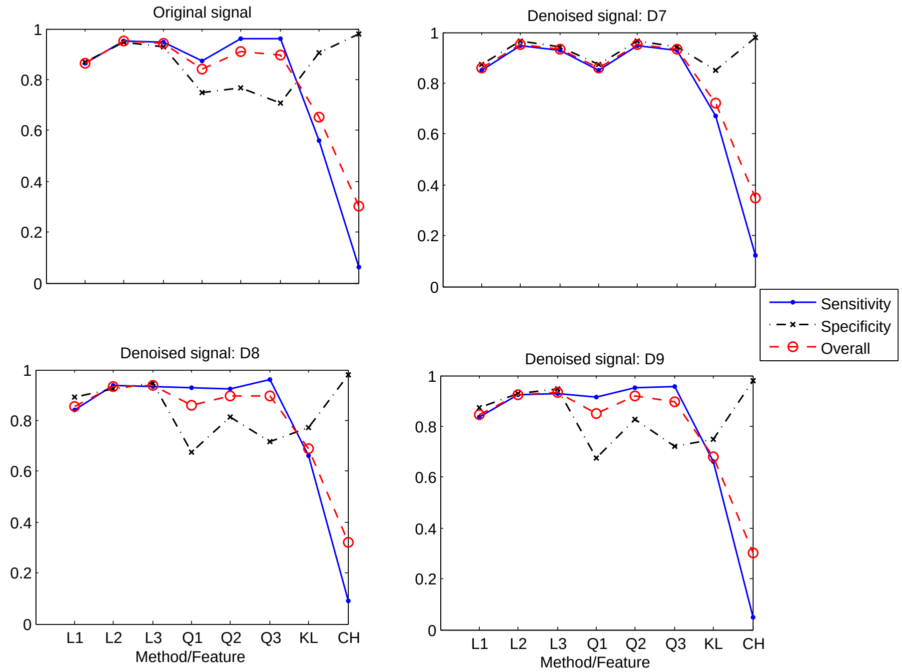


Figure 17: Classification rates 12-lead originals ECGs and denoised, D7, D8 and D9 ECGs



discriminant are labeled accordingly. It is clear that for the linear discriminant procedure, and the Kullback-Leibler information and Chernoff information procedures, denoising the ECG signals has had little or no impact on the classification results. In the case of the quadratic procedure, in some cases there are differences, more so when the ECG signals were denoised with a 5-th order polynomial over 11 points (D7).

4.2. Varying the sampling frequency

As mentioned previously, the sampling frequency of the ECG signals used in the analysis thus far is 1kHz. To compare how sensitive our analysis is to different sampling frequencies, we down sampled the existing ECG signal to obtain signals at sampling frequencies of 0.5kHz and 0.25kHz. We take one observation out of two (four) consecutive observations in order to get 0.5kHz (0.25kHz). The time series lengths are 2^{13}, 2^{12} and 2^{11} for 1kHz, 0.5kHz and 0.25kHz, respectively. But notice, that we are using the same period of time, i.e., we are observing the same beats with the three sampling frequencies.

Figure 18: Overall classification rates 12-lead ECGs of different sampling frequencies

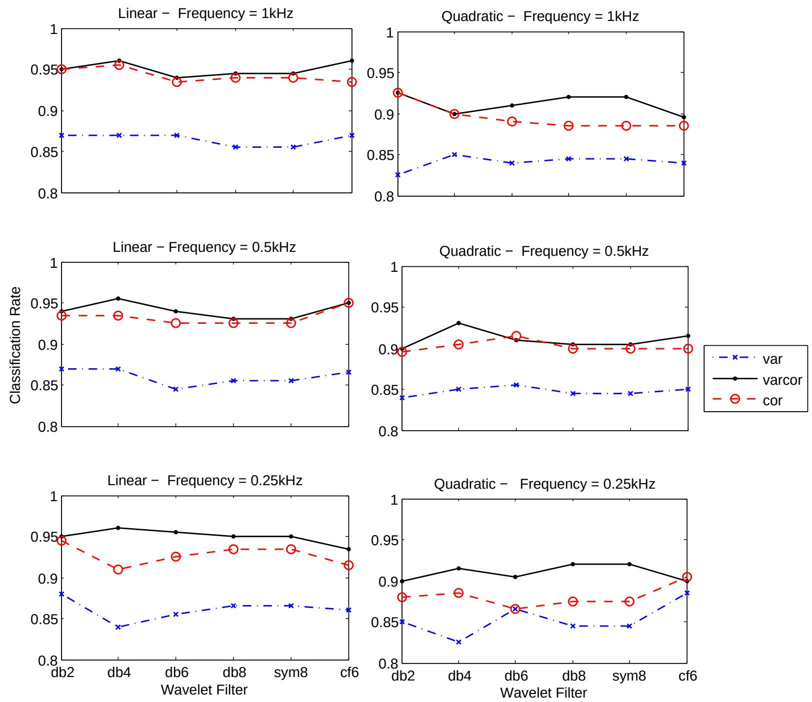
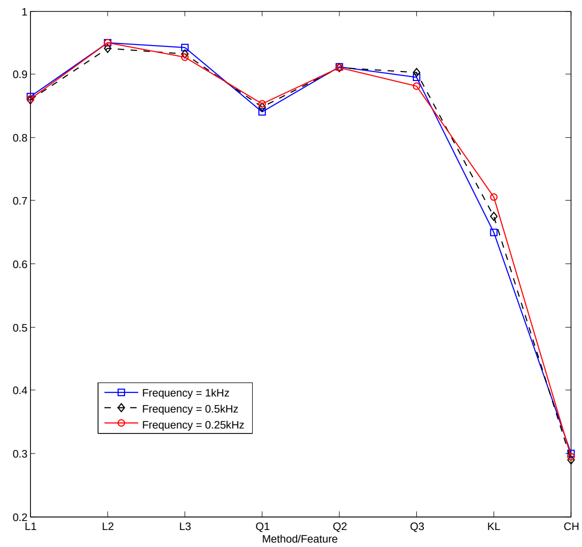


Figure 18 on Page 29 shows the overall classification rates for using linear and quadratic discriminant analysis with the wavelet features using each of the 6 wavelet filters for the different sampling frequencies. We observe that across the different frequencies, there is less variation in the classification rates for the linear procedure than for the quadratic procedure, implying that the linear process is less sensitive to different sampling frequencies than the quadratic process.

Figure 19 on Page 30 shows the average overall classification rates from across the six wavelet filters and the classification rates using the Kullback-Leibler (KL) and the Chernoff(CH) information measures. Note that as before on the horizontal axis, L1, L2, L3 refer to results from the linear discriminant using the features, wavelet variances, combined wavelet variance and correlations and wavelet correlation, respectively for the different sampling frequencies. Likewise the results pertaining to the quadratic discriminant are labeled accordingly. These results imply on average the all of these methods the classification results do not appear to be very sensitive to sampling frequency. Additionally, these results imply that our procedures can handle ECG signals observed at

Figure 19: Wavelet-based average overall classification rates and Kullback-Leibler and Chernoff classification rates for 12-lead ECGs of different sampling frequencies



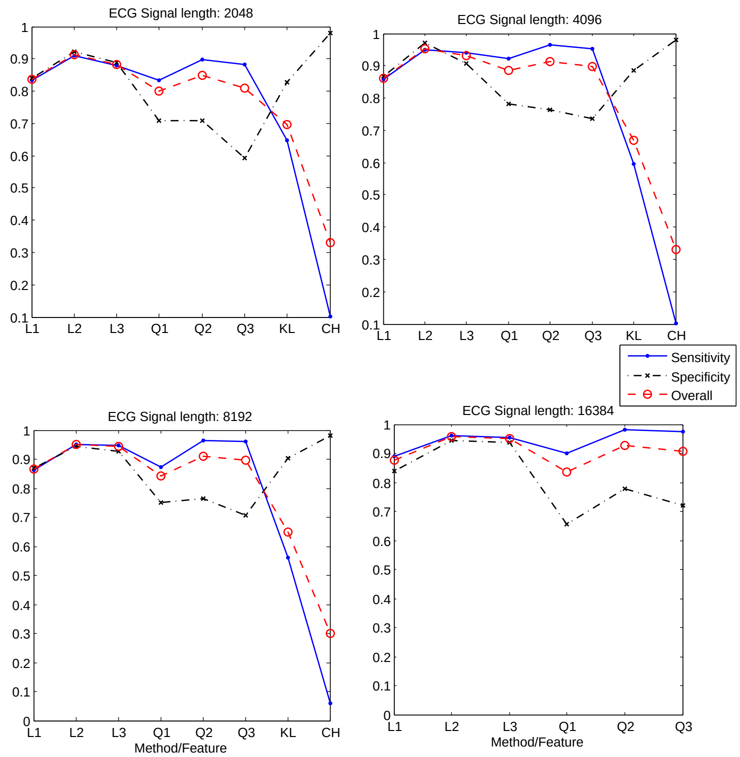
a lower sampling frequencies with similar overall classification rates.

4.3. Varying the signal lengths

In order to evaluate the "robustness" of our procedures with others signals lengths, we considered two smaller lengths, 2^{11} and 2^{12}, and one bigger length, 2^{14}. Notice that, in this case, the sampling frequency is the same for all lengths. So, the periods of observation are around two, four, eight and 16 seconds when we use 2^{11}, 2^{12}, 2^{13} and 2^{14} observations, respectively.

Figure 20 on Page 31 shows the average classification rates across the six wavelet filters for the ECG signals of lengths 2^{11}, 2^{12}, 2^{13} and 2^{14} with the classification rates using the Kullback-Leibler (KL) and the Chernoff(CH) information measures for lengths 2^{11}, 2^{12} and 2^{13}. We were unable to generate results Kullback-Leibler and the Chernoff information measures with 2^{14} observations because of memory constraints in the MATLAB software. Note that as before on the horizontal axis, L1, L2, L3 refer to results from the linear discriminant using the features, wavelet variances, combined wavelet variance and correlations and wavelet correlation, respectively, for the different signal lengths. Likewise the results pertaining to the quadratic discriminant are labeled accordingly.

Figure 20: Wavelet-based average classification rates and Kullback-Leibler and Chernoff classification rates for 12-lead ECGs of different lengths



These results imply that for the signal lengths 2^{12}, 2^{13} and 2^{14} there is little or no difference in the average classification when either the linear or quadratic procedure is used. However the percentages of correct classification are smaller for length 2^{11}. Notice that, 2^{11} observations corresponds to around two seconds of recorded signals, i.e., two or three beats which is a minimal amount of information. Likewise for there is little or no difference in the results when the Kullback-Leibler and the Chernoff information measures were used for 2^{11}, 2^{12} and 2^{13}. Given these consistently good results when the wavelet features are used with the linear discriminant function, we could therefore argue that our proposed method is not very sensitive to the variation in signal lengths, provided, the signal is recorded over a period of more than two seconds.

4.4. Using 8-lead ECG signals

Since there are only two independent leads among six limb and augmented limb leads because Lead iii, aVR, aVL and aVF can be obtained from linear combination of Lead i and Lead ii, we tested eight leads signal (Lead i, ii and V1-V6) to determine if the reduction the computational time does not result in loss of sensitivity and specificity?

Figure 21 on Page 33 shows the overall classification rates for using linear and quadratic discriminant analysis with the wavelet features using each of the six wavelet filters for the 8-lead signals. Figure 22 on Page 33 shows the average overall classification rates from across the six wavelet filters and the classification rates using the Kullback-Leibler (KL) and the Chernoff(CH) information measures for the 8-lead ECG signals. Note that as before on the horizontal axis, L1, L2, L3 refer to results from the linear discriminant using the features, wavelet variances, combined wavelet variance and correlations and wavelet correlation, respectively. Likewise the results pertaining to the quadratic discriminant are labeled accordingly.

These classification rates can be compared with those presented in Figure 11 on Page 23 and Figure 12 on Page 24 for the 12-lead signals. It is clear that the 12-leads based results are generally better than the 8-leads based results. This means that some features associated to leads lead iii, aVR, aVL and aVF have been used as discriminant variables. The linear combination could amplified the variance/correlation differences among the groups with respect to the variance/correlation differences among the groups using only the leads i and ii. For instance, if we use lev=0, which means that the original series is used. When we perform the discriminant analysis using leads i, ii, iii, aVR, aVL and aVF, we have as possible explanatory variables: var(i), var(ii), var(iii)=var(ii-i), var(aVR)=var(-0.5(i+ii)), var(aVL)=var(i-0.5ii), and var(aVF)=var(ii-0.5i) but when we perform the discriminant analysis using only lead i and ii then we have as explanatory variables var(i) and var(ii). Since leads i and ii are not independent, both approaches are not equivalent.

Figure 21: Classification rates for 8-lead ECGs

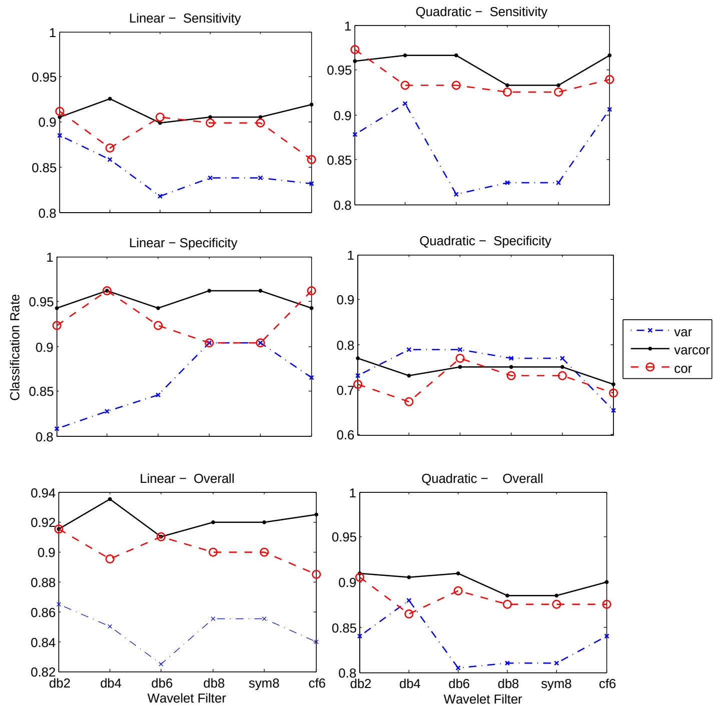
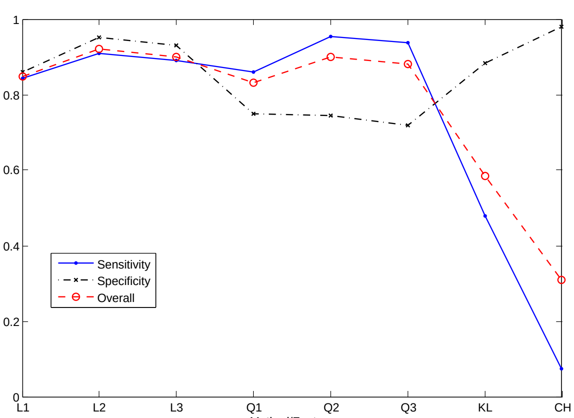


Figure 22: Wavelet-based average classification rates and Kullback-Leibler and Chernoff classification rates for 8-lead ECGs.



4.5. Measuring the repeatability of the classification

The minimum length of the recorded electrocardiograms in the PTB database is 32000 observations. As mentioned at the beginning of this section, we discard the first 2^{12} observations and the following 2^{13} observations were used to develop and test the discriminant procedures. Therefore, we have additional data, at least $32000 - 2^{12} - 2^{13}$ observations, that can be used to evaluate the proposed procedures. It is important to notice that these extra segments of 2^{13} observation were not used in the discriminant procedure's design. So, the classification of these "new" data can be considered as a measure of the repeatability of our previous results. Of course, these "new" data are not independent from the data used to develop the discriminant procedure.

Figure 23 on Page 35 shows the overall classification rates when using linear and quadratic discriminant analysis with the wavelet features using each of the six wavelet filters for a different section of 2^{13} observations of the 12-lead ECG signals compared to what was used to obtain the results shown in Figure 11 on Page 23. For both the linear quadratic processes, apart for a slight increase in the specificity for some of the wavelet filters, the results for this different sections of 2^{13} observations appear to be fairly compatible with those of the original 12-lead ECG signals in Figure 11 on Page 23.

Figure 24 on Page 35 shows the average classification rates from across the six wavelet filters and the classification rates using the Kullback-Leibler (KL) and the Chernoff(CH) information measures for this different section of 2^{13} observations of the 12-lead ECG signals compared to what was used to obtain the results shown in Figure 12 on Page 24. Note that as before on the horizontal axis, L1, L2, L3 refer to results from the linear discriminant using the features, wavelet variances, combined wavelet variance and correlations and wavelet correlation, respectively. Likewise the results pertaining to the quadratic discriminant are labeled accordingly. These results appear to be fairly compatible with those of the original 12-lead ECG signals in Figure 12 on Page 24.

Hence the above assessment of the performance on two different sections of the 12-lead ECG signals, would appear to validate our proposed classification procedure.

Figure 23: Classification rates 12-lead ECGs for a different section of 2^{13} observations.

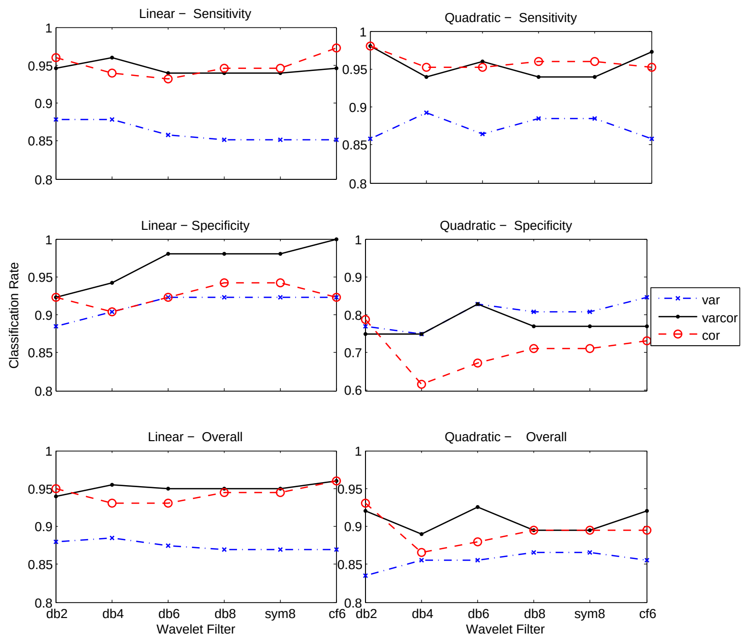
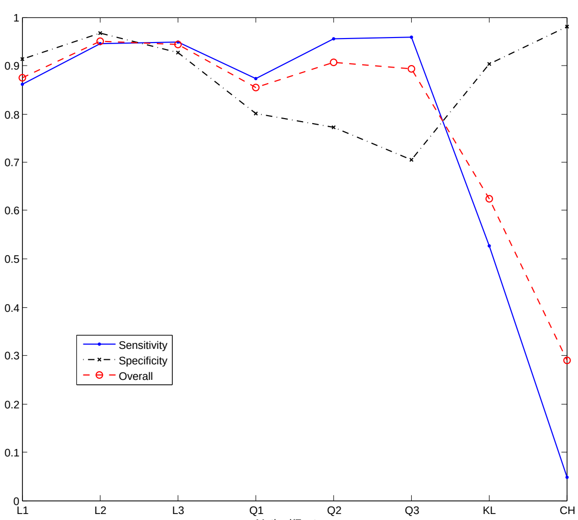


Figure 24: Wavelet-based average classification rates and Kullback-Leibler and Chernoff classification rates for 12-lead ECGs for a different section of 2^{13} observations.



4.6. Comparisons with other author's ECG classification results

Many researchers have analyzed the samples of ECG data for PTB database. In particular, [32], [33], and [34] proposed pattern recognition approaches to classifying the ECG signals between subjects with myocardial infarction and healthy controls from this database.

Banerjee and Mitra [32] developed a algorithm based the Discrete Wavelet Transform (DWT), which extracts features such as the QRS complex and computes the QRS vector of the the 12-lead ECG signals. They applied it to 12-lead signals from 25 subjects with myocardial infarction and 25 healthy controls. They obtained a sensitivity of 100% and a specificity of 92% but they did not use test data or any cross-validation technique to validate the accuracy of their results.

Kopiec and Martyna [33] used a combination of Support Vector Machine (SVM) and Particle Swarm Optimization (PSO) to extract significant features for each of the 12 lead ECG signals from 290 subjects in the database of which 52 were of healthy controls and 148 were myocardial infarction cases and the remaining 90 were of subjects diagnosed with various other heart conditions. Their best test set overall accuracy results for sets of 3 leads, viz., (avl, avr, avf); (i, ii, iii); (v1, v2, v3); (v4, v5, v6); (i, avf, v3) were all around 80%.

Li et al. [34] developed a method referred to as Multiple Instance learning (MIL) which is based on extracted QRS and ST features from the ECG signals and its heartbeats. They applied the methods to 389 ECG records from subjects with myocardial infarction and 79 from healthy controls and obtained a sensitivity of 93% and a specificity of 90%.

Our best results for the 12-lead ECG classification of a 96% sensitivity and a 96% specificity, using the linear discriminant procedure for the 148 records from subjects with myocardial infarction and 52 from healthy controls compare favourably with the results obtained by these other authors who used the ECG signals from the same PTB database, albeit from different samples.

The cardioPATTERN - Telemedical ECG-Evaluation and Follow up procedure [27], using a database of 8500 ECGs from 3781 patients with different cardiac diseases, and 4719 normal persons, obtained a sensitivity of 72% and a specificity of 80% when classifying myocardial infarction. Although this database as well as those used by [2], [?], [4] and [5] are not comparable with the PTB database, we can use their sensitivity and specificity results as reference values. Clearly, the results obtained with our proposed procedure with the 12-lead ECG signals compare quite favourably . Furthermore, given that for most of the time we achieved our best results when both wavelet variances and correlations were input together in the discriminant procedures, it is clearly an indication that the relationships between the components of the multi-lead ECG signals provide useful information for classification.

4.7. Remark

In order to evaluate the performance of the use of discriminant analysis with the wavelet features when applied to multivariate time series with patterns different from that of ECG signals, we conducted two sets of simulation studies in which we used the stepwise implementation of Strauss [15] in linear and quadratic discriminant procedures.

For both studies, the best results were generally achieved when both wavelet variances and wavelet correlations were input together as the discriminating variables. As was also the case in the studies in Sections 3 and 4, when only wavelet variances were the input variables, the overall misclassification rates were generally higher. It is therefore clear that wavelet correlations provide useful information about the interrelationship between the components of each multivariate time series for the purposes of classification when combined with the wavelet correlations. Refer to [29] for the design and results of these studies.

5. Concluding Remarks

We proposed a procedure of classifying multivariate ECG signals based on conventional discriminant analysis with wavelet features, namely, wavelet variances and wavelet correlations as the discriminating variables. We demonstrated in applications to synthetic ECG data and to real ECG data and that when both wavelet variances and correlations were used as the discriminating variables, this proposed procedure generally produces better results in most cases than when only wavelet variances were used or when only wavelet correlations were used. We also demonstrated that overall this procedure outperforms other multivariate discriminant procedures proposed by Kakizawa et al. [11]. Furthermore we demonstrated through a number of experiments that this proposed procedure is not sensitive to preprocessing, different sampling frequencies and different signal lengths, provided that signal lengths are more than two seconds. We also validated the procedure by demonstrating the repeatability of the results with different ECG signal segments.

Using the results of other ECG data classifying methods as reference values, it is clear that our proposed procedure displays favourable performance. Hence, the main contributions of this paper are (i.) the inclusion of the interrelationship features between components of the ECG signals in the form of wavelet correlations provides useful information for the purposes of classifying myocardial infarction; (ii.) the generally favourable performance of our approach when compared to the other well-known methods for classifying multivariate time series.

Acknowledgements

The first author began working on this paper while she was on sabbatical leave at the Universidad Carlos III de Madrid. She acknowledges the support provided by Monash University and the Universidad Carlos III de Madrid during this period.

We thank Professor David Stoffer for his guidance on the implementation of the routines for Kullback-Leibler and Chernoff-based procedures and Professors Gari Clifford and Reza Sameni for their guidance on the use of ECG synthetic generator.

We thank the referees for their comments and suggestions which in many ways helped improve the quality and presentation of the manuscript.

All results for the application and simulations were generated using MATLAB software. To obtain the wavelet variances and correlations, we used the WMTSA Wavelet Toolkit developed by Cornish [28]. To obtain ECG synthetic signals, we used the Open-

Source Electrophysiological Toolbox developed by Sameni and co-authors [23], [25] and [24].

Supplementary Material

Our suite of routines used in this paper is available online at:

http://www.est.uc3m.es/amalonso/esp/addenda.html.

References

[1] LabVIEW for ECG Signal Processing (2012), National Instruments. `http://www.ni.com/white-paper/6349/en`.

[2] Al-Naima, F. and Al-Timemy, A. (2009), Neural network based classification of myocardial infraction: A comparative study of wavelet and Fourier transforms in Pattern Recognition, Y. Peng-Yeng (Ed.). *InTech.pp.338-352*.

[3] De Chazal, P. and Reilly, R.B. (2000), A comparison of the ECG classification performance of different feature sets. *Computers in Cardiology* **27**, 327–330.

[4] Heden, B., Ohlin, H., Rittner, R. and Edenbrandt, L. (1997), Acute myocardial infarction detected in 12-lead ECG by artificial neural networks. *Circulation: American Heart Association Inc.* **96**, 1798–1802.

[5] Bozzola, G., Bortolan, G., Combi, C., Pinciroli, F. and Brohet, C. (1996), A hybrid neuro-fuzzy system for ECG classification of myocardial infraction. *Computers in Cardiology* , 241–244.

[6] Goldberger, AL, Amaral, LAN, Glass, L., Hausdorff, J.M., Ivanov, PCh., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C-K., Stanley, H.E. (2000) PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* , 101(23):e215–e220.

Circulation Electronic Pages:
http://circc.ahajournals.org/cgi/content/full/101/23/e215.
The PTB Diagnostic ECG Database:
http://www.physionet.org/physiobank/database/ptbd.

[7] Bousseljot, R, Kreiseler, D, Schnabel, A. (1995) Nutzung der EKG-Signaldatenbank CARDIODAT der PTB ber das Internet. *Biomedizinische Technik* **40 (S1)**, S317.

[8] Lachenbruch, P.A. and Mickey, M.R. (1968), Estimation of error rates in discriminant analysis.. *Technometrics* **10**, 1–10.

[9] Huang H., Hernando O., Stoffer D.S. 2004. Discrimination and Classification of Nonstationary Time Series using the SLEX Model . *J.Amer. Statist. Assoc.* **99 (467)**, 763–774.

[10] Maharaj, E.A., Alonso, A.M. (2007), Discrimination of locally stationary time series using wavelets. *Comput. Statist.Data Anal.* **52** , 879–889.

[11] Kakizawa Y., Shumway R.H. and Taniguchi M. (1998), Discrimination and clustering for multivariate time series . *J. Amer. Statist. Assoc* **93 (441)**, 324–340.

[12] Shumway, R.H. (2003), Time-frequency clustering and discriminant analysis . *Statist. Probab. Let.* **63 (3)**, 307–314.

[13] Percival, D.B. and Walden, A.T. (2000), Wavelet Methods for Time Series Analysis. *Cambridge University Press, Cambridge.*

[14] D'Urso, P., Maharaj, E.A. (2012), Wavelets-Based Clustering of Multivariate Time Series. *Fuzzy Sets and Systems* **193** , 33–61.

[15] Strauss, R.E. (2004), Statistical Matlab Library.
http://www.faculty.biol.ttu.edu/Strauss/Matlab/matlab.htm.

[16] McLachlan, G.J. (2004), Discriminant Analysis and Statistical Pattern Recognition. *John Wiley and Sons, Inc. Hoboken, New Jersey.*

[17] SERROUKH A., WALDEN A.T., PERCIVAL D.B. (2000), STATISTICAL PROPERTIES AND USES OF THE WAVELET VARIANCE ESTIMATOR FOR THE SCALE ANALYSIS OF TIME SERIES. *J. Amer. Statist. Assoc.* ***95 450*** *, 184–196.*

[18] SERROUKH, A. AND WALDEN, A.T. (2000), Wavelet scale analysis of bivariate time series II: Statistical properties for linear processes . *Nonparametric Statistics* ***13*** *, 37–56.*

[19] CHINGANDA, E.F. AND SUBRAHANIAM, K. (1979), Robustness of the linear discriminant function to nonnormality: Johnson's system. *J. Statist. Planning Inference* **3**, 69–77.

[20] FATTI, L.P., HAWKINS, D.M., RAATH, E.L. (1982), Discriminant Analysis. In Topics in Applied Multivariate Analysis, D.M. Hawkins (Ed.). *Cambridge University Press, Cambridge.pp.1-71.*

[21] THE OPEN-SOURCE ELECTROPHYSIOLOGICAL TOOLBOX (2010) http://www.oset.ir/.

[22] MCSHARRY P. E., CLIFFORD G. D., TARASSENKO L. AND SMITH L. A. (2003), A Dynamic Model for Generating Synthetic Electrocardiogram Signals. *IEEE Trans. Biomed. Eng* ***50****, 289–294.*

[23] SAMENI R., CLIFFORD G. D., JUTTEN C., SHAMSOLLAHI M. B. (2007), Multi-Channel ECG and Noise Modeling: Application to Maternal and Fetal ECG Signals. *IEURASIP Journal on Advances in Signal Processing. Article ID 43407, 14 pages, 2007. doi:10.1155/2007/43407.*

[24] CLIFFORD G., NEMATI, S. AND SAMENI, R. (2010), An Artificial Vector Model for Generating Abnormal Electrocardiographic Rhythms. *JPhysiological Measurements.* ***31 5****, 595–609.*

[25] SAMENI R. (2008), Extraction of Fetal Cardiac Signals from an Array of Maternal Abdominal Recordings," Ph.D. dissertation, Sharif University of Technology -

Institut National Polytechnique de Grenoble, July 2008.
http://www.sameni.info/Publications/Thesis/PhDThesis.pdf.

[26] Shumway R.H. and Stoffer D.S. (2006), Time Series Analysis and Its Applications With R Examples. *Springer, New York.*

[27] cardioPATTERN - Telemedical ECG-Evaluation and Follow up (2009). http://radib.dyndns.org.

[28] Cornish C. (2004), WMTSA wavelet toolkit for MATLAB.
http://staff.washington.edu/dbp/wmtsa.html

[29] Maharaj, E.A., Alonso, A.M. (2012), Discrimination of multivariate time series using wavelets. Working Paper 12-06 Statistics and Econometrics Series (03), February 2012. http://hdl.handle.net/10016/13359.

[30] Maglaveras, N., Stamkopoulos, T., Diamantaras, K., Pappas, C., Strintzis., M. (1998) ECG pattern recognition and classification using non-linear transformations and neural networks: A review. International Journal of Medical Informatics, 52, 191-208.

[31] Orfanidis, S.J.(1996) Introduction to Signal Processing, Prentice-Hall, Englewood Cliffs, NJ, 1996.

[32] Banerjee, S. and Mitra,, M. (2010) ECG feature extraction and classification of anteroseptal ayocardial infarction and normal subjects using discrete wavelet transform. Proceedings of 2101 International Conference on Systems in Medicine and Biology, 16-18 December 2010, IIT, Kharagpur, India.

[33] Kopiec, D. and Martyna J. (2011) A hybrid approach for ECG classification based on Particle Swarm Optimization and Support Vector Machine. E. Corchado, M. Kurzynski, M.Wozniak (Eds): HAIS2011, Part 1, LNAI 6678, 329-337.

[34] Li, S., Lu, Y.L., Yang, K. and Li, S. (2012) ECG Analysis using multiple instance leanring for mycardial infarction detection. IEEE Transactions on Biomedical Engineering, 59, 12, 3348-3356.