CLASSIFICATION OF THE ELECTROCARDIOGRAM USING SELECTED WAVELET COEFFICIENTS AND LINEAR DISCRIMINANTS

P. de Chazal, R. B. Reilly*, G. McDarby** and B.G. Celler***

*Digital Systems Processing Laboratory, Department of Electronic and Electrical Engineering, University College Dublin, Dublin 4, Republic of Ireland

**Biomedical Systems Laboratory, School of Electrical Engineering, University of New South Wales, Sydney, NSW 2052, Australia

ABSTRACT

The classification of the electrocardiogram (ECG) into different pathophysiological disease categories is a complex pattern recognition task. Successful classification is achieved by finding characteristic shapes in the ECG that discriminate effectively between categories. In this investigation selected wavelet coefficient values are used to represent the ECG shape information. A two-stage feature selection scheme was implemented. Firstly, high magnitude coefficients were determined, then, using a forward selection process, a subset of these that maximised the classification accuracy was found. Linear discriminants were used for the final classification.

Twenty-five wavelet coefficients were selected as inputs and cross-validation used to estimate the classifier performance. An overall accuracy of 72.3% was achieved using a database of 500 ECG records independently classified into seven classes. This compared well with published cardiologist classification rates. By introducing a no-classification state, the accuracy increased to 79% with 80% of ECG records classified.

The method presented here is not specific to the ECG domain and may easily be applied to other classification problems.

1. INTRODUCTION

Pattern recognition applied to biomedical signal processing is conventionally based on measurements directly taken from the time domain sampled data. This can lead to a redundant representation of the signal information. One such example of this is ECG classification.

Computer based classification of the ECG can achieve high accuracy and offers the potential of affordable mass screening for cardiac abnormalities. Conventionally the approach is to use measurements of heights and widths of the component waves of the QRS, T and possibly P waves to represent the diagnostic information. Classification is achieved on the basis of these measurements. Measurements based on QRS, T and P sections can vary enormously even among normals and can lead to misclassification. In this study we take an alternative approach and use wavelet coefficients to describe the ECG diagnostic information.

The selection of the significant coefficients from a wavelet transform provides a method of achieving high compression of the ECG with very little shape distortion. Hence, the significant coefficients provide efficient descriptors of the ECG shape. This fact can be exploited for the purpose of classification. A further step following application of the wavelet transform is to find a set of coefficients that efficiently describe the diagnostically important sections of the ECG. Unfortunately a priori knowledge of the significant coefficients is not known. So the problem reduces to one of a systematic search for significant coefficients.

2. METHODS

The Frank lead ECG *record* [1] presents three ECG signals that project the electrical field generated by the muscular tissue of the heart onto three mutually orthogonal planes. The Frank system attempts to compensate for distortions of the electric field introduced by the irregular shape of the human torso. In practice the Frank lead signals are approximate orthogonal views due to the wide variation of shapes and tissue content of human torsos.

A two-stage scheme was investigated for the selection of wavelet coefficients to use as inputs to a linear discriminant classifier. The first stage selected a set of candidate coefficients by choosing coefficients with the highest magnitude. The second stage selected a subset of the candidate coefficients that maximised the classification accuracy by using a forward selection process. To validate the method we investigated the classification of the Frank lead ECG as normal or one of six disease conditions.

2.1 ECG pre-processing.

The ECG is filtered with a 0.5 - 40 Hz linear phase digital bandpass filter to remove unwanted baseline drift and powerline interference. The QRS complexes are detected and a representative beat identified. A data window containing the P-QRS-T complex is isolated using the ECG samples in the range 200ms before the R-wave maximum point to 400ms after the R-wave maximum. The isopotential value is subtracted, and the data window multiplied with a Hanning window to ensure the end points are zero and thus eliminating possible edge effects.

0-7803-6293-4/00/$10.00©2000 IEEE

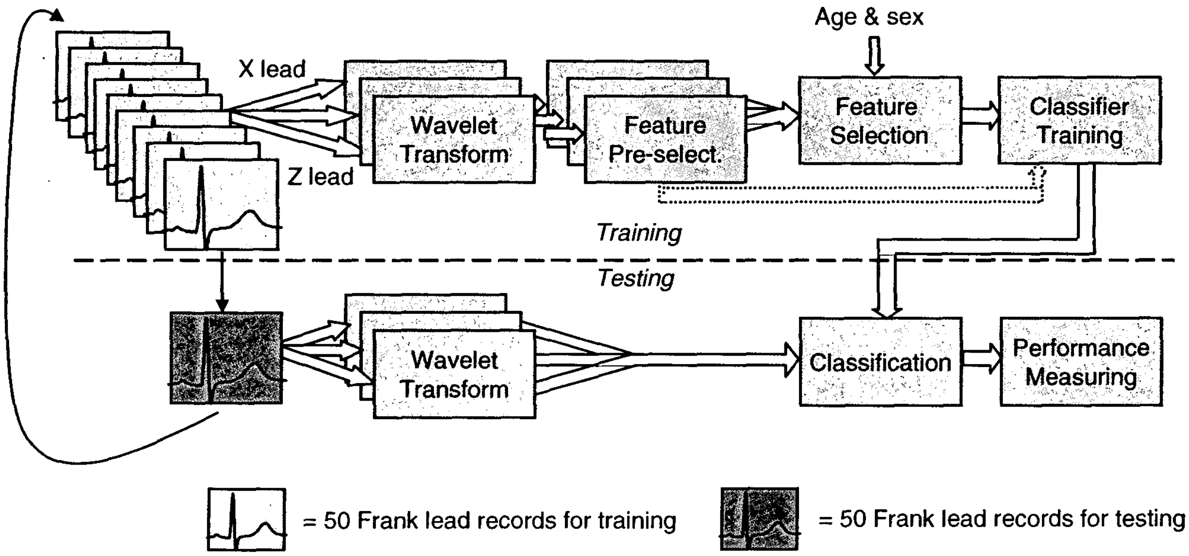

Figure 1: Data processing stages. The available ECG data is divided into a number of sub-groups (folds) as per the cross-validation scheme. The DWT is applied to all data. A two-stage feature selection process is applied to the training data. After feature selection, final training of the classifier is done and the classification performance estimated on the test data.

A 6-level discrete wavelet transform decomposition of each characteristic beat is achieved using a 10th order Daubechies wavelet [2]. The wavelet type and the decomposition level were chosen after some initial trials but they remain an area of the research requiring further investigation.

2.2 Feature pre-selection

Application of the wavelet transform results in 412 coefficients per lead that are possible input features to the classifier. Ranking the magnitude of the wavelet coefficients per lead showed that many of the features have a value close to zero indicating that most of the shape information is contained in relatively few coefficients [Figure 2]. Candidate features for processing by the next stage are found by choosing the most highly ranked coefficients (equivalent to lowest rank-sum) when averaged across the ECG records. The rank-sum is determined as follows:

- For each lead of each record the wavelet transform coefficients were determined and sorted into descending magnitude order. Each coefficient is then assigned a rank expressing its relative magnitude.
- The ranks are summed across the ECG records for each coefficient.
- The coefficients with the lowest ranks-sums are selected.

2.3 Classifier

A supervised training technique was used to derive a classifier. In supervised training, a classifier model with adjustable parameters is chosen and a representative set of labelled examples of the classes are used to optimise the parameters, using an error function. The classification model chosen is based on linear discriminants [3]. Linear discriminants partition the feature space into the different classes using a set of hyper-planes. Although this is a very restrictive condition, linear discriminants are extremely fast to calculate relative to other classifier building techniques. There is one output of the classifier for each class that represents a probability estimate. The final classification is obtained by choosing the class with the highest probability estimate.

Other classifier models such as logistic discrimination and feed forward neural networks impose less restrictive conditions on the partitioning but, in practice, linear discriminants perform as effectively for ECG classification [4].

2.4 Feature selection

The performance of most classifier training algorithms is degraded when one or more of the available features are redundant or irrelevant. Redundant features occur when two or more features are correlated whereas irrelevant features do not separate the classes to any useful degree. The classification performance of a given set of features may often be improved by searching for a subset of the features with higher performance. Finding this optimal subset is generally computationally intractable for anything apart from very small feature sets. This is because the number of possible subsets rises exponentially with size of the feature set. In practice a sub-optimal search procedure is used.

We have used a step-up selection procedure. Beginning with the empty feature set, all possible features sets containing one feature are generated. The classification performance of all these feature sets is evaluated and the set with the best performance retained.

***Table 1:** The overall accuracy and class sensitivities of the linear discriminant classifier processing features sets before and after feature selection. Results are obtained by averaging 10 runs of 10-fold cross validation.*

	Accuracy Overall (%)	Class Sensitivities (%) NOR	LVH	RVH	BVH	AMI	IMI	MIX	# Features
Feature pre-selection	71.6	91	59	30	35	71	82	32	62
Feature selection	72.3	89	61	39	44	75	81	26	25
CSE Cardiologist panel	74.8								-

This set is expanded to sets of two features by adding all remaining features one at a time. The classification performance of these feature sets is evaluated and the set with the best performance retained. This process is repeated until the classification performance no longer improves significantly with the addition of extra features.

2.5 Classification Performance Estimation

When developing a classifier it is important to be able to estimate the expected performance of the classifier on data not used in training. The available data must be divided into independent training and test sets. There are a number of schemes for achieving this and the most suitable for the size of data set used in this study, is n-fold cross validation [5]. This scheme randomly divides the available data into n approximately equal size and mutually exclusive "folds". For an n-fold cross validation run, n classifiers are trained with a different fold used each time as the test set, while the other n-1 folds are used for the training data. The choice of n influences the ratio of data used for training/testing with an optimal value of n in the range 5-20. Cross validation estimates are generally pessimistically biased, as training is performed using a subsample of the available data. Ten-fold cross validation was used in this study.

The randomising process was "stratified" so that all the folds contained the same relative proportions of normals and the six disease conditions. Studies have shown that stratification of the folds decreases both the bias and the variance of the performance estimate [5].

Cross validation estimates are highly variable and depend on the division of the data into folds. A decrease in the variance of the performance estimate may be achieved by averaging results from multiple runs of cross validation where a different random split of the training data into folds is used for each run. For this study ten runs of ten-fold cross validation were employed.

In this study we report the overall classification accuracy and the individual class sensitivities. The overall accuracy is the percentage of total cases correctly classified. The individual class sensitivities are the percentage of cases correctly classified of a particular class.

3. RESULTS

The ECG database used throughout this study contains 500 records with 155 normal (NOR) and 345 abnormal cases. The classification of every record is known with 100% certainty. The validation was based on ECG independent clinical information, consisting of data derived from cardiac catheterisation, coronary angiography, left ventriculography, echocardiography and results from coronary care and cardiac surgery. The abnormal cases comprise 79 left ventricular hypertrophy (LVH), 21 right ventricular hypertrophy (RVH) and 25 bi-ventricular hypertrophy (BVH) cases; 77 anterior myocardial infarction (AMI), 111 inferior myocardial infarction (IMI) and 32 combined myocardial infarction (MIX) cases. Each case contained between eight to ten seconds of digitally sampled (500 Hz sampling rate) data from simultaneously recorded Frank lead ECGs.

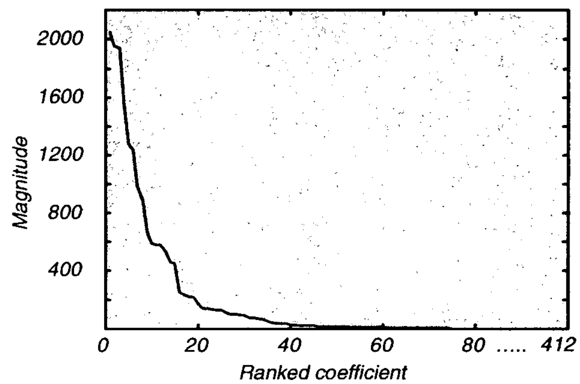

***Figure 2:** Plot of the ordered wavelet coefficient magnitudes computed from a representative ECG record.*

Figure 2 depicts a typical plot of the decay of the magnitude of the wavelet coefficients determined from an ECG lead. The magnitude of the largest 20 coefficients decay very quickly. The magnitude of the next 20 coefficients decay more slowly and beyond about the 40th ranked coefficient the magnitudes are very close to zero.

A pre-selection feature set containing 62 features was formed by choosing the 20 coefficients with the lowest rank-sums for the X, Y and Z leads and adding the features describing the age and sex of the patient. These features were used as inputs to the linear discriminant classifier. Using 10-fold cross-validation, the overall accuracy was 71.6% as shown in the first row of table 1. The individual class sensitivity results reveal that the classification rates for normals was high (91%); for the ventricular hypertrophies classes it was low (LVH-59% RVH-30% and BVH-35%) and for infarctions was high for the AMI (71%) and IMI (82%) classes and low for MIX (32%). The three classes, RVH, BVH and MIX, have 32 or fewer examples each and probably are not large enough for effective training of the classifier.

The next step was to find a subset of the above features with equivalent or better classification performance. The forward selection process was applied separately to each training-set and selection was stopped after 25 wavelet coefficients were chosen. Once the coefficients were determined the classifier was independently validated on the test-set. Again using 10-fold cross-validation, the overall accuracy was 72.3% as shown in the second row of table 1. Hence, feature selection marginally improved the accuracy while reducing the required inputs from 62 to 25. Analysis of the class sensitivities reveals that feature selection has increased the classification rate of most of the disease classes. For the hypertrophy classes LVH improved from 59% to 61%, RVH from 30 to 39% and BVH from 35% to 44%. For the infarction classes, AMI improved from 71% to 75%, IMI was effectively the same at 81% while MIX decreased from 32% to 26%. There was a slight reduction in the classification rate of the normal class from 91% to 89%. In general terms feature selection has tended to equalise the class sensitivities.

It is possible to compare our results to the classification performance achieved by cardiologists assessed in the Common Standards in Electrocardiography (CSE) project [6]. The cardiologists diagnosed a similar ECG database into the same seven categories considered in this study. The proportions of the classes were similar making a direct comparison of overall accuracy possible. In the CSE project, five cardiologists classified the ECG and achieved accuracy rates between 67.5% and 74.4% [6]. Our result compares favourably.

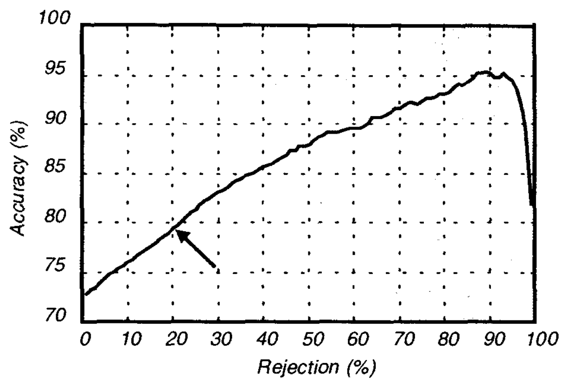

Figure 3: Plot of the classification accuracy versus the percentage of records rejected. The arrow indicates a possible setting that could be used in the final classifier.

The outputs of the classifier represent the posterior probabilities of each class and hence provide a confidence in the decision. A useful post-processing step is to threshold the outputs so that if none of the outputs exceed the threshold then no attempt is made to classify the ECG record.

The effect of varying the threshold on the overall classification rate was investigated and the results are shown in figure 3. Note that the x-axis is shown in terms of percentage of records rejected and this increases as the threshold is raised. Initially the graph shows a linear improvement of the classification rate as more records are rejected. At approximately 90% rejection the accuracy reaches a maximum of 95%. Beyond this point rejecting more records reduces the accuracy. Although high accuracy is obtained at the maximum the classifier would be of little practical use, as only 10% of records are classified. The arrow on figure 3 indicates a more realistic setting. At this point, 80% of records are classified and the overall accuracy is 79%.

4. SUMMARY

A two-stage feature selection scheme was successfully used to select wavelet coefficients for processing by a linear discriminant classifier. Processing just 25 wavelet coefficients it achieved an overall accuracy of 72.3%. This compares very well to the published performance of cardiologists (67.5% - 74.4%). By introducing a no-classification state the overall accuracy increased to 79% with 80% of the ECG records being classified.

Currently, possible enhancements to classification performance using other wavelets are being investigated together with the use of the wavelet packet transform.

5. REFERENCES

[1] E. Frank, "An Accurate and Clinically Practical System for Spatial Vectorcardiography," *Circulation*, vol. 13(May), pp. 737-749, 1956.

[2] M. Vetterli and C. Herley, "Wavelets and Filter Banks: Theory and Design," *IEEE Trans. on Signal Processing*, vol. 40, No. 9, September, pp. 2207-2232, 1992.

[3] B. D. Ripley, *Pattern recognition and neural networks*, Cambridge, England: Cambridge University Press, 1996.

[4] J.L. Willems and E. Lesaffre, "Comparison of multigroup logistic and linear discriminant ECG and VCG classification", *Journal of Electrocardiology*, vol. 20(2), pp 83-92, 1987.

[5] R. Kohavi, "A study of cross validation and bootstrap for accuracy estimation and model selection", *In Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 1137-1143, 1995.

[6] J. Willems et al, "Comparison of Diagnostic Results of ECG Computer Programs and Cardiologists", *Proceedings of the Computers in Cardiology Conference*, Durham, North Carolina, pp. 93-96, 1992.

This work was supported under the Enterprise Ireland Strategic Research Programme as ST/98/023.