AUTOMATIC CLASSIFICATION OF ECG BEATS USING WAVEFORM SHAPE AND HEART BEAT INTERVAL FEATURES

P. de Chazal, R. B. Reilly

Digital Signal Processing Research Group, Department of Electronic and Electrical Engineering,
University College Dublin, Republic of Ireland

ABSTRACT

This paper presents the classification performance of an automatic classifier of the electrocardiogram (ECG) for the detection of normal, premature ventricular contraction and fusion beat types. Both linear discriminants and feed forward neural networks were considered for the classifier model. Features based on the ECG waveform shape and heart beat intervals were used as inputs to the classifiers. Data was obtained from the MIT-BIH arrhythmia database. Cross-validation was used to measure the classifier performance. A classification accuracy of 89% was achieved which is a significant improvement on previously published results.

1. INTRODUCTION

Studying the electrocardiogram (ECG) signal provides an insight to understand life-threatening cardiac conditions. This typically is centered on the study of arrhythmias, which can be any disturbance in the rate, regularity, and site of origin or conduction of the cardiac electric impulse. Not all arrhythmias are abnormal or dangerous but some do require immediate therapy to prevent further problems.

A subject's ECG information can be recorded using a portable Holter monitor which is worn by the subject. A Holter monitor typically employ a few electrodes and store a recording of the subject's heart rhythm as they go about their daily activities over a 24 to 48 hour period. The Holter monitor is then returned to a cardiologist who examines the recordings and determines a diagnosis. Examining these recordings is a time-consuming and hence any automated processing of the ECG that assists the cardiologist in determining a diagnosis would be of assistance.

Beat classification is an important step in arrhythmia analysis as many arrhythmias simply consist of a single aberrant beat as opposed to a sustained rhythm disturbance. A beat classifier attempts to classify a heartbeat into a normal beat or into a class representing one of many different arrhythmias. The rhythm of some ECG signals can be determined by knowing the beat classification of a number of consecutive beats in the signal.

Beat classification is a candidate task for automatic pattern recognition as it involves the labeling of beats on the basis of ECG waveform shape and temporal position relative to surrounding beats. Other authors have described previous work on this problem

Senhadji et al. [1] explored the use of the discrete wavelet transform to discriminate between three beat types. Using the Daubechies' orthogonal wavelets, Spline wavelets, and Morlet type wavelets they employed a beat classifier modelled on linear discriminants processing input features derived from distributions of energy and local extrema in the details corresponding to different levels of decomposition. Their study was conducted on a database set of 53 beats consisting of 20 normal, 13 premature ventricular contractions and 20 beats with an S-T segment deviation. The data was divided into training (25 beats) and testing (28 beats) data sets. The classifier achieved an accuracy of 98% in classifying the beats and was found to outperform a classifier processing features derived from the maximum magnitude of the P, QRS, and T waves, the P-R and ST intervals and power spectral density measurements. Until the results are validated on a significantly larger database it is difficult to draw any real conclusions from this work.

The beat classifier designed by Yeap et al. [2] was modelled using a feed forward neural network. The classifier's performance was tested on the American Heart Association database; beats were classified into normal or premature ventricular contractions beat types. Four of the 80 available ECG records were used to train the classifier; the remaining records (excluding the ventricular tachycardia records) were used to test the classifier. The neural network consisted of two hidden layers each with 20 processing units. The input feature vector consisted of five features: the QRS width, the R wave's amplitude, a measure of the QRS offset, the T wave slope and a measure of the R-R interval with respect to its mean value. In testing, an accuracy of 93.3% was achieved with a sensitivity of 67.6% and a specificity of 97.9%.

The classification rates of automatic beat classifiers presented in the literature to date have not been high enough for the classifiers to gain wide spread clinical acceptance. Hu et al. [3] notes that certain beat types are sufficiently rare that to date not enough ECG data has been collected to obtain a representative sample of the these populations and hence classifier training procedures are unable to properly model these classes. In order to boost the classification performance, they suggested customising a beat classifier to a specific patient (known as a local classifier) and then combining it with a global classifier designed from a large database of ECG signals. They modeled a global classifier on a feed-forward neural network with one hidden layer of seven processing units. They used self-organising maps and learning vector quantisation to design the local classifier. The two classifiers were then combined using a mixture of experts (MOE) approach. The MIT-BIH Arrhythmia database was used to examine the MOE classifier. Thirteen recordings were used to train the global classifier and 20 recordings were used to simulate the records of 20 patients. The feature vector consists of the QRS width, the instantaneous RR interval, the average R-R interval and 9 elements representing the QRS template and the classifier considered normal, premature ventricular contraction and fusion beats only. The global classifier achieved an accuracy of 62% on the second set of recordings. The local classifier significantly enhanced the performance of the global classifier with the MOE classifier achieving 94% accuracy on the same data set. In practice this method requires a cardiologist to annotate a segment of a patient-specific ECG in order implement the MOE approach. The main drawback of this approach therefore lies in the expert input required to customise this approach to each patient.

2. AIM

The aim of this work was to investigate the performance of an aotomatic beat classifier categorising ECG recordings from the MIT-BIH Arrhythmia database into different beat classes. For this study the same classes as used by Hu et al. [3] were considered i.e. normal (N), premature ventricular contraction (PVC) and fusion (F) beats. Our goal was to produce a classifier requiring no expert input with similar performance to the MOE approach used by Hu.

Two types of classifier models were considered: a linear discriminant model and a feed forward neural network model. In assessing the performance the following criteria were considered: 1) Division of the available data to obtain unbiased performance measures 2) the processing required to extract the features, 2) the processing requirements of the classifier, 3) the class sensitivities achieved and 4) the overall accuracy achieved.

3. METHODS

Data from the MIT-BIH Arrhythmia database [4] was used in this study, which includes recordings of many common and life-threatening arrhythmias along with examples of normal sinus rhythm. The database contains 48 recordings each containing two ECG signals. The data is band-pass filtered at 0.1-100Hz and sampled at 360Hz. There are over 109,000 labelled ventricular beats from 13 different beat classes. The ECG data associated with beats belonging to the normal, PVC and fusion beat classes was selected. The size of the classes is respectively 75,054; 7,129 and 803 beats.

Due to the large numbers of normal and PVC class examples relative to the fusion class we weighted the contributions of each example to the training process according to its class as follows. The examples from these classes were weighted so that both classes contributed the equivalent of 1000 examples to the training process. To weight a class the required weighting rate (w) was calculated and then the contribution to the likelihood error function data of each example of that class weighted by w. For example, the normal class contains 75,054 beats so each example was weighted with a factor of 1000/75,054 = 0.0133 in the likelihood function. For the fusion beat class no weighting was used. The purpose of the weighting was to ensure that the large classes did not dominate the learning process.

ECG segmentation. The arrhythmia database provides QRS detection times for all classified beats and these were used as a starting point for the signal processing used in this study. The QRS detection times occur at the instant of the major local extremum of the QRS complex (i.e either the time of the R wave maximum or S wave minimum).

	Beat	ECG signal 1				ECG signal 2			
Feature	interval	int.	mag.	area	flag	int.	mag.	area	flag
pre RR	Y								
post-RR	Y								
avg. RR	Y								
local avg RR	Y								
P-R		Y				Y			
QRS		Y				Y			
Q-T		Y				Y			
P wave		Y	Y	Y	Y	Y	Y	Y	Y
Q wave		Y	Y	Y		Y	Y	Y	
R wave		Y	Y	Y		Y	Y	Y	
S wave		Y	Y	Y		Y	Y	Y	
R' wave		Y	Y	Y		Y	Y	Y	
S' wave		Y	Y	Y		Y	Y	Y	
T wave			Y	Y			Y	Y	

Table 1: Features processed by the classifiers
Key: int. – interval; mag. – magnitude. Y – feature included

The ECG segmentation algorithm of Laguna et al [5,6] was used to provide estimates of QRS onset and offset and T wave offset times and, if present, the P wave onset and offset time for the two ECG signals provided for each annotated beat of the database. This algorithm has been validated on the CSE multilead database [5] and the MIT-BIH QT database [6]. In both cases the accuracy of the method in determining waveform boundary points was comparable with the inter-expert variation.

Feature extraction. Table 1 summarises the features used to characterize the ECG in this study. Features relating to RR intervals were calculated for each beat and features relating to the P, QRS and T waves were calculated for the two ECG signals available for each beat. A total of 52 features were calculated for each beat.

RR interval features. RR intervals were defined as the interval between successive heart beats. Heart beats were identified by locating the QRS complexes in the ECG. Due to poor signal quality resulting in heart beats being missed some of the RR intervals generated were physiologically unreasonable. All RR intervals with duration greater than 2 seconds were replaced with a code indicating the interval was invalid.

Four features were extracted for each beat from the RR sequence. The pre-RR interval was the RR interval between a given beat and the previous beat. The post-RR interval was the RR interval between a given beat and the following beat. The average RR interval was the mean of the valid RR intervals for a recording. This feature had the same value for all beats in a recording. Finally, the local average RR interval was determined by averaging the valid RR intervals of the ten RR intervals surrounding the beat.

P wave features. Five features relating to the P-wave were determined for each ECG signal for each beat. The first feature was a binary flag indicating the presence or absence of the P wave. When a P wave was not present the following three features were set to a code indicating their value was invalid otherwise they were calculated as follows. The P-R interval was defined as the time interval between the P wave onset and the

QRS onset. The P-wave duration was the time interval between the P-wave onset and P-wave offset. The P-wave area was the area enclosed by the P-wave relative to the P-wave baseline. The P-wave amplitude was the amplitude of the maximum deviation of the P-wave from the P-wave baseline. The P-wave baseline was calculated using the procedures recommended by the CSE working party for ECG waveform measurement [7].

QRS complex features. Sixteen features relating to the QRS complex were determined for each ECG signal for each beat. Before features were extracted the QRS complex for a given beat was segmented into its component waves using the standards for waveform determination and naming recommended by the CSE working party [7]. After determining the QRS complex baseline any Q, R, S, R’ and S’ waves were identified. Any waves beyond the S’ wave in the QRS complex were ignored. For each identified wave the following features were calculated. The wave duration defined as the time interval between the onset and offset of the wave. The wave amplitude defined as the maximum deviation of the wave from the QRS complex baseline and the wave area defined as the area enclosed by the wave between its onset and offset relative to the QRS complex baseline. In the case of a wave not being identified the wave duration, amplitude and area were set to a code indicating an invalid measurement. The sixteenth feature was the QRS duration which was defined as the time interval between the QRS onset and QRS offset.

T wave features. Three features relating to the T-wave were determined for each ECG signal for each beat. The Q-T interval was defined as the time interval between the QRS offset and the T-wave offset. The T-wave area was the area enclosed by the ECG trace between the QRS offset and the T-wave offset relative to the QRS complex baseline. The T-wave amplitude was the amplitude of the maximum deviation of the ECG signal between the QRS offset and T-wave offset from the QRS baseline.

Feature sets. Two feature sets were formed each with 28 features. Feature set 1 (FS1) contained the RR interval features and the P, QRS and T wave features derived from ECG signal 1. Feature set 2 (FS2) contained the RR features and the P, QRS and T wave features from ECG signal 2. These two feature sets were formed to determine the effect of lead placement on classification performance.

Classifier models. Two statistical classifier models were chosen in this study so that the effect of classifier model on performance could be examined.

Linear discriminants (LD) [8] partition the feature space into the different classes using a set of hyper-planes. The parameters of this classifier model were fitted to the available training data by using the method of maximum likelihood. Using this method the calculations required for training is achieved by direct calculation and is extremely fast relative to other classifier building techniques such as neural networks. This model assumes the feature data has a Guassian distribution for each class.

Neural networks (NN) [8] implementing logistic discriminants impose fewer conditions on the feature space partitioning than linear discriminants. The model assumes the feature data has a class distribution belonging to one of the family of exponential distributions. This family includes many of the common distributions such as the Gaussian, binomial, Bernoulli and Poisson. Direct optimisation of the model parameters is not possible and an iterative numerical optimisation technique is required. The logistic discriminant model was implemented with feed-forward neural network. A network with one layer of hidden units and a softmax output stage was used and the network parameters optimized by minimizing the (negative) log-likelihood error function. The number of hidden units controls the flexibility of the feature space partitioning with more hidden units allowing greater flexibility. Optimisation of the parameters (weights) of the network was achieved with a gradient-descent algorithm with an adaptive learning rate and momentum constant. Training was stopped when the successive iterations no longer resulted in a significant reduction in the error function. The weights of hidden units were optimised with the back-propagation algorithm.

In response to input features, both models calculate a probability estimate of each class. The final classification is obtained by choosing the class with the highest probability estimate.

Classifier's performance. In this study the performance of the classifier is quoted using the specificity, the class sensitivities and the overall accuracy. The sensitivity of the classifier to a particular beat class is the fraction of beats in the class that are correctly classified. The specificity is the sensitivity calculation applied to the normal class. The overall accuracy is the fraction of the total number of beats classified correctly.

Estimating the classifier performance. Training of the classifier involves the optimisation of classifier model parameters using available training data hence care must be taken when estimating the classifier performance to obtain unbiased figures. One approach to this problem is to use independent data for training and testing. The n-fold cross validation scheme achieves this by randomly dividing 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.

Combining Classifiers. A classification based on processing information from both feature sets simultaneously was obtained by combining the posterior probabilities obtained from each feature set. To classify an ECG beat, the classifier processes the feature information of each ECG signal separately and a set of probabilities for each beat is determined. To obtain the final classification, the probabilities for each class are averaged across the two ECG signals and the class with the highest average probability estimate chosen. By using diagnostic information from all available signals more efficient use of the available ECG diagnostic information is made.

Thresholding the posterior probabilities. The outputs of the two classifiers 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 beat.

4. RESULTS

The feature data was divided into 48 folds with each fold containing data from one record. Seven hidden units were used in

Feature Set	Model	Test Acc	N	PVC	F	Train Acc
FS1	LD	85.7	86	88	79	88.8
	NN	88.5	88	92	65	95.3
FS2	LD	86.3	87	84	49	87.9
	NN	86.3	87	90	14	93.7
FS1+FS2	LD	89.1	89	88	68	
	NN	89.1	89	94	60	

Table 2: Classification results for the different feature set and classifier model combinations as estimated using cross-validation. The accuracy, specificity and sensitivies are shown for the testing set while only accuracy is shown for the training set. All figures are percentages.

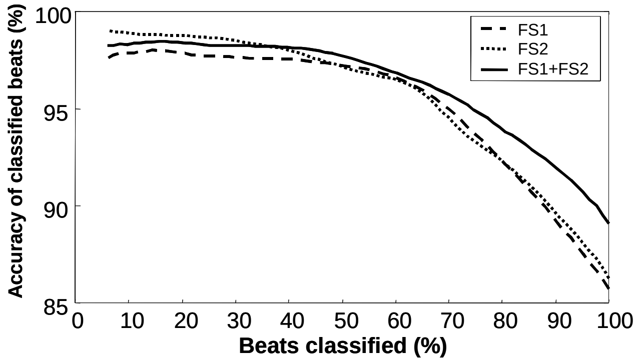

Figure 1: Plot of classification accuracy versus the percentage of records classified as the posterior probability threshold is varied for the LD classifier.

the NN model. For the LD classifier the prior probabilities were set to 1/3 for each class.

Results are shown in Table 2. Separate classifier performance figures were obtained for the four combinations of classifier models and ECG signals. Results for the two classifier models processing both feature sets simultaneously are also shown. The effect of varying the posterior probability threshold on the overall classification rate was investigated and the results are shown in Figure 1. Note that the x-axis is shown in terms of percentage of records classified and this decreases as the threshold is raised.

5. DISCUSSION

There was only a moderate difference in the classification performance obtained from the classifiers for the two feature sets. For FS1 the LD classifier resulted in an accuracy of 85.7% and for FS2 the accuracy was 86.3%. The main difference was the lower sensitivity result for the fusion beat class of 49% of FS2 compared to 79% for FS1. A similar trend was observed for the NN classifier results with the sensitivity for fusion beat class being 14% for FS2 compared to 65% for FS1. The overall accuracy performance of the NN classifier was 1.8% higher than the LD classifier for FS1 and the same overall accuracy was obtained for both classifiers for FS2 (86.3%).

Combining the posterior probabilities from the two feature sets proved to be a good strategy with the LD and NN classifier both obtaining an overall accuracy of 89.1%. This result is significantly better than the figure of 62% obtained by Hu et al. [3] for their global classifier and only 5% less than the result they obtained for MOE approach which required expert annotation of a section of the ECG before final classification.

As the posterior probability threshold is raised the number of beats classified reduces and the classification accuracy increases. At the point where 50% of beats are classified the accuracy for FS1, FS2 and FS1+FS2 is approximately 97% and any further increase of the threshold makes only a small increase in the accuracy.

6. CONCLUSION

The choice of classifier model did not have a significant influence on the overall classification results. Processing ECG data from the two ECG signals simultaneously the linear discriminant and the neural network model achieved an overall accuracy of 89.1%. The sensitivities of the linear discriminant model were more evenly balance than the results from the neural network model. The results obtained were a significant improvement on comparable classifiers reported in the literature.

7. REFERENCES

[1] D. E. Ingalls, "Image Processing for Experts," IEEE Trans. ASSP, vol. ASSP-36, pp. 1932-1948, 1988

[2] L Senhadji, G Carrault et. al, “Comparing Wavelet Transforms for Recognizing Cardiac Patterns”. IEEE Engineering in Medicine and Biology Mar/Apr pp 167-173, 1995.

[3] TH Yeap, F Johnson, M Rachniowski, “ECG Beat Classification by a Neural Network”. Proceedings Annual International Conference of the IEEE EMBS Society pg 1457-58, 1990.

[4] Y. H. Hu, S. Palreddy and W. J. Tompkins, "A Patient-Adaptable ECG Beat Classifier Using a Mixture of Experts Approach", IEEE Transactions on Biomedical Engineering, pg. 891-900, vol. 44, 1997.

[5] “MIT-BIH Arrhythmia Database, 3rd Edition.” May, 1997.

[6] P Laguna, R Jané, P Caminal. “Automatic Detection of Wave Boundaries in Multilead ECG Signals: Validation with the CSE Database”. Computers and Biomedical Research 27(1):45-60, 1994.

[7] R Jané, A Blasi, J García, P Laguna. “Evaluation of an automatic threshold based detector of waveform limits in Holter ECG with the QT database” Computers in Cardiology 24:295-298, 1997.

[8] The CSE Working Party “Recommendations for measurement standards in quantitative electrocardiography,” European Heart Journal, vol. 6(10), pp. 815-825, 1985

[9] BD Ripley. “Pattern Recognition and Neural Networks”, Cambridge, England: Cambridge University Press, 1996.

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