

Beat Classification for Use in Arrhythmia Analysis

M O'Dwyer, P de Chazal, RB Reilly

University College Dublin, Ireland

Abstract

An important step in the classification of arrhythmias is beat classification. The aim of this work is to classify the beats in the MIT-BIH Arrhythmia database and to investigate pertinent features for beat classification. A comparison is made of the classification performance using features obtained after three different processing tasks on the electrocardiogram: QRS detection, QRS onset/offset detection and P wave onset detection. Linear discriminants and quadratic discriminants were employed to model the classifier. The best classification performance of 82% was achieved by employing the QRS width as an input feature to the linear discriminant classifier.

1. Introduction

Classifying arrhythmias involves the recognition of characteristic patterns of the electrocardiogram (ECG). Beat classification is an important step in designing an arrhythmia classifier as many arrhythmias simply consist of a single aberrant beat as opposed to a sustained rhythm disturbance. Previous studies have employed different features in most beat classifiers. Senhadji [1] investigated features extracted from the wavelet coefficients using linear discriminants. Hu [2] used the amplitude of points surrounding the QRS complex as features and a neural network model as the classifier. Yeap [3] also employed neural networks as the classifier model and used the QRS width and amplitude along with three other measurements made on the ECG as features. It is difficult to compare the results as these studies employed different data sets.

The aim of this work is to investigate different combinations of features for beat classification. The following criteria were considered: 1) the processing required to extract the features, 2) the use of a feature selection method, and 3) the overall accuracy achieved. It is hypothesised that features suitable for beat classification will form the basis of those used in an arrhythmia classifier. The classification accuracies achieved in these studies are un-biased, as the data used to test and train the classifiers are independent statistical samples.

2. Methodology

The classification method developed here involves a number of steps: 1) Processing of the ECG signal. 2) Feature extraction. 3) Feature selection. 4) Training of beat classifiers. 5) Testing the beat classifiers. The MIT-BIH Arrhythmia database [4] consists of ECG data band-pass filtered at 0.1-100Hz and sampled at 360Hz. This data was used to evaluate the performance of the classifiers in this study. A data set was selected from the beginning of records 100, 101, 103, 105, 106, 118, 119, 200, 202, 205, 207, 208, 209, 212, 213, 214, and 215. Episodes of ventricular flutter in record 207 were excluded as this beat type only appeared in this record. Table 1 shows a summary of the data set.

Table 1. Summary of the Data Set Used in this Study

Beat Type	No. Subjects	No. Beats	% Beats
APC	9	70	0.74
Fusion	3	191	2.01
LBBB	2	691	7.28
Normal	14	6747	71.04
RBBB	3	1015	10.10
PVC	12	784	8.25
Total	17	9498	100

2.1. Processing of the ECG signal

In order to extract features from the ECG signal certain processing tasks are necessary. The first of these tasks involves measurement of the R-R intervals, the P-R intervals and the QRS width. The detection of the P, Q, R, and S wave in the electrocardiogram is a complex problem since they have a time varying morphology and are subject to physiological variations due to individual patient variation and corruption from noise. Considerable research has been carried out in the development of QRS and P wave detectors; a review of these detectors can be found in [5] and [6]. However in this study the QRS width, the R-R and the P-R intervals were marked manually with the aid of the annotation files.

The second of these processing tasks involves extracting templates from the ECG waveform. The ECG waveform is band-pass filtered (0.5-40 Hz) to remove baseline wander and power line interference. Then the

waveforms are down-sampled at 80Hz. A window is taken around each beat, 400ms before the R wave and 600ms after. For each beat a beat template is formed using the amplitudes of the resulting 80 points in the windowed segment. For each beat a QRS template is formed using the amplitudes of 9 points extracted from a segment windowed 50ms before and after the R wave.

2.2. Feature extraction

In order to develop an efficient beat classifier one needs to establish features to distinguish between the various beats. The first set of features is extracted from the ECG after the beat (QRS) detection process. This set consists of features based on the R-R interval (RR), the amplitude of points in the beat template (B amp) and the amplitudes of points in the QRS template (QRS amp). After the QRS onset/offset of each beat detected is determined, features based on the QRS width (QRS w) are extracted. The remaining set of features is extracted after the P wave onset of each detected P wave is determined. This set consists of features based on the P-R interval (PR). Table 2 shows a summary of the processing required prior to feature extraction and the features available.

Table 2. Available Features

Processing to Feature Extraction	Features Based on	No. Features
QRS Detection	R-R interval	10
	Beat amplitudes	80
	QRS amplitudes	9
QRS Onset/Offset Detection	QRS interval	5
P wave Onset Detection	P-R interval	2

2.3. Feature selection

The performance of most classifier training algorithms is degraded when one or more of the available features are redundant, i.e. when two or more features are correlated. Features are irrelevant if they do not separate the classes to any useful degree. Data is initially screened for correlated features, those with a correlation of 99% or greater are removed. 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. In this study a step-up selection procedure was used. 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. This set is expanded to sets of two features by adding all remaining features one at a time. The performance of these feature sets is evaluated and the set with the best performance retained. The procedure is repeated until the classification performance no longer improves.

2.4. Training a classifier

A supervised training technique has been used to derive the beat classifiers. A review of classification methods suitable for ECG signals can be found in [7] and [8]. For this study, linear discriminants [9] and quadratic discriminants have been used as the classifier models. 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 such as neural networks. Quadratic discriminants, which are less restrictive than linear discriminants partition the feature space into different classes using a set of hyper-quadratics. Quadratic discriminants are guaranteed to outperform linear discriminants for very large data sets since they are less restrictive, however they can be outperformed by linear discriminants for moderate data set sizes [9]. In response to a set of input features, the output of the classifier modeled on either of these discriminants is a set of numbers representing the probability estimate of each class. The final classification is obtained by choosing the class with the highest probability estimate.

2.5. Testing a classifier

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. In this study n -fold cross validation scheme was employed; 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 sub-sample of the available data.

In this study two schemes were used for dividing the data into folds. In scheme one each ECG beat is considered as a separate entity; the beats are randomly divided into training and testing data. Although this method is commonly used in the literature [2], a

significant problem with this approach is that intra-subject beats are highly dependent and hence dependent beats may appear in the training and testing sets. An improved scheme is to divide the data on a per-subject basis so that training and testing fold data contains beats from separate subjects.

3. Results

Two classifiers were used to generate the results; one was modelled on linear discriminant analysis (LDA) and the other on quadratic discriminant analysis (QDA). Beat classifiers were trained and tested using sets of features corresponding to increasing levels of processing. Table 3 summarises results obtained from the beat classifiers using data divided into folds under scheme one, i.e. all beats are considered as separate entities. Column 6 of this table shows the average LDA classification accuracy. Using the 10 features based on the R-R interval, the LDA classifier achieves an accuracy of 78% whereas an accuracy of 94% is achieved if the beat template amplitudes are employed as features along with the R-R intervals by the LDA classifier.

Table 3. Average LDA Testing Accuracy. All beats are treated as separate entities, different combinations of features are used, and no feature selection is performed.

RR	Features Employed				Acc (%)	
	B amp	QRS amp	QRS w	PR	LDA	
10					77.92	
	80				92.21	
		9			84.41	
10	80				94.06	
10		9			87.94	

The beat classifiers using data divided into folds under scheme two generated the remaining results in this paper. Scheme two divides data into folds on a per subject basis. Results in tables 4a-c were generated by directly applying different combinations of features as inputs to the beat classifiers. Column 6 of this table shows the average LDA classification accuracy. No feature selection was performed. Table 4a shows the testing accuracy for the LDA classifier using combinations of features extracted after QRS detection. Table 4b shows the testing accuracy for the LDA classifier using combinations of features extracted after QRS detection and QRS onset/offset detection. Table 4c shows the testing accuracy for the LDA classifier using combinations of features extracted after QRS detection, QRS onset/offset detection, and P wave onset detection.

For all configurations, except for using the two PR features the QDA classification accuracy was misleading. The extra flexibility in the QDA model is more sensitive to correlation of features in the smaller classes, which

leads to numerical instability. Hence, the classifier could not be modeled successfully for many feature combinations. Successful modeling was achieved for the PR features resulting in an accuracy of 77%.

Tables 4a-c. Average testing accuracy using features extracted after detection from each beat of a) the R waves, b) the R waves and the QRS Onset/Offset points, and c) the R waves, QRS Onset/ Offset points and the onset of any detected P waves.

RR	Features Employed				Acc (%)	
	B amp	QRS amp	QRS w	PR	LDA	
10					72.86	
	80				75.23	
		9			74.31	
10	80				71.23	
10		9			73.05	

RR	Features Employed				Acc (%)	
	B amp	QRS amp	QRS w	PR	LDA	
			5		75.32	
10			5		72.80	
10	80		5		73.67	
10		9	5		76.91	

RR	Features Employed				Acc (%)	
	B amp	QRS amp	QRS w	PR	LDA	
				2	79.75	
10				2	75.50	
10	80			2	75.09	
10		9		2	75.23	
10			5	2	75.93	
10	80		5	2	78.02	
10		9	5	2	78.43	

Results in tables 5a-c were generated by applying the feature selection process to each of the different combinations of features in tables 4a-c above. Columns 6 and 7 of table 5 show the average classification accuracy for the LDA and the QDA classifiers. For each combination considered the features that gave the best separation between the classes were determined and used to train the beat classifier. Table 5a shows the testing accuracy for the LDA and QDA classifiers using features selected from combinations of features extracted after QRS detection. In table 5b, the results correspond to features selected from combinations of features extracted after QRS detection and QRS onset/offset detection. Results in table 5c correspond to features selected from combinations of features extracted after QRS detection, QRS onset/offset detection, and P wave onset detection. For example applying feature selection method to the 80 beat amplitude features results in an average accuracy of 78% (LDA) when 10 of the 80 features are selected.

Tables 5a-c. Average testing accuracy using feature selection on features extracted after detection from each beat of a) the R waves, b) the R waves and the QRS Onset/Offset points, and c) the R waves, QRS Onset/Offset points and the onset of any detected P waves. The number of features selected is indicated by the following symbols: †=2, ‡=3, °=4, ê=5, ÷=7, ' =8, **=9, *=10. No symbol indicates only one feature was selected.

RR	Features Employed			PR	Classifier Acc (%)	
	B amp	QRS amp	QRS w		LDA	QDA
10					76.66	75.99 [†]
	80				78.81 [*]	73.57 [†]
		9			74.63 [†]	69.74 ^{**}
10	80				77.75 [†]	77.77 [*]
10		9			74.74 [†]	75.61 ^{**}

RR	Features Employed			PR	Classifier Acc (%)	
	B amp	QRS amp	QRS w		LDA	QDA
			5		82.22	80.35
10			5		82.22	80.35
10	80		5		82.22	77.81 [†]
10		9	5		82.22	80.35

RR	Features Employed			PR	Classifier Acc (%)	
	B amp	QRS amp	QRS w		LDA	QDA
				2	77.95	77.71
10				2	77.98 [†]	77.71
10	80			2	78.47 ^ê	77.77 [*]
10		9		2	74.04 [†]	76.14 ^{**}
10			5	2	82.22	80.35
10	80		5	2	82.22	77.81 [†]
10		9	5	2	82.22	80.35

4. Discussion

The results obtained by dividing data into folds under scheme one (i.e. divided randomly, see table 3) are significantly higher when compared to the results obtained by dividing data into folds under scheme two (i.e. divided on a per subject basis, see table 4a), as intra-subject beats are highly dependent. Considering all beats as separate entities may result in dependent beats appearing in the training and testing sets. Therefore, more realistic results are obtained by employing scheme two.

By comparing table 4 with table 5, it is apparent that feature selection greatly enhances the classifier's performance by removing the irrelevant features. The step-up feature selection procedure employed in this study does not always find the optimal feature set, but it does find a close approximation in a very short time.

By increasing the level of processing required from QRS detection to QRS onset/offset detection the classification accuracy can be increased substantially. Using features extracted after QRS detection, the

maximum classification of 79% was achieved by the LDA classifier employing 10 of the beat template amplitudes. The QDA classifier achieved an accuracy of 78% selecting 10 features from the R-R intervals and the beat template amplitudes. An extra level of processing enables features based on the QRS width to be extracted. Combining these features with features extracted after QRS detection increases the LDA classification accuracy to 82% and the QDA classification accuracy to 80%. Introducing the features (PR) extracted after the next level of processing, P-wave onset detection, the classification accuracy was not increased. However, by solely employing the P-R interval as an input feature both the LDA and QDA classifier achieve an accuracy of 78%. The best accuracy is achieved using one feature: QRS width. The LDA classifier using this feature resulted in an average accuracy of 82.2%.

5. Conclusion

The QRS width achieves the best accuracy and does not require significant processing to acquire this interval. Dividing data into folds using scheme two results in unbiased beat classification as the testing and training fold data contain beats from separate subjects. It is proposed that higher accuracies may be achieved if a separate classifier was used for all beat types and an additional algorithm is employed to combine the results.

References

- [1] Senhadji L, Carrault G, Bellanger JJ, Passariello G. Comparing Wavelet Transforms for Recognizing Cardiac Patterns. *IEEE Eng in Med and Biol* 1995;167-173.
- [2] Hu YH, Tompkins WJ, Urrusti JL, Afonso V. Applications of artificial neural networks for ECG signal detection and classification. *J of Electrocardiology* 1993;26:66-73.
- [3] Yeap TH, Johnson F, Rachniowski M. ECG Beat Classification by a Neural Network. *Proc. Annu. Int. Conf. IEEE EMBS* 1990;1457-8.
- [4] MIT-BIH Arrhythmia Database, 3rd Ed. May 1997. <http://ecg.mit.edu/dbinfo.html>
- [5] Friesen GM, Jannett TC, Jadallah MA, Yates SL, et al. A Comparison of the Noise Sensitivity of Nine QRS Detection Algorithms. *IEEE Trans Bio Eng* 1990;37:85-98.
- [6] Nugent CD, Webb JAC, Wright GTH, Black ND. *Electrocardiogram 1: Pre-processing Prior to Classification*. Automedica 1998;16:263-82.
- [7] Kors JA, van Bommel JH. Classification methods for computerised interpretation of the electrocardiogram. *Methods of Information in Medicine* 1990;29:330-6.
- [8] Nugent CD, Webb JAC, Black, et al. *Electrocardiogram 2: Classification*. Automedica 1999;17:281-306.
- [9] Ripley BD. *Pattern Recognition and Neural Networks*. Cambridge, England: Cambridge University Press, 1996.

Address for correspondence:

Maria O' Dwyer.

Dept. of Electronic and Electrical Engineering.

University College Dublin. Belfield, Dublin 4, Ireland.

maria@ee.ucd.ie