Automatic Arrhythmia Classification: A Pattern Recognition Approach

Diana Batista
Master of Science Degree in Biomedical Engineering
Instituto Superior Técnico, Universidade de Lisboa
November 2014

Keywords: Arrhythmia, ECG, Pattern Analysis, Artificial Neural Network, k-Nearest Neighbours, Support Vector Machine

Abstract: Automatic analysis of electrocardiograms (ECGs) has become a major issue due to the large amount of data recorded by numerous cardiac monitoring tools. Particularly, distinguishing between different rhythms is of extreme importance. This work concerns an automatic ECG analysis methodology, validated on benchmark data, and with initial tests already performed on a pervasive ECG acquisition setup. The main focus is on rhythm classification, which was dealt with using a pattern recognition approach. Spectral and time domain features were extracted and the performance of three classifiers was evaluated: k-nearest neighbour (kNN), artificial neural network (ANN) and support vector machine (SVM). When distinguishing between normal sinus rhythm and atrial fibrillation, an overall accuracy of 99.08 % was achieved with the SVM classifier. Classification of eight types of rhythms, divided in five classes, was achieved with a correct classification rate close to 94 % for SVM and kNN. Promising results were obtained from initial experiments carried out with a pervasive ECG acquisition setup.

1 INTRODUCTION

An electrocardiogram (ECG) is a recording of the heart’s electrical activity. This recording can be obtained in a non-invasive manner by placing electrodes on the surface of the chest. The basic components of the ECG waveform are depicted in Figure 1. The RR interval, time period between consecutive R waves, is used to compute the heart rate. Its regularity / irregularity is one of the first steps when analysing an ECG strip.

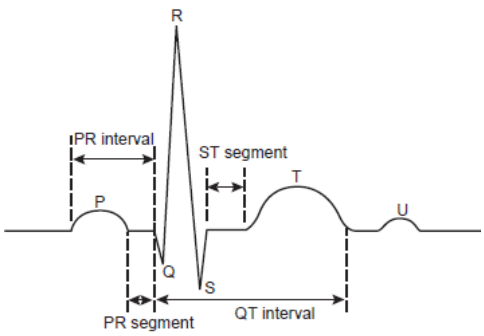

Figure 1: Basic components of the ECG waveform (Huff, 2006).

Besides the standard 12-lead ECG, widely used in clinical practice, a number of other cardiac monitoring tools have been developed in the last few decades. Portable ECG devices that allow the diagnosis of arrhythmias (disturbances in rate, rhythm, or conduction) include Holter monitors, mobile cardiac outpatient telemetry systems, event recorders and patch monitors. An enormous amount of data can be collected by such devices and it is therefore essential to develop algorithms that aid in the analysis of the records.

The problem of automatic detection of arrhythmic events from ECG records has been largely addressed. Many authors have focused on the classification of beat types and, to a lesser extent, rhythm classification has also been attempted. The types of beats / rhythms included and the methodologies adopted vary widely. A truthful comparison of the results is rather difficult since databases used for validation are not always publically available.

In this study a methodology to automatically classify ECG records according to rhythm type is presented. Since the ultimate goal is to perform this analysis on systems with a pervasive acquisition setup, all the steps, from acquisition to classification, are addressed. Algorithm validation is achieved on a benchmark database and initial tests with the BITalino system (www.bitalino.com) are presented.

The remaining of this paper is organized as follows. In section 2 the proposed methodology is presented. In section 3 the results obtained are presented and discussed. Finally, section 4 contains the conclusions.

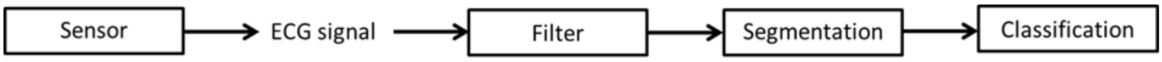

Figure 2: Automatic signal analysis methodology.

2 METHODOLOGY

The methodology proposed in this thesis is schematized in Figure 2, encompassing the following steps: signal acquisition and processing (including filtering and segmentation steps) and finally rhythm classification. These steps are detailed next.

2.1 Signal Acquisition and Processing

2.1.1 BITalino System

The BITalino is a low-cost biosignal acquisition system (Guerreiro *et al*., 2013). It includes sensors for multiple physiological signals: electromyography, electrodermal activity and ECG. It further provides an accelerometer, a light sensor and a light-emitting diode. Software for real time acquisition and visualization is also available (Alves *et al*., 2013).

Since a pervasive acquisition setup is sought, this system was used to acquire ECG records at the fingers. A sampling rate of 1000 Hz was used. It has been shown that the ECGs acquired with the BITalino system at the fingers are highly correlated with lead I from traditional 12 leads systems (Carreiras *et al.*, 2013). It is therefore reasonable to expect that cardiac conditions that can be diagnosed by analyzing the lead I of the ECG can also be detected with this pervasive acquisition system.

2.1.2 Filter

The filter process adopted here has been previously used with ECG data acquired with the BITalino system, mainly for biometric applications. First, two median filters are applied to remove the baseline, with window sizes of 0.2 and 0.3 seconds. A finite impulse response low-pass filter with cutoff frequency of 40 Hz is then used to deal with high frequency noise. Due to the characteristics of the acquisition system, muscular activity is frequently picked up, thus a final moving average filter, with a window of 28 milliseconds, is applied.

2.1.3 R Peak Detection

An algorithm based on the work of (Hamilton, 2002) was used to detect R peaks. A filter step is applied prior to peak detection and a set of decision rules is then followed to classify peaks either as QRS complex or noise. These rules take into consideration peak height, peak location (relative to the last QRS peak) and maximum derivative. The accuracy of the algorithm depends on the computation of a detection threshold, defined using QRS peaks and noise peaks heights.

2.2 Classification

The classification scheme is shown in Figure 3, encompassing the following steps: feature extraction, feature normalization, classifier training and testing. In the following subsections, these steps will be detailed.

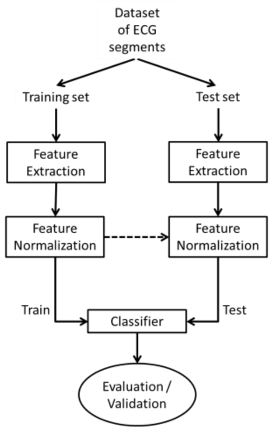

Figure 3: Automatic classification scheme.

2.2.1 Feature Extraction

Two types of features were explored in this analysis: spectral parameters, derived from the wavelet decomposition of the ECG signals; and time domain parameters, translating heart rate characteristics.

Spectral Parameters

Spectral parameters were extracted following the scheme shown in Figure 4. The power spectral density (PSD) of the wavelet decomposition of the signals was computed and two different feature sets were constructed.

Figure 4: Feature extraction process of the spectral parameters.

Signals were decomposed until the sixth level using the quadratic spline wavelet, depicted in Figure 5. Details of this wavelet function and the coefficients of the corresponding finite impulse response filters are given in (Mallat and Zhong, 1992). Figure 6 shows a 10 seconds extract of the decomposition of a normal sinus rhythm ECG. The decomposition was achieved with the redundant discrete wavelet transform (RDWT), or *algorithme à trous* (Fowler, 2005).

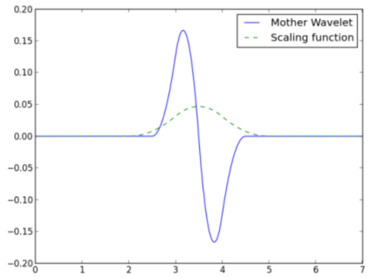

Figure 5: Mother wavelet and scaling function of the quadratic spline wavelet.

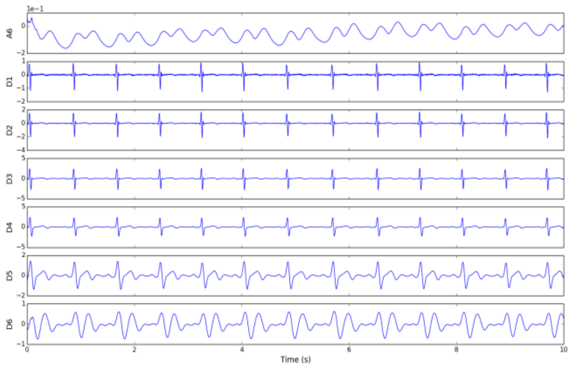

Figure 6: Approximation (A6) and detail (D1 to D6) coefficients of the wavelet decomposition of a 10 s normal sinus rhythm ECG.

For each one of the 7 signals, corresponding to the coefficients of 6 detail and one approximation signals, the PSD was computed. Welch's method, relying upon the concept of modified periodograms, was adopted (Welch, 1967). Segments of 256 samples with 50 % overlap were used, and a Hanning window was employed. Figure 7 depicts the PSD of the signals shown previously.

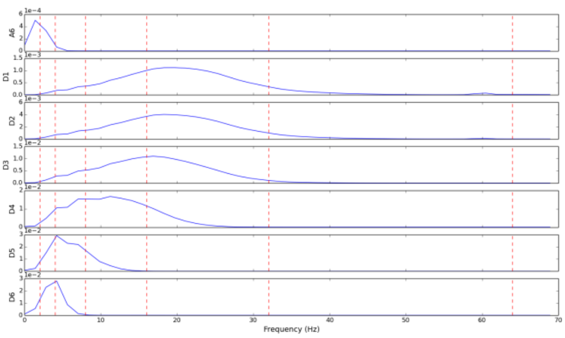

Figure 7: PSD of each one of the approximation (A6) and detail (D1 to D6) wavelet coefficients. Dotted red lines delimitate the sub-bands of feature set A.

Two feature sets of wavelet-based features were extracted:

- For each one of the 7 PSD signals, the average value of the PSD over predefined sub-bands was computed. The 6 sub-bands considered were: [0, 2]; [2, 4]; [4, 8]; [8, 16]; [16, 32]; and [32, 64] Hz. These sub-bands are depicted in Figure 7. This feature set, henceforth referred to as feature set A, contains therefore 42 features (6 values for each one of the 7 signals). The same features are referred by Kara *et al.* (2007).

- For each one of the 7 PSD signals, the integral over the range [0, 55] Hz was calculated. This computation was performed using the trapezoidal rule. A total of 7 features are in this way selected to represent each pattern. This feature set shall be referred to as feature set B.

Time Parameters

To complement the information given by spectral features, two time domain parameters were selected: average RR interval and standard deviation of RR intervals. These ought to be particularly interesting in the distinction of rhythms such as atrial fibrillation and normal sinus rhythm due to the inherent irregularity of atrial fibrillation. Feature set C contains these two parameters.

2.2.2 Feature Normalization

An important step in classification tasks is feature normalization. This can highly influence the classifier's performance. Once the dataset was divided into training and test sets, features from the training set were normalized and the same transformation was then applied to the test set. Two normalization schemes were considered: feature scaling to the range [0, 1] and feature standardization. These operations are detailed in Equations (1) and (2).

$$x_{\text{scaled}} = \frac{x - \min x}{\max x - \min x} \tag{1}$$

$$x_{\text{standardized}} = \frac{x - \mu(x)}{\sigma(x)} \tag{2}$$

2.2.3 Classifiers

The performance of three supervised learning classifiers was assessed: k-nearest neighbour (kNN), multilayer perceptron (MLP) and support vector machine (SVM). The kNN classifier simply assigns to a new pattern the label of the majority of the k closest neighbours. The Euclidean distance was used as a measure of similarity between patterns, and all features were weighted equally. A more detailed explanation of this classifier can be found in (Duda, Hart and Stork, 2001).

MLPs are the most common type of artificial neural networks (ANNs). The network first goes through a learning stage when labelled patterns are presented to it and weights between neurons are adjusted according to the desired output. Further details about MLPs, the backpropagation training algorithm and acceleration techniques can be found in (Beale and Fiesler, 1997). Here, a single hidden layer was used and the most suitable number of neurons in this layer was found for each classification task. The activation function used in this analysis was the logic sigmoid. The learning rate was set to 0.01 and a momentum term of 0.1 was used to accelerate the training phase.

In the last few years SVMs have become increasingly popular on the nonlinear classification of patterns. By using the so-called kernel trick, data is mapped unto a higher dimension where it can be linearly separated (Fletcher, 2009). The radial basis function kernel, given by equation 3 and dependent on the kernel parameter γ, was used in this analysis. The penalty parameter C, which controls the tradeoff between smoothness of the decision boundary and misclassifications, takes also a user-defined value. These parameters were experimentally tuned for best performance.

$$\mathrm{k}(\mathbf{x}_\mathrm{i}, \mathbf{x}_\mathrm{j}) = \mathrm{e}^{-\left(\gamma \|\mathrm{x_i} - \mathrm{x_j}\|^2\right)} \tag{3}$$

2.2.4 Validation Setup

Cross-validation was implemented to test the algorithm. A stratified 4-fold cross-validation was used: for each one of the 4 possible combinations, 3 folds were used as training data and the fourth served as test. This process was repeated for 50 runs, and for each run a balanced dataset was generated by randomly sampling on the existing data.

To evaluate the classifier's performance a couple of accuracy measures were computed. The test error simply refers to the proportion of misclassifications on the entire test set. Inversely, the accuracy represents the proportion of correct classifications.

Using the usual notation for true positives, true negatives, false positives and false negatives (that is TP, TN, FP and FN) we can define the precision (or positive predictive value) and the recall (or sensitivity) as shown in equations (4) and (5). The F_1 score, also known as F-score or F-measure, is the harmonic mean of precision and sensitivity and can be obtained by equation (6).

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{4}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{5}$$

$$\mathrm{F}_1 = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{6}$$

3 RESULTS AND DISCUSSION

3.1 Experimental Setup

The validation of the proposed methodology was performed on benchmark data and ECG acquisitions performed in collaboration with the Hospital de Santa Marta, as described next.

3.1.1 BITalino Acquisitions

ECG recordings were acquired with the BITalino at the Santa Marta Hospital, in Lisbon, during the months of June and July. Subjects were either inpatients or came to the hospital for diagnosis (stress test or tilt table test) or rehabilitation purposes. All subjects gave written informed consent before the acquisition. They were asked to hold one electrode in each hand, between thumb and index finger. Simultaneous recordings were obtained with traditional 12-lead ECG and the BITalino system to allow an accurate ECG interpretation. Since automatic rhythm classification was to be performed, an attempt was made to acquire at least 60 seconds of data with the BITalino system.

A total of 294 records were acquired, corresponding to 180 men with an age of approximately 46.78 ± 30.21 years and 114 women with an age of approximately 46.09 ± 30.65 years.

3.1.2 Benchmark Data

For benchmarking purposes, the MIT-BIH (Massachusetts Institute of Technology – Beth Israel Hospital) arrhythmia database was used (Moody and Mark, 2001). A total of 48 two-channel Holter records are available, each approximately 30 minutes long. The upper signal is usually a modified limb lead II (MLII) but occasionally a modified lead V5. The lower signal is most often a modified lead V1 (occasionally V2 or V5, and in one instance V4). For this analysis only MLII records were used (records number 102 and 104 were therefore excluded). All signals were digitized at a sample rate of 360 Hz. The database includes different sets of annotations verified by more than one cardiologist. All beats are identified and labeled according to their type (i.e. normal beat, premature ventricular contraction…). Annotations that mark the beginning of a rhythm type are also available.

3.2 Results R Peak Detection

In order to evaluate the performance of the algorithm used for R-peak detection, the MIT-BIH arrhythmia database was used, taking advantage of the beat annotations present.

The Waveform Database (WFDB) Software Package, available in (Moody, 2014), was used for validation. For each ECG record, R peaks were detected and an annotation file containing their positions was generated. These positions were then compared to the reference annotations from the database. Standard comparison options were used: the comparison started 5 minutes after the beginning of the record and the match window was set to 0.15 second. Sensitivity and positive predictive values were, respectively, 99.59 and 99.69 %.

Additional tests were carried out to understand if rhythm and beat type influenced the algorithm's performance

Regarding rhythm type, the smallest values of sensitivity and positive predictive values were obtained for ventricular tachycardia. Trigeminy and bigeminy, which are characterized by an alternation of normal and ventricular beats, also presented lower sensitivity values. Finally, the sensitivity for segments of atrial flutter and atrial fibrillation was also below average.

Regarding beat types, it became clear that some beat types are much more easily detected than others. Whilst the error rate was below 1 % for the majority of beats, there were some noteworthy exceptions. The detection of premature ventricular contractions had an error rate close to 3 %, which explains the results obtained previously for ventricular tachycardia, bigeminy and trigeminy. The worst results corresponded however to aberrated atrial premature beats and unclassifiable beats, with error rates of, respectively, 33.33 and 53.33 %.

3.2 Classification on Benchmark Database

3.2.1 Database Assembly

The same database used to validate the R peak detection algorithm was used in this analysis. In order to perform rhythm classification, each record was split into multiple segments according to rhythm annotations. Additional cuts were made in a non-overlapping manner to obtain segments of predefined length. The total number of segments obtained for each type of rhythm and segment length is presented in Table 1. For the features based on the location of the R peaks, the position annotations present on the database were used. This assures that the performance of the algorithm is not affected by eventual mistakes on the detection of the peaks.

Table 1: Number of segments per rhythm type for different segments' length.

Type of rhythm	10 s segments	30 s segments	60 s segments
Atrial bigeminy	7	2	1
Atrial fibrillation	752	226	98
Atrial flutter	60	14	6
Ventricular bigeminy	144	31	7
2° heart block	68	21	9
Idioventricular rhythm	12	3	1
Normal sinus rhythm	6046	1920	911
Nodal rhythm	7	0	0
Paced rhythm	316	99	45
Pre-excitation	36	6	1
Sinus bradycardia	180	60	30
Supraventricular tachyarrhythmia	12	1	0
Ventricular trigeminy	72	12	3
Ventricular flutter	12	3	1
Ventricular tachycardia	8	2	0

3.2.2 Normal vs. Atrial Fibrillation

The first set of experiments focused on the distinction of normal sinus rhythm and the most common arrhythmia, atrial fibrillation**.** Segments with a length of 60 seconds were used. For this length, 98 atrial fibrillation segments and 911 normal sinus rhythm segments are available. Thus, for each run, 98 segments of each class are selected and the validation process is carried out.

Study of Feature Sets

The three feature sets described in the previous section were considered, first individually and then in combination. For each experiment, multiple tests were performed in order to choose the most suitable classifiers' parameters and only the best results are reported. For the ANN this consisted of varying the number of neurons in the hidden layer. For the kNN we varied the number of neighbours considered for classification, k. Regarding the SVM, the penalty parameter, C, and the kernel parameter, γ, were varied in a logarithmic scale, respectively between $[10^{-2}, 10^{8}]$ and $[10^{-5}, 10^{3}]$. For each experiment, the classifiers' parameters that led to the best performance are summarized in Table 2.

The individual performance of each feature set was assessed first. Figure 8 shows the performance, in terms of test error, of the three classifiers, for the three feature sets used. It is clear that time domain features performed better than spectral features. Feature set A, containing average values over sub-bands, achieved better results than feature set B, containing the large range power features. In terms of classifiers, SVM proved to be the most promising choice whilst ANN achieved the worst results for all sets of features.

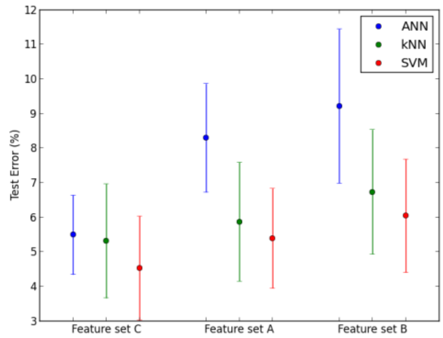

Figure 8: Means and standard deviations of test errors for feature sets C (time parameters), A (average PSD values) and B (large range power values).

Spectral and time domain features were then combined. Referring to Figure 9, it is clear that combining time domain and spectral features considerably improved the classifiers' performance. This was true for the two combinations and for all three classifiers. The best results were obtained when combining feature sets B and C and using the SVM classifier. In this case, the overall accuracy was 99.08 % and F-scores above 99 % were obtained for both classes.

Table 2: Best classifiers' parameters for each classification task.

Feature set	Classifier	Normalization	Parameters	Feature set	Classifier	Normalization	Parameters
A	ANN	Scaling	55 Hidden Neurons	A + C	ANN	Scaling	35 Hidden Neurons
	kNN	Standardization	$k = 1$		kNN	Scaling	$k = 1$
	SVM	Standardization	$C = 10^3 ; \gamma = 10^{-2}$		SVM	Standardization	$C = 10^8 ; \gamma = 10^{-2}$
B	ANN	Standardization	15 Hidden Neurons	B + C	ANN	Standardization	14 Hidden Neurons
	kNN		$k = 1$		kNN		$k = 1$
	SVM		$C = 10^2 ; \gamma = 1$		SVM		$C = 10 ; \gamma = 1$
C	ANN	Standardization	10 Hidden Neurons				
	kNN		$k = 3$				
	SVM		$C = 1 ; \gamma = 10$				

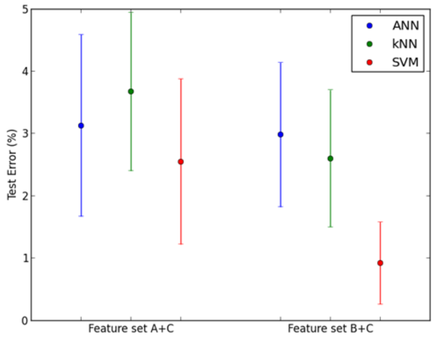

Figure 9: Means and standard deviations of test errors for feature sets A+C and B+C.

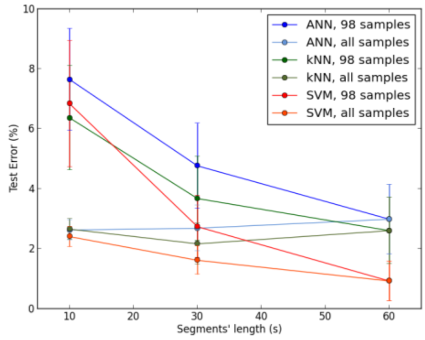

Figure 10: Mean and standard deviation of the test error as a function of the segments' length.

Influence of Segment's Length

The study of features presented above was carried out with ECG segments of 60 seconds. The influence of the segments' length on the performance of the classifiers is explored here. Features used for classification were the ones that held better results: large range power features in combination with time parameters (feature sets B + C). The classifiers' parameters were set to the values obtained previously with this set of features (confer Table 2).

As shown in Table 1 the number of segments of different lengths varies greatly. This fact must be taken into account when performing this analysis; therefore the two following experiments were considered. First, for each classifier, segments' length of 60, 30 and 10 seconds were used while using as many samples per rhythm as available (i.e. 98, 226 and 752 respectively for 60, 30 and 10 seconds). Secondly, the number of samples was kept unchanged, 98, varying only the segments' length. A graphical representation of the test error for both experiments and all classifiers is shown in Figure 10.

It is clear that reducing the segments' length results in a loss of information, which affects the classifiers' performance. Depending on the classifier and the length of the segments, this loss of information is, in part, compensated by a larger number of samples available for training. It therefore seems as if there may be an 'ideal' segment length which will represent a compromise between the information that can be extracted and the total number of segments available. This 'ideal' length will be different for each classifier.

Influence of Segmentation Algorithm

All the experiments carried out until this point used reference R peak annotations, included in the database. This allowed us to evaluate the performances of the classification algorithm in such a way that it was not affected by segmentation errors. It is now important to check the real influence of these R peak detection errors. In order to do so, the classification algorithm was reapplied, using this time R peak annotations given by the segmentation algorithm described in the methodology section. The difference between the results obtained here and the ones obtained previously are illustrated in Figure 11.

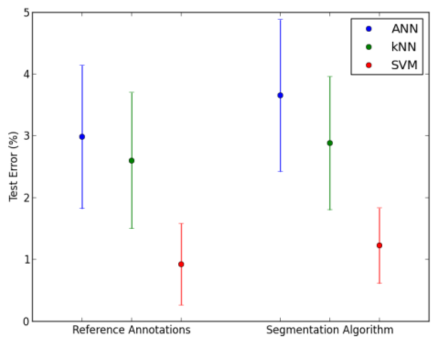

Figure 11: Comparison of rhythm classification error rate when using reference annotations and the segmentation algorithm.

Analyzing the figure above, it is clear that the results of the segmentation algorithm influence the performance of the classification. For all classifiers, the test error increased when including the R peak detection algorithm in the analysis. However, the results are still considerably good and the difference is not large. We can conclude that, although improvements can be achieved, the segmentation algorithm is reliable enough. It should nonetheless be expected that, with noisier data (e.g. acquisitions with the BITalino at the fingers), the segmentation will affect the results in a larger extent.

3.2.3 Multiple Rhythms

The first set of tests focused only on distinguishing normal sinus rhythm from the most common arrhythmia, atrial fibrillation. To test the performance of the algorithm in a more realistic situation, it is necessary to include other types of rhythms in the classification task. The results obtained with multiple rhythms are presented in this section. A few tests were performed maintaining a length of 60 seconds but, in order to have enough samples to include more rhythms, the segments' length was then shortened. The results presented here were obtained using the features that held best results: large range power features in combination with time parameters (feature sets B and C). The classifiers' parameters were set to the corresponding values.

60 Seconds Segments

The following four rhythms were considered: normal sinus rhythm, atrial fibrillation, paced rhythm and sinus bradycardia. A total of 30 samples per rhythm were selected in each run. The results obtained are presented in Table 3.

By including two more rhythms, and consequently decreasing the number of samples used for training and testing the classifier, the performance of the classifiers was affected. Still, the overall accuracy for the SVM classifier was around 97 %. It is interesting to note that precision and recall values for paced rhythm and sinus bradycardia are quite high, but the recognition of normal and atrial fibrillation segments becomes less obvious. This is true for the three classifiers.

30 Seconds Segments

A test meant to deal with bigeminal and trigeminal rhythms was considered next. Because of the similarities between these two arrhythmias, and considering the small number of segments available, the following experiments were carried out. First, using, respectively, 31 and 12 samples per rhythm, bigeminy and trigeminy were considered, individually, along the four rhythms considered previously. Secondly, using 12 samples per rhythm, an attempt was made to distinguish between the six rhythms. Finally, a class that contains segments of trigeminy and bigeminy was constructed, and the distinction between this class and the four others was attempted. A total of 43 samples per rhythm were used in this case. A graphical representation of the test error for these four situations is shown in Figure 12.

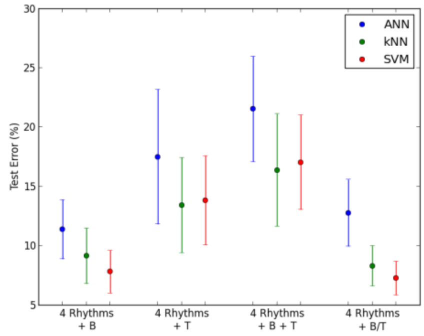

Figure 12: Mean and standard deviation of the test error for different approaches to bigeminal (B) and trigeminal (T) rhythms.

When including a fifth rhythm, bigeminy or trigeminy, the performance of the classifiers worsened. This is understandable since not only is there one additional class to consider, but also the number of training patterns was reduced. As expected, the results are poorer with trigeminy segments than with bigeminy. Certainly, this can again be explained by the smaller number of training patterns (31 versus 12).

Table 3: Results obtained with feature sets B + C in the distinction of four rhythms.

Classifier	**Rhythm**	**Precision (%)**	**Recall (%)**	**F-score (%)**	**Test error (%)**
ANN	Normal	97.16 ± 3.46	82.54 ± 7.65	88.32 ± 5.54	5.71 ± 2.23
	Atrial Fibrillation	86.64 ± 4.95	96.99 ± 3.30	91.09 ± 3.45	
	Paced	98.44 ± 2.10	99.86 ± 1.0	**99.08 ± 1.33**	
	Sinus bradycardia	98.66 ± 2.07	97.79 ± 1.59	98.09 ± 1.51	
kNN	Normal	**97.91 ± 2.83**	82.72 ± 7.28	88.77 ± 5.14	5.18 ± 2.23
	Atrial Fibrillation	87.32 ± 4.92	**98.14 ± 2.92**	91.98 ± 3.48	
	Paced	98.14 ± 2.15	**100.0 ± 0.0**	99.00 ± 1.15	
	Sinus bradycardia	99.48 ± 1.24	**98.46 ± 1.67**	**98.90 ± 1.21**	
SVM	Normal	95.05 ± 2.92	**94.41 ± 3.33**	**94.39 ± 2.50**	**3.24 ± 1.25**
	Atrial Fibrillation	**94.39 ± 3.37**	97.47 ± 2.62	**95.64 ± 2.56**	
	Paced	**99.58 ± 1.05**	98.52 ± 2.15	98.97 ± 1.29	
	Sinus bradycardia	**99.63 ± 1.00**	96.65 ± 0.22	98.02 ± 0.51	

When trying to distinguishing between the six rhythms, the results are quite poor. The kNN classifier performs best but the overall accuracy is barely above 83 %. Even though paced and sinus bradycardia rhythms present precision and recall values above 90 % for all classifiers, the recognition of the remaining rhythms is considerably worst.

By constructing a new class that encompasses bigeminy and trigeminy segments, the results improved. An overall correct classification of almost 93 % was achieved with the SVM classifier. The biggest advantage of combining these two rhythms is that the number of samples available for training increases substantially. It should be mentioned that the two rhythms share similar characteristics (alternation between premature ventricular contractions and sinus beats) and this can justify considering them under the same class.

10 Seconds Segments

With segments of 10 seconds more samples per rhythm are available and it is therefore possible to include more classes in the experiment. Besides the rhythms considered above, second heart block and atrial flutter were included in the analysis. Multiple experiments were carried out to determine the best solution in terms of classes to be considered.

Atrial flutter presents some similarities with atrial fibrillation, namely the fast electrical discharge pattern in the atria. As was the case with bigeminy and trigeminy, it was possible to verify that considering these two rhythms under the same class resulted in higher recognition rates.

An additional experiment was carried out to try to find a better solution for bigeminy and trigeminy segments, which seemed to be affecting the classifiers' performance. The possibility of forming a class with normal, bigeminy and trigeminy segments was explored and the results confirmed the potential of this approach. Therefore, when the occurrence of bigeminal or trigeminal rhythms is of no practical relevance, these rhythms may be included in the normal class instead of considered individually.

The results obtained when considering 5 classes encompassing a total of 8 rhythms are summarized inTable 4.

An overall accuracy close to 94 % was achieved for the kNN and the SVM classifiers. Generally, the kNN classifier reached better recall values for the arrhythmic classes, whereas the precision values for these rhythms were better with the SVM classifier. Depending on the application, it may be more useful to opt for one or the other.

3.3 Classification with BITalino Records

A total of 13 normal sinus rhythm and 13 atrial fibrillation records acquired with the BITalino were selected to perform a few classification tests. These had a length of approximately 60 seconds and were not too seriously affected with artefacts.

Due to the small number of samples used in this experiment, adjustments to the validation setup were made. Instead of a 4-fold cross validation, a leave-one-out approach was adopted. That is, each sample was used once as test sample while the remaining formed the training set.

Table 4: Results obtained with feature sets B + C in the distinction of five classes (68 samples per rhythm).

Classifier	Rhythm	Precision (%)	Recall (%)	F-score (%)	Test error (%)
ANN	Normal + Bigeminy + Trigeminy	87.54 ± 3.44	81.40 ± 3.98	83.96 ± 3.05	7.63 ± 1.33
	Atrial Fibrillation / Flutter	84.57 ± 3.49	**90.85 ± 3.69**	87.27 ± 3.07	
	Paced	97.42 ± 1.49	99.59 ± 0.66	98.45 ± 0.85	
	Sinus bradycardia	95.40 ± 2.25	90.88 ± 2.08	92.85 ± 1.68	
	2nd heart block	99.05 ± 1.02	99.38 ± 0.73	99.20 ± 0.62	
kNN	Normal + Bigeminy + Trigeminy	**91.75 ± 3.15**	82.10 ± 4.38	86.21 ± 3.36	**6.05 ± 1.26**
	Atrial Fibrillation / Flutter	88.16 ± 3.00	90.53 ± 3.71	89.04 ± 2.73	
	Paced	97.24 ± 1.59	**99.74 ± 0.56**	98.44 ± 0.95	
	Sinus bradycardia	94.75 ± 2.41	**97.76 ± 1.51**	**96.12 ± 1.50**	
	2nd heart block	99.17 ± 0.82	**99.85 ± 0.44**	**99.50 ± 0.48**	
SVM	Normal + Bigeminy + Trigeminy	86.10 ± 3.33	**87.39 ± 3.76**	**86.32 ± 2.63**	6.18 ± 1.00
	Atrial Fibrillation / Flutter	**91.18 ± 3.26**	89.35 ± 3.05	**89.90 ± 2.29**	
	Paced	**98.45 ± 1.47**	99.15 ± 0.89	**98.77 ± 0.84**	
	Sinus bradycardia	**95.45 ± 2.37**	94.82 ± 2.62	94.97 ± 1.88	
	2nd heart block	**100.0 ± 0.0**	98.53 ± 0.0	99.24 ± 0.0	

Before the classification step, signals were processed, including filtering and segmentation, as detailed in the methodology section.

The results obtained are summarized in Table 5. The best results were obtained with the ANN classifier, with a test error of 9.23 % and precision and recall values above 90 %.

Although the results were not as good as the ones obtained previously, this experiment proved that it is possible to distinguish atrial fibrillation from normal sinus rhythm ECG records acquired with the BITalino system. With a more thorough analysis, including a fine tuning of the classifiers' parameters, a larger amount of samples, and the inclusion of more rhythms, we should be able to successfully apply the automatic ECG analysis methodology described in this work.

4 CONCLUSIONS AND FUTURE WORK

In this work, a methodology for automatic ECG analysis was proposed, encompassing data acquisition, processing and classification. Automatic arrhythmia classification was attempted using both spectral and time domain features. The performance of three supervised learning classifiers was compared. Validation of the proposed method was performed recurring to benchmarked data from a widely used arrhythmia database. Additionally, initial experiments were carried out to assess the feasibility of classifying ECG records acquired with the BITalino at the fingers, in collaboration with Santa Marta Hospital, in Lisbon.

At first, only normal sinus rhythm and atrial fibrillation ECG records were considered. A number of tests were carried out to determine the best set of features and corresponding classifiers' parameters. It was proven that the combination of spectral and time domain features offered the best results. SVM reached an overall accuracy of 99.08 % with F-scores above 99 % for both classes. An additional test was conducted to try to understand the influence of the segments' length. It was shown that it is important to find a balance between the number of samples available for training and the information contained in each segment.

The inclusion of other rhythms on the classification task was then attempted. In order to do so, and due to the small number of samples per rhythm available, the segments' length was reduced. It was not possible to reach accuracies as high as before but interesting results were obtained and a few conclusions were drawn. In the last experiment, when 5 classes representing a total of 8 rhythms were considered, the kNN classifier reached an overall classification rate close to 94 %.

As the number of classes increased (and the number of samples diminished), the classifiers were less able to distinguish between different rhythms. This is not to say that the recognition of all classes was equally difficult. In fact, rhythms such as paced rhythm or 2nd heart block were easily classified.

As a future development, it may be interesting to develop a step-by-step classification process. That is, to perform the classification in different phases, determining which rhythms ought to be distinguished in each phase. Additionally, it may be worthwhile to adapt the features to the type of rhythms we sought to distinguish. For instance, due to the characteristics of the rhythms, it is possible that distinguishing between normal sinus rhythm and sinus bradycardia could rely solely on the heart rate. When selecting the features to use, one should take into account the amount of time necessary to extract them (in this case, spectral features required a lengthier processing method).

In terms of classifier, SVM and kNN proved to be the most suitable choices. Interestingly, the latter provided higher values of recall for arrhythmias whereas precision values were higher for the first. Again, depending on the application, it may be more useful to opt for one or the other. When making this choice, one should also consider the time available for classification since the length of training and testing phases is different for each classifier.

Finally, the experiments performed with records acquired with the BITalino proved that it is possible to develop algorithms that will automatically analyze and classify ECG records acquired from the fingers with this system. With further developments, more rhythms can be included in the analysis and real time applications can be considered. For instance, the BITalino can be embedded in day-to-day objects and alarms can be triggered when an arrhythmia is present.

Table 5: Results obtained with BITalino records.

Classifier	Rhythm	Precision (%)	Recall (%)	F-score (%)	Test error (%)
ANN	Normal	90.80 ± 5.86	**91.23 ± 5.10**	**90.84 ± 3.86**	**9.23 ± 4.00**
	Atrial Fibrillation	**91.39 ± 4.69**	90.31 ± 6.85	**90.65 ± 4.25**	
kNN	Normal	**91.67**	84.62	88.0	11.54
	Atrial Fibrillation	85.71	**92.31**	88.89	
SVM	Normal	**91.67**	84.62	88.0	11.54
	Atrial Fibrillation	85.71	**92.31**	88.89	

REFERENCES

Alves, A., Silva, H., Lourenço, A. and Fred, A. (2013). BITtalino: A Biosignal Acquisition System based on the Arduino. *BIODEVICES*, pp.261--264.

Beale, R. and Fiesler, E. (1997). *Handbook of neural computation*. Bristol: Institute of Physics Pub.

Carreiras, C., Lourenço, A., Silva, H., and Fred, A. (2013). Comparative study of medical-grade and off- the-person ecg systems. In *Int. Congress on Cardio- vascular Technologies*.

Duda, R., Hart, P. and Stork, D. (2001). *Pattern classification*. New York: Wiley.

Fletcher, T. (2009). Support vector machines explained. Tutorial paper. [online] Available at: http://www.tristanfletcher.co.uk/. [Accessed 5 Nov. 2014]

Fowler, J. (2005). The redundant discrete wavelet transform and additive noise. *Signal Processing Letters, IEEE*, 12(9), pp.629--632.

Guerreiro, J., Martins, R., Silva, H., Lourenço, A. and Fred, A. (2013). BITalino-A Multimodal Platform for Physiological Computing. *ICINCO (1)* pp.500--506.

Hamilton, P. (2002). *Open Source ECG Analysis Software Documentation*. 1st ed. [ebook] E.P. Limited. Available at: http://www.eplimited.com/osea13.pdf [Accessed 26 Oct. 2014].

Huff, J. (2006). *ECG workout*. Ambler, PA: Lippincott Williams & Wilkins.

Kara, S. and Okandan, M. (2007). Atrial fibrillation classification with artificial neural networks. *Pattern Recognition*, 40(11), pp.2967--2973.

Mallat, S. and Zhong, S. (1992). Characterization of signals from multiscale edges. *IEEE Transactions on pattern analysis and machine intelligence*, 14(7), pp.710--732.

Moody, G. (2014). *WFDB Applications Guide*. [online] Physionet.org. Available at: http://www.physionet.org/physiotools/wag/wag.htm [Accessed 29 Oct. 2014].

Moody, G. and Mark, R. (2001). The impact of the MIT-BIH arrhythmia database. *Engineering in Medicine and Biology Magazine, IEEE*, 20(3), pp.45--50.

Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, 15(2), pp.70--73.