Pervasive Electrocardiography and Health Monitoring

Ana Rute Banha
ana.banha55@gmail.com

Instituto Superior Técnico, Lisboa, Portugal

July 2018

Abstract

Heart diseases are one of the most dangerous and present conditions worldwide. The common method for the detection and diagnosis of this type of disease is electrocardiography (ECG), which is a simple, non-invasive and cost-effective technique for diagnosing and monitoring purposes.

This work focused on two main areas: Denoising and outlier removal of ECG heartbeats, and Classification of arrhythmic heartbeats. For the first task several, methods were tested, namely two median filters plus a 40 Hz Low pass filter, wavelet based denoising methods, using several mother wavelets, and a moving average filter. In the present work, the best results were obtained with the technique consisting of using a decomposition level 6 quadratic spline wavelet based denoising which achieved a 0.835 cosine similarity. After an outlier removal step, the result was improved to 0.930.

The classification task consisted in separating a dataset into 5 classes and using the Support Vector Machines (SVM) as classifier. Two databases were used for this task, the MIT-BIH database that was used as a benchmark, achieving an accuracy of 88.70% and the Santa Marta (StM) database, which achieved an 81.06% accuracy. Also, two extra tests were done using this database, one for testing the effect of the best denoising method on the classification task, which improved the results in terms of accuracy and other testing a different classifier, namely the K-nearest neighbors (KNN) classifier.
Keywords: Arrythmic heartbeats, ECG, Heart Rhythm, Support Vector Machine, K-Nearest Neighboor, Wavelet Decomposition, Outlier Removal, Denoising, Classification.

1. Introduction

Heart diseases are one of the most dangerous and present conditions worldwide. The common approach for the detection and diagnosis of this type of disease is based on electrocardiography (ECG), which is a simple, non-invasive and cost-effective technique for diagnosing and monitoring purposes. In this work the main focus is on arrhythmic heartbeats, a disturbance in rate, rhythm or conduction of the electric signal through the heart. While some arrhythmias are harmless, others can compromise the functioning of the heart, being potentially heart threatening, a fact that increases the importance of the development of automated ways to classify the heart activity and get a fast and accurate diagnosis to the patient. While research on heart rhythm and heartbeat automatic classification areas have gained an increased interest, with many new approaches and techniques being proposed every day, still much work has to be done towards increased diagnosis accuracy. In particular, in the area of IoT, with sensors being integrated into everyday objects (such as smart-phones or garnments), analysis of the ECG data collected with these sensors poses additional difficulties, mainly due to the low quality of the acquired signals. These present artifacts and a high level of noise contamination, due to the intermitence and/or poor contact of electrodes with the skin (more notably associated with movements) and contamination with other signals (such as muscle activity). In a step towards pervasive signal acquisition systems, most of the works are based on the placement of electrodes at the hands, wrist or chest level, corresponding to lead I placement according to the standard [1]. Hence the main focus is on diagnosis based on a single lead, mainly lead I.

This thesis focuses on two main objectives: ECG denoising techniques and heartbeat classification from one-lead ECG records.

For the first task of ECG denoising, several methods are tested and compared, namely two median filters plus a 40 Hz low pass filter, wavelet based denoising techniques and moving average filters. Besides these methods, an outlier removal technique is developed and proposed, aiming to remove all the heartbeats that are not relevant for the classification task at hand and the artifacts corrupt the signals. The denoising methods were tested on a database developed at IT - Instituto de Telecomunicaes [2], using a BITalino device [10].

For the classification task, which was considered the final purpose of the thesis, three different sets of features are tested for each experiment: temporal features, morphological wavelet based features and a combination of both. The aim of this task was to develop an algorithm that could separate 5 types of heartbeats: Beat originating in the SA node(N), left bundle branch block beats(L), supraventricular ectopic beats(S), ventricular ectopic beats(V) and fusion beats(F). Firstly, the best set of features for this task was chosen, using the MIT-BIH Arrhythmia Database, which is widely used for heartbeat and rhythm automatic classification experiments. The feature set that showed the best results was used after for the validation dataset, gathered at the Santa Marta Hospital, in Lisbon, the Santa Marta (StM) Database.

2. Background

2.1. Arrhythmia and ECG heartbeat types

Arrhythmia, or dysrhythmia, is a very general term that can refer to all rhythms of the heart other than the normal one. It can represent any disturbance in the rate, regularity, site of origin or conduction of the cardiac impulse. The cardiac impulse starts in the sinoatrial (SA) node but, under some circumstances, cardiac cells in any other part of the heart can serve as a pacemaker of the heart. When this happens, we are in the presence of an ectopic pacemaker, that is, a pacemaker that is any other than the sinus node. The result may be ectopic beats or rhythms. Depending on the location of the pacemaker (natural - SA node - or ectopic), the heart rhythms can be identified as:

- Sinus Node Rhythms, which are due to impulse conduction from the sinus node. This include not only arrhythmia (Sinus arrhythmia) but also the normal sinus rhythm, i.e, the usual cardiac rhythm of the heart;
- Atrial Arrhythmia that originate from ectopic sites in the atria;
- Junctional Arrhythmia and Atrioventricular Blocks originate in the area around the AV node and the bundle of His, called AV junction;
- Ventricular Arrhythmia and Bundle-Branch Blocks that create the impulse in the ventricles, below the bundle of His.

An important step towards identifying an arrhythmia is the classifications of heartbeats, which helps classifying the rhythm of an ECG signal, given by the classification of several consecutive heartbeats in that signal [6].

2.2. Wavelet Transform

Recently, wavelet transform has been widely used in several applications, including the processing of non-stationary signals, feature extraction for classification tasks, heart rate variability analysis, ECG data compression, among others. The wavelet transform is similar to the Discrete Fourier transform. Their main difference is the fact that fourier transform (FT) will tell what frequencies are present in the signal but a wavelet transform (WT) will tell what frequencies are present and where (or at what scale). This transform is a multi-resolution analysis and, consequently, overcomes the difficulty of finding an optimal resolution for the analysis of the signal. It allows the examination of the low frequency content over a larger amount of time without compromising the accurate time domain localization of the high frequencies. Generally, the wavelet transform is expressed by equation 1:

$$F(a,b) = \int_{-\infty}^{\infty} f(x)\psi^*_{a,b}(x)dx \tag{1}$$

where ψ^* is the complex conjugate analyzing wavelet function and a and b are the dilatation and time parameters, respectively. For a function to be considered an analyzing wavelet, or mother wavelet, it should satisfy some mathematical properties, in particular, the function needs to have finite energy and respect the admissibility condition that implies a zero average. A family of functions can be derived from a mother wavelet by applying translations and dilatation. In the last couple of decades, wavelet theory has been widely used for the applications mentioned above, testing several mother wavelets such as Daubechies, Morlet, spline, raised cosine and quadratic spline wavelets. The wavelet transform was used both for the denoising task and for the feature extraction in the classification task.

3. Implementation

This thesis aimed to develop a system for automatic classification of heartbeats. This was done by focusing on two main tasks: denoising and filtering techniques and classification of arrhythmic heartbeats. Three different databases were used in this work, depending on the type of task at hand. This chapter describes, firstly, the types of heartbeats that were to be classified and then the proposed methodology for both the denoising and the classification tasks. The mehodology used for both the denoising and the classification tasks are schematized in figures 1 and 2, respectively.

3.1. Types of Arrythmia

In this work, the main goal as mentioned before, was to develop a system for automatic classification of arrhythmia in ECG recordings. The classes that were used in this task have as a base the As-

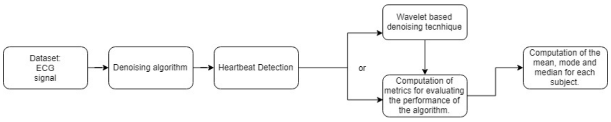

Figure 1: Denoising methodology.

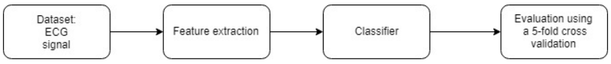

Figure 2: Steps for the classification task.

sociation for the Advancement of Medical Instrumentation (AAMI) [5] recommended practice. This combines the MIT-BIH heartbeat types into only five heartbeat classes, described above in the above section. Class N consists of beats that originate in the sinus node (normal and bundle branch block beat types), class S consists of supraventricular ectopic beats, class V are ventricular ectopic beats, F contains beats that result from fusing normal and V beats and, lastly, class Q contains unknown beats including paced beats. For this work, some changes were made to this practice (Table 1):

- Label for left bundle branch block beat (L) was added to the classification system. The L heartbeat is an arrhythmia type that should also be taken into account because, even if most times it's harmless, it can also be an indication of heart disease. This type also has a similar morphology to the Ventricular ectopic beats (V class) and it's interesting to measure the performance of the system in distinguish these two classes. The right bundle branch block beat (R) was also considered but its classification is not best done using the lead I signals, which is the one that was used in this work so, this type of bundle branch block heartbeat was considered to be a class N heartbeat.

- Since the purpose was to identify pathological beats, the Q label was excluded from all the experiments, since they were mainly unknown and unclassifiable beats, which wasn't of interest for this study.

Table 1: Classes considered for this thesis.

AAMI heartbeat class	**N**	**Q**	**S**	**V**	**F**
Description	Beats originating in the sinus node except for left bundle branch block beat	Unknown Beat	Supraventricular ectopic beat	Ventricular ectopic beat	Fusion beat
MIT-BIH heartbeat types	Normal beat (NOR)	Paced beat (P)	Atrial premature beat (A)	Premature ventricular contraction (V)	Fusion of ventricular and normal beat (F)
	Right bundle branch block beat (RBBB)	Fusion of Paced beat and normal beat (fPN)	Aberrated atrial premature beat (a)	Ventricular escape beat (E)	
	Left bundle branch block beat (LBBB)				
	Atrial escape beat (AE)	Unclassified beat	Nodal (junctional) premature beat (J)		
	Nodal (junctional) escape beat (NE)		Supraventricular premature beat (S)		

3.2. Denoising and Outlier Removal

The main goal in this part of the denoising and outlier removal method, was to take a signal that was corrupted with noise, from now on referred to as the "Noisy Signal" and use several denoising techniques in order to get the cleanest signal possible, measuring the results against a baseline signal with least possible noise added to it, referred to as the "Clean Signal" from this point on. Both this signals correspond to the same time period and have the same heartbeats, acquired by different sensors.

Two approaches for signal denoising are proposed, based on:

1. Data filtering;
2. Wavelet Decomposition.

In 1, the methods described in 4.3.1 and 4.3.3 are included. 2 includes the method described in 4.3.2.

Before all the approaches described below there was a need to remove the baseline wonder from the signals. For that purpose, 2 median filters were applied both to the noisy signal and the clean signal, using window sizes of 0.2 and 0.3 s.

The general steps for evaluating the denoising method were as follows:

1. R peak detection for both signals, that returns 2 arrays of R peaks positions for both signals;
2. compare the locations from both signal's peaks and consider a real peak one that is located in the exact same position on both arrays or with a difference of + or - 20 samples;
3. compute the heartbeat related to each "true" peak. For this, a window of 200 ms left and 400 ms right of the R-peak was used. This fixed-window definition of the heartbeat was used always throughout the developed work;
4. calculate the Signal-to-Noise Ratio - SNR and cosine similarity - Cossim between each computed beat, comparing the beats related to the similar R-peak from both the noisy and the clean signals;
5. for each subject, the mean, mode and median of each measurement are calculated for all the detected beats. This metrics are described in Equations xx,y,z, respectively. Since the mode metric is the most common value on a set of values and the intent of using this metric was to have a better grasp of the most common values in comparison to the mean, before computing the mode, for the cossim, all the values for each subject were approximated up to 2 decimal cases and the SNR to 0 decimal cases.

The denoising methods that were evaluated in this work were the following:

1. 2 Median filters plus a 40 Hz Low-Pass filter (to remove high frequency noise);

2. Wavelet based denoising, using the following mother wavelets:

   - Function = "Sym12" and level = 3 which, according to [4] gives the best performance of denoising for both universal and interval-dependent thresholds;
   - Function = "Db4" and level = 5, as in [3];
   - Function = "Quadratic spline" and level 6, as in [16].

   Every wavelet was tested both applying them directly to the noisy signal or combining it with the 40 Hz Low Pass filtering;

3. Moving Average;

4. Outlier removal step combined with the denoising method that showed the best results.

3.2.1 Outlier removal

Besides noise contamination, signals may present artifacts (acquisition / connectivity failures; ...) that lead to too corrupted segments, that will be considered as outliers that should be removed from the signals to achieve better results. An outlier removal tecnhique was used during this thesis, and consisted in the development of beat prototypes. After calculating the RR peaks and computing the heartbeat using the same fixed window as described above, the outlier removal was done, comparing each detected beat to each of the prototypes and which considered a sample as an outlier if it's cosine similarity to every prototype was less than 0.8, a value that was chosen in order to have better results without losing a big part of the signal. Since, as mentioned above, only 5 type of beats are of interest for this experiment, 50 prototypes were developed (10 for each label) using a labeled database. In order to obtain prototypes for each of the classes (pathological or normal), the corresponding available labeled data was first segmented into individual heart beats, as explained before and K-means clustering was applied over these heartbeats. The value of K was empirically set to 10, the number of clusters per class. Given the dependence of the K-means on the initialization, this algorithm was run 20 times, selecting as final solution the partition corresponding to the smaller mean square error, and the corresponding centroids were chosen as class prototypes.

3.3. Feature Extraction

For data representation, we explored 2 types of features,temporal features and morphological wavelet based features, described next. These features were chosen based on [7] and [13].

3.3.1 Temporal Features

Four temporal features were extracted from the RR-intervals (distance between 2 consecutive R peaks) of the pre-processed ECG signals. The features used are described below:

- Pre-RR-interval: Distance from a given heartbeat to the previous one;
- Pos-RR-interval: Distance between a given heartbeat and the following one;
- Average RR-interval: The mean of all the RR-intervals from a recording of a subject;
- Local average RR-interval: Mean of the RR-intervals computed over 10 heartbeats centered at the given point.

3.3.2 Morphological Features

Wavelet-based morphological features were also used in the task of classifying ECG beats, which were extracted from each ECG cardiac cycle by selecting the above mentioned window of -200ms to +400ms around the R-peaks. The choice of the mother wavelet and the decomposition level for this work was based on [7].So, the Daubechies wavelet of order 2 (Db2) was chosen due to it's similar morphological structure with the heartbeats. This wavelet can be seen in Figure 3.

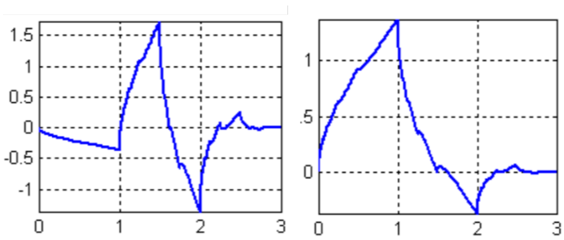

Figure 3: Mother wavelet (right) and scaling function (left) of the Db2 wavelet

Also, a level of decomposition of 4 was chosen, which means that the heartbeat signal was decomposed into the detail coefficients D1-D4 and one approximation coefficient A4.

The wavelet coefficients represent the distribution of the signals energy in time and frequency. From these coefficients, the following statistical values were extracted and used as features:

Table 2: Experiments specification for Classification Task

Experiment	DATABASE	Labels used	Specification
1	MIT-BIH	N, S, V, F	This first experiment was made to evaluate the algorithm and the 3 different combinations of features for the detection of the location where a certain heartbeat is originated. Its based on the AAMI recommended practice, taking out the Q type, for the reasons mentioned above.
2	MIT-BIH	N, L, S, V, F	In the second experiment we distinguish the L type from the N type mentioned above, for it is also an important beat to take into account. This label set, as mentioned above, is the one that we aim to optimize and validate after, on the StM Database.
3	MIT-BIH	L, V	In order to verify if the L and V were being properly identified, since they have a similar morphology, this extra test was made.
4	StM	N, L, S, V, F	Using the combination of both time and morphological features, this experiment was done to validate the results from point 2 in a different database.
5	StM	N, L, S, V, F	Experiment 4 is then repeated but using a different mother wavelet, namely the quadratic spline wavelet, with a level of decomposition = 6.
6	StM	N, L, S, V, F	From experiment 4 and 5, the best set of features was chosen and retested after applying the best denoising method found from the denoising experiments, in order to see the effect of denoising on heartbeat classification.
7	StM	N, L, S, V, F	Same setup was used but the KNN algorithm was tested instead of the SVM.

- Maximum value for each of the wavelet coefficients;
- Mean value for each of the wavelet coefficients;
- Minimum value for each of the wavelet coefficients;
- Standard deviation value for each of the wavelet coefficients;

Therefore, 20 morphological wavelet based features were extracted from the ECG heartbeat.

Besides the Db2 wavelet, an extra technique was implemented, using quadratic spline level 6 mother wavelet to see if it improved the classification results.

3.4. Classification

Two classifiers were used for the classification task, described below: k-nearest neighbor and support vector machine. These classifiers were implemented using Python programming language and the scikit-learn Python module, a machine learning kit developed for this language [9].

The experiments specifications made for the classification task on both databases are specified in Table 2. A stratified 5-fold cross-validation was used to test the implemented algorithms for both databases, where for each one of the five possible combinations, 4 folds were used as training data and one was used as test data. Due to the unbalancing of the number of label samples for the databases, a balancing step was performed before running the algorithm, using as criteria for the number of samples from each label the minimum common denominator, i.e, the value that matched the label with the smallest number of samples. In the case of the StM database, since the F label only had 33 samples, the second smallest value was chosen for the other labels, keeping the 33 F samples. This is going to be further discussed in the Results section.

The algorithms and feature sets were tested using the MIT-BIH database, which served as benchmark data, and were validated on the StM one. To evaluate the performance of the classifier, accuracy, precision, recall and F_1-score metrics were used. Using the typical notation for true positives (T_P), true negatives (T_N), false positives (F_P) and false negatives (F_N), these metrics can be defined by Equations 2, 3, 4 and 5, where the F_1-score is the harmonic mean of precision and recall.

$$Accuracy = \frac{T_P + T_N}{N} \tag{2}$$

$$Precision = \frac{T_P}{T_P + F_P} \tag{3}$$

$$Recall = \frac{T_P}{T_P + F_N} \tag{4}$$

$$F_1 = \frac{2T_P}{2T_P + F_N + F_P} \tag{5}$$

4. Experimental Setup and Results

4.1. Experimental Setup

The proposed methodology was validated using three different databases that are described next.

4.1.1 BITalino Bicycle (BB) Database

This database was gathered for the study of the influence of physical effort on a bicycle with the integration of a BITalino device [11] in the ECG signal morphology. It had 2 goals:

1. Evaluate the performance of the system and algorithms used for biometric recognition when there are changes in the heart rate (HR) and pattern of the ECG signal;
2. Evaluate the use of heart rate variability (HRV) as a biometric trait.

It collected a number of different biological signals, from ECG signals acquired three different ways, to blood volume pulse data and respiration data, among others. In this work, this database was used to test the performance of different denoising methods on an ECG signal. So, only ECG signals were used. This dataset had recordings for ECG signals gathered three different ways, but only the two below were used for this work.

- ECG signal collected on the chest, using a chest band, with electrodes placed on each side of the sternum and the fifth or sixth intercostal spaces. This was used as the baseline clean signal, i.e., the signal we want to get at the end of the denoising experiments.
- ECG signal acquired at the hands, using a stationary bicycle with a BITalino device coupled to it. Due to amount of movement, sweat, among other things, this signal is highly corrupted by noise and was the signal that was aimed to be denoised in order to get it as close to the baseline clean signal described above.

For both of the used recordings, a sampling rate of 1000 Hz was used. The protocol used for this recordings is specified in [14]. Twenty five subjects were recorded and studied in the present work.

4.1.2 MIT-BIH Arrhythmia (MIT-BIH) Database

The MIT-BIH Arrhythmia database was built from a set of over 4000 long-term Holter ECG recordings, gathered by the Beth Israel Hospital Arrhythmia Laboratory between 1975 and 1979. It contains 48 recordings of slightly over 30 minutes long, where 23 of them are randomly chosen from the set and 25 are selected from the same set to include a variety of rare but clinically important cases that would not be well represented by a small random sample of the Holter recordings [15]. The first 23 records serve as a sample of the variety of waveforms and artifact that an arrhythmia detector might encounter in a routine clinical use. The last 25 records were selected to include complex ventricular, junctional, and supraventricular arrhythmia and conduction abnormalities.

The analog outputs were filtered to limit analog-to-digital converter (ADC) saturation and for anti-aliasing, using a pass-band filter from 0.1 to 100 Hz. This filtered signals were then digitized at a 360 Hz sampling frequency.

More specific characteristics, recording and annotation details can be found in [8].

This database is used in the classification experiments since it is one of the most used databases for this type of tasks and serves as a base for comparison of the algorithms implemented in this thesis and also to compare to the results from the Santa Marta database, described below.

4.1.3 Santa Marta (StM) Database

The ECG signal acquisition was done by the clinicians of the Santa Marta Hospital in Lisbon, using a Philips Page Writer Trim III recorder, with a sampling frequency of 500 Hz.

This database was used for designing prototypes for the outlier removal task that's going to be described bellow and for validating the feature set chosen for this thesis, from the benchmark data (MIT-BIH database). In total, 11429 recordings were used for these experiments, a much higher amount than the number of recordings for the benchmark data.

5. Results for Denoising Techniques

The denoising experiments were done using the BB database. Due to problems during the recording of the signals, some subjects missed the beginning of the signals, either in the noisy or in the clean version of the signal, which compromised the beat detection because of the fact that the algorithm that was used, based on the work of [12], makes an initial estimate for the detection by detecting maximum peaks in eight consecutive 1-second intervals. Because of this, 3 subjects were taken off the denoising experiment. These subjects, namely subjects 7,18 and 19 are displayed in Figure 4.

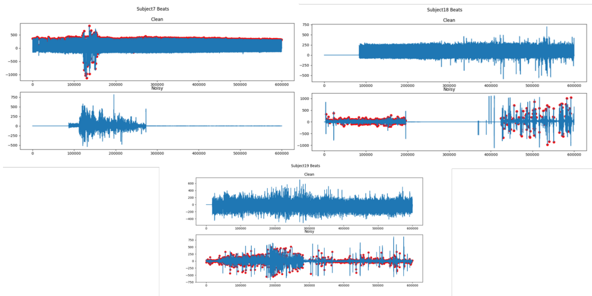

Figure 4: Subjects that were removed from denoising experiment.

From all the denoising experients made before the outlier removal, the one that presented the best results, in terms of all the metrics used, was the quadratic spline wavelet based denoising technique. The results are presented in Table 3 for the 4 subjects that presented the higher results after only the 40 Hz low-pass filtering step.

Table 3: SNR and CosSim for the previously selected subjects before and after wavelet denoising technique, Quadratic Spline, is applied to the heartbeats after the 40 Hz low pass filtering to the noisy signal.

		2 Median Filters 40 Hz LP Filter			**Quadractic Spline Level 3**		
	SUBJECT	**MEAN**	**MODE**	**MEDIAN**	**MEAN**	**MODE**	**MEDIAN**
SNR	**1**	6,904	8,000	7,055	6,951	8,000	7,179
	11	3,604	5,000	3,941	3,616	5,000	3,960
	16	4,593	7,000	5,482	4,568	7,000	5,301
	20	4,010	5,000	4,196	4,066	5,000	4,229
CosSim	**1**	0,834	0,970	0,900	0,835	0,970	0,903
	11	0,752	0,890	0,848	0,752	0,900	0,850
	16	0,794	0,930	0,895	0,794	0,930	0,894
	20	0,717	0,880	0,794	0,721	0,910	0,797

Afterwards, an outlier removal method was applied to the signals, combined with this wavelet based method, to measure the improvement to the signal denoising results. Figures 5 and 6 show the improvement in terms of SNR and CosSim mean for all the subjects that presented a CosSim mean value higher than.

As it can be seen in the figures, practically most subjects present an improvement on both the CosSim and SNR metrics. This shows the importance of an outlier removal step and that the developed prototype method can be a promising method to further developed.

5.1. Classification Results

The classification task were done firstly in the benchmark data, namely the MIT-BIH database,

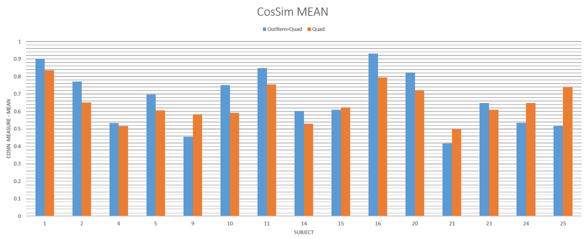

Figure 5: CosSim mean values ofr quadratic spline wavelet based technique with and without outlier removal for subjects whose CosSim mean was higher then 0.5.

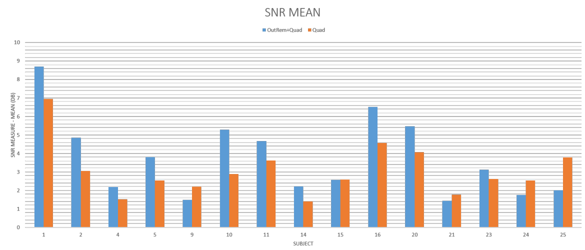

Figure 6: SNR mean values ofr quadratic spline wavelet based technique with and without outlier removal for subjects whose CosSim mean was higher then 0.5.

being validated after in the StM Database, where a different set of features was also tested. The experiments made are described in **??** and their results are analyzed in this section.

To begin the classification task of the thesis, a feature extraction process needed to be done, using the features described in the above section.

The main task of this work was the classification of heartbeats in the validation setup, using the StM database. Even though the number of samples for each label isn't that big between this dataset and the previous one, the number of different recordings is much higher, which represents a bigger variability in the data both from subject, placement or situation. This variability can result in worst values for the metrics used in this thesis and to the classifier working on recognizing, for example, the subject for the heartbeat in question instead of the type of heartbeat, independently of external conditions. Secondly, a different mother wavelet was tested, namely the Quadratic Spline wavelet with decomposition level 6. After that, two more experiments were done, one using the outlier removal plus best denoising method from the ones studied in the previous experiments and another trying a different classifier, KNN, on the experimental setup that showed the best results for this dataset. In all the experiments a balancing step was performed for the dataset, where the second smallest value for the number of samples was used, due to the fact that the smallest number of samples was for class "F", that had only 33 samples. This is a number too small to be able to train and test an algorithm. So during these classification tasks, there were only 33 "F" samples while the rest of the classes had 1028 samples (or 31 "F" samples and 681 for the rest of the classes in the outlier removal experiment). For this reason, one last test was made, using the method from experiment 1, the one that had the feature set that gave the best results, in other to see the effect of this unbalancing on the experiments results, using both the SVM and the KNN classifiers.

5.1.1 Classification using SVM Classifier

After the balancing step, 1028 random samples from classes "N", "L", "S" and "V" and the 33 samples for the "F" class were used to train and test the algorithm using a 5 fold cross validation using the combination of time and wavelet based features, the feature set that had the best results for Experiment 2 using the MIT-BIH database. The results for this experiment are showed in Table 4.

As it can be seen, the accuracy values still were above 80% for this setup but decreased in relation to the results obtained in experiment 2. Also, while for the MIT-BIH database the precision, recall and $F_1 - score$ values are very close to the accuracy values, being also above 80%, using the StM Database these 3 metrics decrease significantly. These results can be due to the 33 samples for the "F" label, that may be leading to a large number of false positives for the other labels. To confirm that this is affecting the results, a test was made using only the samples for the "N, "L", "S" and "V" classes and the same algorithm. The results are showed in Figure 7.

Even though the improvement in the accuracy values aren't highly significant, there is a high increase in the other metrics, which shows the importance of the balancing of the dataset, especially if one of the labels has a smaller number of samples as in this case. Without the F-class which has only 33 samples, the number of false positives decreases and the number of true positives increases, leading to these higher precision, recall and F_1 values.

The fact that the values obtained for this dataset are lower than the ones for the MIT-BIH one can be explained by the much higher number of recordings used, which has the advantage of having a higher number of data for training and testing the algorithms but has the disadvantage of having a bigger variability among the recordings as explained above.

After these results were obtained, the same set of features were tested after a denoising and outlier removal step.

Table 4: Results for classification of the 5 labels using the Db2 wavelet with a decomposition level 4.

Feature set	C/gamma Parameters	Accuracy (%)	Precision (%)	Recall (%)	F_1
Db2, level=4	10/0.00001	81.06%	74.23%	67.04%	67.94%

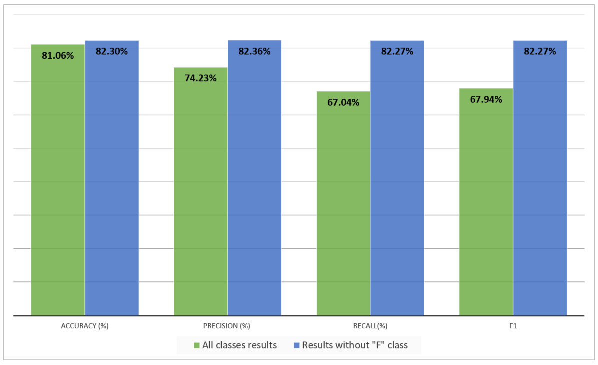

Figure 7: Comparison between the results for the classification of the 5 different classes(green) and taking the "F" class beats from the dataset (blue).

In the denoising task of this thesis, the method that achieved the best results was the one that used the outlier removal technique plus the quadratic spline wavelet based denoising technique. To investigate the effects of these techniques on the classification task, the same features were extracted from the dataset after a denoising and outlier removal step. After performing the outlier removal on the StM dataset, the number of samples in each class decreased.

Since the second smaller number of samples in a class now is 681, the balancing step was done so that each class had this number of random samples, from the dataset after outlier removal, except for the "F" class which had 31 samples, the only available in the dataset. After this balancing step, the wavelet based denoising method was applied to each heartbeat before extracting the wavelet based features. The results for this experiment are shown in table 5.

Table 5: Results for classification of the 5 labels after an outlier removal and denoiing steps.

Feature set	C/gamma Parameters	Accuracy (%)	Precision (%)	Recall(%)	F_1 (%)
Db2, level = 4 + Outlier Removal + Denoising	10/0.00001	80.56%	75.21%	67.39%	68.45%

The main purpose of this experiment was to see if the algorithm performed better after denoising and outlier removal than without these steps. As it can be seen, the accuracy value is lower then in the previous experiment (the same one without denoising and outlier removal) but the other values are slightly better. This can be explained by the decrease in the number of samples from one experiment to another. Since there is less data to train the algorithm, the accuracy is affected but since the difference between the number of samples in the "F" class and the other classes is smaller, the precision, recall and F_1 values are higher. To better see the results of using the denoising methods, experiment 4 was redone but this time using a balanced dataset that had the same number of samples for each class as the ones for the current experiment, i.e., 681 random samples from the dataset with no outlier removal step for the "N", "L", "S", and "V" classes and 31 for the "F" one. The results are presented in Figure 8.

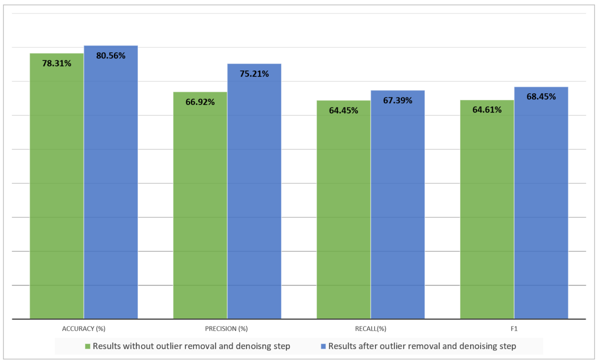

Figure 8: Comparison between the results for the classification with 681 samples per class balancing (and 31 for the "F" class) (green) and for classification after outlier removal plus quadratic spline wavelet based denoising step.

5.1.2 Classification using KNN Classifier

To also test the performance using a different classifier, this last experiment was made using a KNN algorithm. This experiment used the same setup as Experiment 4, the only difference being the classification algorithm used. The results are shown if Table 6 and show worst results (Accuracy = 77.83%) than the ones that used the SVM classifier (Accuracy = 81.06%). The precision, recall and F_1 values are again lower then the accuracy ones, probably because of the same reason as before, that is, the under sampling of the "F" class. In Figure 9, the results using only 4 classes, without the "F" are showed and compared to the ones with all 5 classes. From this figure, it can be seen that taking out the "F" class performs better than using it, giving higher precision, recall and F_1 values, probably due to the decrease in false positives and increase of true positives.

From the experiments' results, it can be seen that the algorithm that performed better was the SVM one, for both the StM and the MIT-BIH databases.

Table 6: Results for classification of the 5 labels using the Db2 wavelet with a decomposition level 4 and the KNN classifier.

Feature set	**K**	**Accuracy (%)**	**Precision (%)**	**Recall(%)**	F_1
DB2, level = 4	1	77.83%	64.40%	64.42%	64.35%

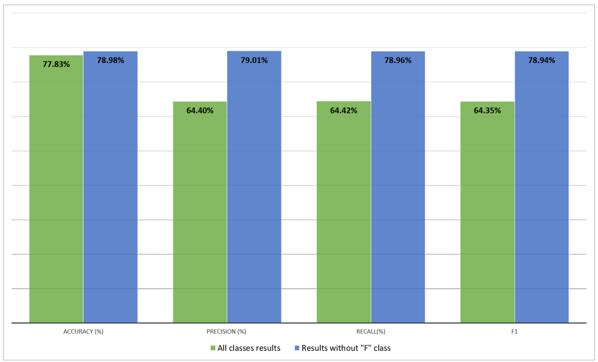

Figure 9: Comparison between the results for the classification of the 5 different classes(green) and taking the "F" class beats from the dataset (blue), using the KNN classifier.

Also, the testing of both wavelet features showed that the Db2, a it said in the literature, were in fact best for this type of task. Another important result is the problem with ill balanced datasets for these algorithms, showed when the task was done with and without the F type heartbeats, showing an increase in all 4 metrics after removing this label. One last important result was the improvment of the classification task after using the best denoising method tested in the denoising section (outlier removal plus quadratic spline wavelet based denoising), showing the importance of a denoising step before starting classification tasks using ECG signals.

6. Conclusions

Medical and biomedical advances are continuously being made in the field of cardiac monitoring. This leads to an increase of collected data that needs to be processed and analyzed. Automatic classification and denoising algorithms can help in this task. Denoising methods have been developed in order to improve the quality of ECG signals, facilitating their readability and analysis. Classification algorithms have been studied to help and automatize the recognition of arrhythmia and pathological heartbeats. In this work, several methodologies for the denoising of ECG signals were studied, as well as automatic classification of heartbeats algorithms, using both time and morphological features.

The denoising part of the work studied 3 different types of methods: the baseline one consisted in 2 median filters for baseline wander removal plus a 40 Hz low-pass filter, which achieved good results, namely a cosine similarity up to 0.834. After this, three wavelet based denoising techniques were tested, using the Sym12, Db4 and Quadratic Spline wavelets, The quadratic spline wavelet was tested and achieved a slightly better result then the baseline median filters plus 40 Hz low pass filter, achieving a cosine similarity of up to 0.835. Even thought the improvement wasn't very significant, this method still performed slightly better for every subject. Lastly, using denoising method that showed the best results, with the quadratic spline mother wavelet, an outlier removal step was added to the denoising process, in order to test its effect on the final results. After developing the prototypes, the heartbeats detected by the R-peak detector used were considered outliers if their cosine similarity to all the prototypes was bellow 0.8. After this, the quadratic spline method was applied to each heartbeat and the results were measured, achieving values up to 0.930 for the cosine similarity, a very significant improvement when compared to all other previous methods.

The second part of the work consisted in developing a feature set and an algorithm for automatic classification of heartbeats and test them against the database collected by the IT group. T For this, firstly the different feature sets were tested in the benchmark data, the MIT-BIH database, widely used in the studies of arrhythmia and heartbeat classification. The main focus of this part of the work was to classify the heartbeats into these 5 labels: "N", "L", "S", "V" and "F".

The results for the mains focus of the classification task, i.e., classification of the 5 different types of heartbeats mentioned before, were validated using the StM database. For this dataset, a balancing first step was also performed but, due to the fact that the "F" class, which was the minimum common denominator, only had 33 samples while the second smallest class had 1028, the balancing was done for the 1028 samples per class, except for the"F" class that was represented only by those 33 samples. This resulted in an accuracy of 81.06% and a precision of 74.23%.

To evaluate the effects of denoising on the dataset, an experiment that used the best method from the previous task was done. First, an outlier removal step was done on the dataset, removing all the heartbeats that had a cosine similarity lower than 0.8 to the developed prototypes. Next, the balancing step was done the same ways as before for the StM database, seeing as the minimum common denominator continued to be the "F" class, with 31 samples after the outlier removal. So the balancing was done choosing 681 samples from each of the other classes and using the 31 "F" ones. This resulted in a 80.56% accuracy. Since the number of sables decreased, to be able to compare the results using this denoising method and without it, the same balancing was done to the dataset with no outlier removal and the features were extracted without the denoising method. This showed results of 78.31% for accuracy, 2.25% less, which proves the importance of removing the outliers and denoising the signals before using the classification algorithms.

Since there are several classification algorithms, another one was tested for this task to classify between the 5 classes, namely KNN classifier. The results for this test showed an accuracy of 77.83%, lower than fo the SVM ones. The precision was also lower than this value (64.40%) so an extra test was made for this classifier taking out the "F" class, reaching a precision value of 79.01% and an accuracy of 78.98%. Still these values were smaller than the SVM results, which makes SVM a better choice for this type of classification tasks.

There is still a lot of room for improvement in this field, with different features to be explored and different classifiers. As a future development, these experiments could be done taking into account other ECG leads besides the Lead I used for this thesis, that also give important information about the signals and can help in classifying the heartbeats into more than just this five labels. Other features could also be used that gather more morphological information from the signal. The time features already prove to have good results for classification so, if combined with the right set of morphological features, they can lead to better results than the ones presented in this thesis. Also, one improvement that can be made is the use of other classifiers. Recently, deep neural networks have been showing promising results for this type of tasks and can be a good improvement to this methodologies, despite their high computational cost.

References

[1] ECG standards. https://thephysiologist.org/study-materials/the-ecg-leads-polarity-and-einthovens-triangle/.

[2] IT - Instituto de Telecomunicações. https://www.it.pt/ITSites. Accessed: 2017-10-15.

[3] M. Alfaouri and K. Daqrouq. ECG signal denoising by wavelet transform thresholding. *American Journal of Applied Sciences*, 5(3):276–281, 2008.

[4] M. AlMahamdy and H. B. Riley. Performance study of different denoising methods for ECG signals. *Procedia Computer Science*, 37(Supplement C):325 – 332, 2014.

[5] Association for the Advancement of Medical Instrumentation. Testing and reporting performance results of cardiac rhythm and st segment. *ANSI/AAMI*, 1998.

[6] P. de Chazal and R. Reilly. A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval feature. *IEEE Trans. Biomed. Eng.*, Vol. 53:2535–2543, 2006.

[7] F. A. Elhaj, N. Salim, A. R. Harris, T. T. Swee, and T. Ahmed. Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals. *Computer Methods and Programs in Biomedicine*, 127(Supplement C):52 – 63, 2016.

[8] A. G. et al. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation Electronic Pages*, Circulation 101(23):e215–e220, 2000.

[9] F. Pedregosa, et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[10] J. Guerreiro, R. Martins, H. Plcido da Silva, A. Lourenco, and A. Fred. Bitalino: A multimodal platform for physiological computing, 07 2013.

[11] H. Silva, A. Fred and R. Martins. Bitalino - biosignals for everyone. *IEEE Pervasive Computing*, 13(4):64–71, 2014.

[12] P. Hamilton. Open source ECG analysis. In *Computers in Cardiology*, pages 101–104, 2002.

[13] O. T. Inan, L. Giovangrandi, and T. A. Kovacs. Robust neural network based classification of premature ventricular contraction using wavelet transform and time interval features. *IEEE Trans. Biomed. Eng.*, Vol. 53:2507–2515, 2006.

[14] M. Martinho. Protocol for the study of the influence of physical effort on a bicycle with the integration of a bitalino device in the ECG signal morphology. *Report/Protocol for IT Experiments*, 2017.

[15] M. R. Moody GB. The impact of the MIT-BIH arrhythmia database. *IEEE Eng in Med and Biol*, 20(3):45–50, 2001.

[16] P. Ranjith, P. Baby, and P. Joseph. ECG analysis using wavelet transform: application to myocardial ischemia detection. *ITBM-RBM*, 24(1):44 – 47, 2003.