UNIVERSITY OF OULU

FACULTY OF INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING

Marcela Tobón Cardona

AUTOMATIC DETECTION OF EARLY REPOLARIZATION PATTERN IN ECG SIGNALS WITH WAVEFORM PROTOTYPE-BASED LEARNING

Master's Thesis
Degree Programme in Biomedical Engineering
August 2018

Tobón Cardona M. (2018) Automatic detection of Early repolarization pattern in ECG signals with prototype-based learning. Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland. Master's thesis, 52 p.

ABSTRACT

Early repolarization (ER) pattern was considered a benign finding until 2008, when it was associated with sudden cardiac arrest (SCA). Since then, the interest in the medical community on the topic has grown, stating the need to develop methods to detect the pattern and analyze the risk of SCA. This thesis presents an automatic detection method of ER using supervised classification. The novelty of the method lies in the features used to construct the classification models. The features consist of prototypes that are composed by fragments of the ECG signal where the ER pattern is located. Three different classifiers models were included and compared: linear discriminant analysis (LDA), k-nearest neighbor (KNN) algorithm, and support vector machine (SVM). The method was tested in a dataset of 5676 subjects, manually labeled by an experienced analyst who followed the medical guidelines.

The algorithm for the detection of ER is composed of different stages. First, the ECG signals are processed to locate characteristic points and remove unwanted noise. Then, the features are extracted from the signals and the classifiers are trained. Finally, the results are fused and the detection of ER is evaluated.

Accuracies of the different classifiers showed results over 90%, demonstrating the discrimitative power of the features between ECG signals with and without the ER pattern. Additionally, dimensionality reduction of the features was implemented with Isomap and generalized regression neural networks (GRNN) without affecting the performance of the method. Moreover, analysis of critical cases that are difficult to label was performed based on the distances to the classifier decision boundary, improving the sensitivity of the detection. Hence, the method presented here could be used to discriminate between ECG signals with and without the ER pattern.

Keywords: ECG, Early Repolarization, Slur, Notch, Waveform, Prototypes, Features, Classifiers, KNN, LDA, SVM, Isomap, GRNN, Boundary

TABLE OF CONTENTS

ABSTRACT

TABLE OF CONTENTS

FOREWORD

ABBREVIATIONS

1. Introduction **6**

2. ECG and Early Repolarization Pattern **8**
2.1. Anatomy and Physiology of the Heart 8
2.2. Basics of ECG 8
2.3. Early Repolarization Pattern 10
2.3.1. Automatic detection methods of ER 11

3. Supervised learning **13**
3.1. Supervised classification 13
3.1.1. K-nearest neighbors 13
3.1.2. Support vector machines 13
3.1.3. Linear discriminant analysis 14
3.2. Comparison of classifiers 15
3.3. Features dimensionality reduction 16

4. Automatic detection of ER **19**
4.1. ECG preprocessing 19
4.1.1. ECG resampling 20
4.1.2. Baseline offset removal 20
4.1.3. R-peak detection 21
4.1.4. S-waves and low-SNR filter 21
4.2. Waveform prototype feature extraction 22
4.3. Lead-based classification 24
4.4. Classifier fusion 25
4.5. Classifier validation 25
4.5.1. Data splitting for cross validation 25
4.5.2. Balancing of the data 26
4.6. Borderline cases assessment 26
4.7. Features dimensionality reduction 28

5. Results **30**
5.1. Study data 30
5.2. Acquisition of the features 32
5.3. Lead-based classification 34
5.4. Classifier fusion and validation 34
5.5. Comparison of the classifiers 36

5.6. Borderline cases assessment 38
5.7. Dimensionality reduction 41
5.8. Comparison with existing automatic method for detection of ER . . . 42

6. Discussion **44**

7. Conclusion **48**

8. REFERENCES **49**

FOREWORD

This thesis represents to me something more than an academic accomplishment. It represents the whole journey of coming to Finland and meeting new people. It symbolizes how much I learned and developed my skills since the start of my studies. It is a reflect of my interest and passion for research and developing technology that can help doctors and patients.

In this thesis I present a method to automatically detect the early repolarization pattern in electrocardiographic signals. It combines my previous knowledge on signal processing with my acquired understanding of supervised learning systems. I hope this work will encourage people to use and develop health-care technology for the benefit of the patients.

It was a very enriching journey for me to get the guidance of my supervisor Professor Tapio Seppänen who was willing to dedicate his time and share his knowledge since the beginning of my studies. Also, my gratitude to Dr. Tuomas Kenttä for his expertise and guidance during this process. My wonderful family is also represented in this thesis with their support, inspiration and love no matter the distance between us.

Hereby, I would like to express my gratitude and appreciation to my examiners, advisers and supervisor.

Oulu, June 25th, 2018.

University supervisor: Dr. Tech. Prof. Tapio Seppänen.

Second examiner: PhD. Tuomas Kenttä.

ABBREVIATIONS

ECG	Electrocardiography
SA	Sinoatrial
AV	Atrioventricular
ERP	Early Repolarization Pattern
ER	Early Repolarization with terminal QRS slur/notch
SCA	Sudden Cardiac Arrest
SCD	Sudden Cardiac Death
IVF	Idiopathic Ventricular Fibrillation
SNR	Signal to Noise Ratio
KNN	K-Nearest Neighbors
SVM	Support Vector Machines
LDA	Linear Discriminant Analysis
ISOMAP	Isometric Feature Mapping
GRNN	Generalized Regression Neural Networks
TPR	True Positive Rate
SPC	Specificity
TP	True Positives
TN	True Negatives
FP	False Positives
FN	False Negatives

1. INTRODUCTION

In recent years, there has been a growing interest on the Early Repolarization Pattern (ERP) in electrocardiographic (ECG) signals. Controversy exist on the term ERP because it was first used to describe ST-segment elevation in ECG signals, and later to refer to terminal QRS slurring or notching with or without the presence of ST-segment elevation. Moreover, ERP caught the attention of the medical community when it was associated with sudden cardiac arrest (SCA) and idiopathic ventricular fibrillation (IVF) [1], [2]. Although ERP is a common electrocardiographic finding, IVF is a rare disease. On the other hand, SCA is often the presenting symptom in IVF [3]. Hence, the recognition of ERP has importance for specialists, general cardiologists and physicians to distinguish patients at risk for potentially lethal ventricular arrhythmias [2], [3].In this work, automatic detection of the pattern, focused in subjects with QRS slurring or notching (ER), is proposed for the identification of subjects at risk of SCA.

Methods to automatically detect the ER pattern had been proposed before [4],[5], [6]. However, they are based on the measurement and analysis of the QRS downslope. In this study, ER is automatically detected by using the signal itself, instead of the morphological measurements. Research in cognition has shown that humans categorize objects based on hierarchical comparison using a generic prototype [7]. Therefore, in this approach, a fragment of the ECG signal is then used as a reference or prototype. Comparison of new data with the prototype-based features allows to perform the automatic detection of the ER. In this way, the pattern is compared with several prototypes allowing to adapt to user-specific changes on the signal more easily.

Supervised classification methods had been used before in medical applications with ECG signals. The pattern recognition systems generally implemented consist of components such as preprocessing, feature extraction and classification. In the preprocessing stage, unwanted noise of the signals is removed and some transformations can be applied, to reduce the influence of external patterns different than the target. Then, the features from the signals are extracted and usually consist on physiological measurements from the ECG signal. In this study, the boundary between preprocessing and feature extraction is somewhat arbitrary because there is no $feature\ extractor$ whose purpose is to reduce the data by measuring certain $features$. Our feature extraction process consist on selecting the fragment of the signal that it is going to be used. Lastly, in the classification stage, the classifiers are trained to evaluate the presence of the pattern in terms of the features. Some pattern recognition systems also implement a post-processing stage to decide on the final recommended action.

The implicit hypothesis to build a supervised classifier model is that individuals belonging to the same class have similar features, thus the multidimensional feature space is separated in regions according to the classes [8]. Nevertheless, one of the problems with classification is the overlapping data. When the data is overlapping in the feature space, the classifier can fail to detect the correct class. Consequently, this study proposes a method to evaluate critical cases on the detection of the ER, based on the distance to the separation boundary in the feature space.

The basic structure of this work is the following. Chapter 2, presents the literature review on the ECG and ER pattern. Following that, Chapter 3, contains the technological literature review of the supervised classifiers used, how to compare them and optimize them with dimensionality reduced features. Chapter 4 introduces the meth-

ods used for signal processing, extraction of the prototype-based features, classification and the post-processing methods to fused them, compared them and assess their classification as well as the overlapped cases. The fifth chapter presents the results that were achieved in the detection of the ER. Chapter 6, contains the discussion on the results obtained and future work. Finally. the last chapter presents a short conclusion on the study.

2. ECG AND EARLY REPOLARIZATION PATTERN

2.1. Anatomy and Physiology of the Heart

The heart can be considered as two separate pumps (right and left) that work independently of each other. The right side of the heart contains two chambers, the right atrium and right ventricle. The left side of the heart is also divided in two chambers, the left atrium and left ventricle. The atria receive the blood and pump it into the ventricles which then contract and pump it into the vessels.[9]

The heart muscle (myocardium) is responsible to rhythmically contract and drive the circulation of blood throughout the body. The myocardium is triggered by a wave of electrical current that passes through the entire heart and produces the contraction of the muscle, also know as systole. Diastole occurs when the myocardium relaxes and the heart fills with blood to prepare for the next contraction. [10]

The cardiac cycle comprises the relaxation and contraction of both the atria and ventricles. The atria begin to contract (atrial systole) after the atrial depolarization to release the blood into the ventricles. Then, the ventricles begin to contract (ventricular systole) which rises the pressure and allows the blood to move into the lungs and the rest of the body. Following ventricular repolarization, the ventricles begin to relax (ventricular diastole) and the pressure drops which allows the blood to flow back into the atria and repeat the cardiac cycle.[11]

Each heartbeat is triggered by an action potential which originates from a rhythmic pacemaker. Normally, the impulse comes from the sino-atrial (SA) node and spreads across the atria to the atrioventricular node (AV). It reaches the ventricles by passing through the Bundle of His and terminate in the Purkinje fibers, which spread the impulse to the ventricular myocardium and cause the contraction. [12],[13]

This results in a measurable change in potential difference on the body surface of the subject. The resultant amplified and filtered signal is known as an electrocardiogram and represents the heart's electrical activity.

2.2. Basics of ECG

An electrocardiogram (ECG) is obtained by measuring electrical potential between various points of the body. A lead records the electrical signal of the heart between one positive and one negative pole. Each lead represents a view of the heart's electrical activity. A total of 12 leads, or views, are represented on the modern ECG. The first six leads represent the frontal plane (or view) of the heart. They are called the limb leads and are named I, II, III, aVR, aVL, and aVF. The next six leads represent the horizontal plane (or view) of the heart. They are called precordial (in front of the heart) leads, and are named V1,V2, V3, V4, V5,and V6.[9]

Conventionally, the ECG leads are also divided accordingly to the views of the four walls of the left ventricle. Leads II, III and aVF are know as inferior leads because they correspond to the inferior wall. Leads I, aVL, V5 and V6 to the lateral wall; leads V1 and V2 to the septal wall and leads V3 and V4 to the anterior wall. [14]

Each lead generates a characteristic waveform according to the direction of the electrical current. However, there is a common pattern that can be observed in the ECG signal and is characterized by three main waves (Figure 1)[12]:

- The P wave represents the spread of the impulse from the SA node across the atria (often referred to as atrial depolarization).
- The QRS complex represents the spread of the impulse to the ventricles through the Purkinje fibers(ventricular depolarization). The QRS complex generates the largest deflection in the ECG signal because the ventricles contains the largest mass of myocardium. in order to have the strength to pump the blood through the whole body. Because the Purkinje fibers are located in the endocardium (inner layer), the impulse spreads from endocardium to epicardium (outer layer) [13].
- The T wave represents ventricular recovery (often referred to as repolarization).

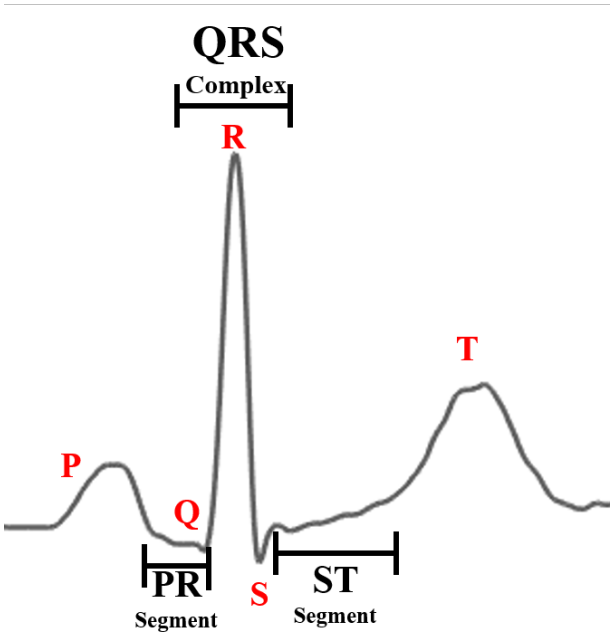

Figure 1. ECG signal morphology [15].

In addition to the main three waves, the normal ECG signal also presents the following isoelectric lines:[13]

- The PR segment represents the delay that the AV node generates between the conduction from the atria to the ventricles, and the spread of the impulse in the His-Purkinje system. Therefore, it does not generate any electrical signal and appears as an isoelectric or flat line after the P wave.
- The QRS complex is followed by a flat line known as the ST segment. The J point (J junction) represents the end of the QRS and the beginning of the ST segment.

When the ECG signal is in regular rhythm, the interval between P waves is somehow constant and there is a QRS following each P wave at a somehow constant time, it is called a sinus rhythm [9]. If the rhythm is not regular, the intervals between two R peaks are too big or two small, or the typical ECG morphology is lost; it is a called an arrhythmia.

2.3. Early Repolarization Pattern

The definition of the Early Repolarization Pattern (ERP) has been a controversy over the last years. The term ERP was originally used to refer to ST-segment elevation without chest pain, but it was later used to describe terminal QRS slurring or notching without the presence of ST-segment elevation. In 2016, the American Heart Association (AHA) made an statement to clarify the confusion and proposed to include those two different phenomenon in the ERP term [3]. Figure 2 shows the examples of ERP with terminal QRS slur/notch, with and without ST segment elevation.

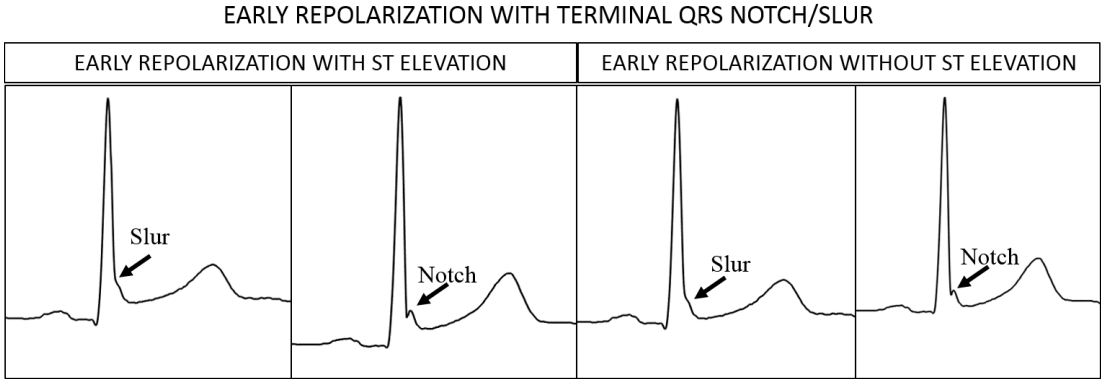

Figure 2. Examples of Early Repolarization pattern.

Early repolarization with terminal QRS slur/notch (ER) was defined by Haïssaguerre et al. as QRS slurring, when there is a smooth transition from the QRS to the ST segment; and QRS notching, when there is a positive J deflection inscribed on the S wave [1]. The American Heart Association proposed to define terminal QRS notch as a low-frequency deflection at the end of the QRS complex, and terminal QRS slur as an abrupt change in the slope of the last deflection at the end of the QRS[3]. Another definition of ER was published in [2] to delineate the electrocardiographic measures of the pattern. According to the definition, ER is present if:

1. An end-QRS notch is a notch that occurs on the final 50% of the downslope of an R-wave in the final segment of the QRS complex, and links with the ST-segment.

2. An end-QRS slur is an apparent slowing of the inscription of the waveform in the final 50% of the R-wave downslope of the QRS complex that merges with the ST-segment of the complex.

3. The pattern lies above the baseline

4. J peak (Jp) is $\geq$ 0.1 mV in 2 or more contiguous leads of the 12-lead ECG, excluding leads V1 to V3

5. QRS duration is < 120 ms.

The underlying biological mechanisms and pathogenesis of the ERP are also an area of controversy to whether the electrocardiographic findings represent early repolarization, late depolarization, or neither [3]. Repolarization is not a propagating phenomenon, and because the duration of the action impulse is much shorter at the epicardium than at the endocardium, the activity appears as if it were propagating from epicardium towards the endocardium [16]. If the epicardial action potential starts "early", then notched QRS complex may arise. Not everyone agrees with this explanation and it has been suggested that the pattern may be due to late ventricular depolarization or genetics [3][17].

ERP was considered a benign finding until 2008 when it was associated with Sudden Cardiac Arrest (SCA) and idiopathic ventricular fibrillation (IVF)[1][2]. Sudden cardiac death (SCD) generally refers to an unexpected death from a cardiovascular cause, whereas SCA describes SCD cases in which the individual survived the cardiac arrest [18]. SCA remains a major public health problem where only 3 to 10% of patients who have an out-of-hospital cardiac arrest are successfully resuscitated [1]. SCD is the manifestation of a fatal heart rhythm disorder due to an unexpected cardiac arrhythmia leading to a hemodynamic collapse [19]. SCA occurs in persons without structural heart disease or with well-recognized electrocardiographic abnormalities that affect ventricular repolarization (e.g., long or short QT intervals or the Brugada syndrome). However, it can also occur in other cases in which there are no signs during sinus rhythm and are described as IVF [1].

Haïssaguerre and colleagues suggested an increased prevalence of ER in subjects with IVF [3],[1]. Another study also revealed that inferolateral ER was more frequent in 45 patients with IVF. Additional studies have demonstrated an association between ERP and arrhythmia in patients with known structural heart disease. A study in Finland found that inferolateral ER was present in 5.8% of the population and was associated with an increased risk of SCD, particularly in those with inferior ER $\geq$ 2 mm.[3]

Therefore, the recognition and correct diagnosis of the ECG pattern of ER has importance for specialists, general cardiologists and physicians[2]. The detection of ER is based on the analysis of the QRS morphology and the measurement of Jp. Automatic methods to detect ER pattern had been proposed and are characterized by the analysis and quantification of the notch or slur slope [4],[5],[6]. In contrast, electrocardiographic machines implemented commonly in clinical practice use interpretation algorithms to detect ST-segment elevation in addition to elevation of the J point[20], but do not include the ER notch/slur analysis.

2.3.1. Automatic detection methods of ER

One of the automatic methods to detect ER is to calculate a tangent from the peak of the R-wave through the following downslope and retain the details of the slope [5],[2]. Other technique available, is to automatically measured various points on the notches and slurs. Combined with the logic that detects a delta wave, by looking for

slope changes in early QRS, and effectively mirrored for the down-slope. It finds a slur at the end of the QRS complex or a notch, according to the slope variations [6]. A more recent algorithm was based on quantification of the characteristic slurring or notching slope. A terminal slope is a line fitted using the yield point, that describes the slurring of the J point (Figure 3), and classifies each lead as notched, discrete, slurred, indeterminate (cannot be classified) or negative [4].

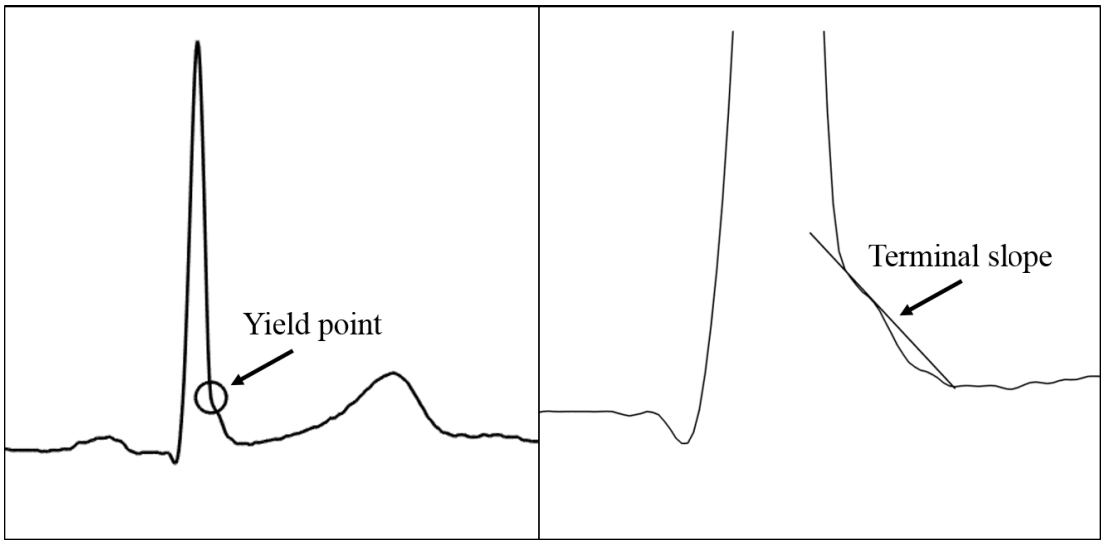

Figure 3. Automatic detection of ER based on the quantification of the steepest slope, the yield point and terminal slope.

A notch was detected if the yield point amplitude was >0 mV, notch amplitude was $\geq$ 0.09 mV, and the peak of the notch occurred within a timeframe defined by the yield point and QRS offset [4]. A lead was considered slurred if no S wave was present, yield point was $\geq$ 0.1 mV and terminal slope was greater than a predefined angle [4].

3. SUPERVISED LEARNING

The goal of supervised learning is to build a concise model of the distribution of the class labels, in terms of the predictor features, that can be used to assign new labels to the testing data [21]. In medical applications, specifically using ECG signals, the aim of supervised learning is to assign each individual heartbeat to its specific class or to detect a patient's pathology using the information contained in the features [8].

3.1. Supervised classification

The implicit hypothesis to build a supervised classifier model is that individuals belonging to the same class have similar features, thus the multidimensional feature space is separated in regions according to the classes [8]. Several methods exist to perform classification like neural networks, Bayesian classifiers, linear discriminant analysis (LDA), decision trees, k-nearest neighbors (KNN), support vector machines (SVM), etc. In this work, three types of supervised classifiers were selected for the detection of ER: KNN, SVM and LDA.

3.1.1. K-nearest neighbors

The KNN algorithm is based on the idea that feature vectors will generally exist in close proximity of predicted features of same properties [22]. Therefore, the label of the testing sample is assigned accordingly to the most common label of the surrounding training samples. KNN uses a distance metric in order to find the surrounding k samples of the testing data. Ideally, the distance metric should minimize the distance between similar classes and maximize it for different classes [22]. Different distance metrics can been used like euclidean distance, cityblock metric, mahalanobis distance, among others.

KNN is a very simple and robust algorithm which proves to be very effective when the number of samples in the training dataset is large [23]. KNN has the disadvantages that it requires large storage space, it is sensitive to the distance metric that is chosen and there is no a principled way to choose k, more that trial and error with cross validation [22].

3.1.2. Support vector machines

SVM are one of newest and the most widely used supervised classifiers [23], [22]. They have good generalization capabilities and, with opportune transformation, can deal also with nonlinear problems [22]. In SVM, a hyperplane is computed to separate the classes in the high dimensional space using the support vectors. Those vectors are defined by the data points that maximize the margin between the classes, and are found by comparing the distances between data samples and the separation hyperplane (Figure 4). The fact that only the support vectors are used to train the SVM, gives the

advantage that the complexity of the model is not affected by the amount of subjects in the feature vector.

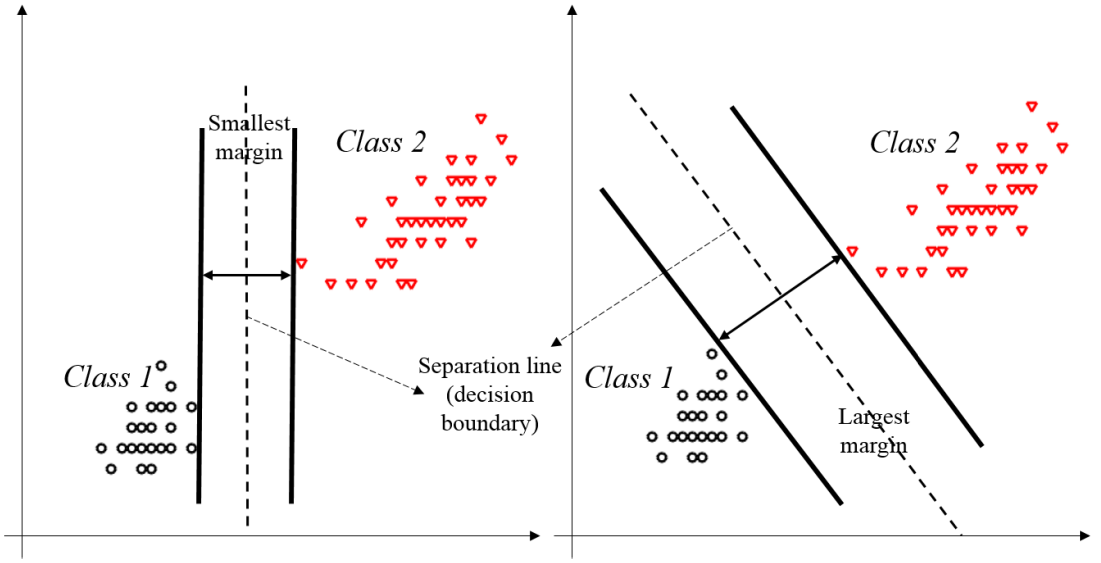

Figure 4. Two separating decision boundaries: a good one with a large margin (right) and a less acceptable (left).

The hyperplane is optimal when the data is linearly separable. However, when the data is no linearly separable, the samples are transformed to a higher dimensional feature space where they can be separated. To transform the data, the kernel trick is used to compute the inner product of the data samples, instead of calculating the coordinates in the high dimensional feature space. Different types of kernels can be used but the most common are gaussian radial basis function (rbf) and polynomial. In order to be able to find the hyperplane, SVM usually uses a soft margin that allows classification errors but that helps the algorithm to converge. It also reduces the risk of overfitting. Similar to the KNN classifier, SVM is sensitive to the kernel function that is selected and its parameters.[22]

The equation used to train the SVM classifier with gaussian kernel is:

$$f(x) = \sum_{i=1}^{n} a_i y_i G(x_i, x) + b \tag{1}$$

where $(a1, \ldots, a_n, b)$ are the estimated SVM parameters and $G(x_i, x)$ is the dot product in the Gaussian predictor space between x and the support vectors .

3.1.3. Linear discriminant analysis

LDA finds the projection hyperplane that minimizes the interclass variance and maximizes the distance between the projected means of the classes[24]. It is the oldest classifier inspired in the one devised by Sir R. Fisher [25]. Fisher's LDA builds linear discriminant functions, that estimate discriminant scores and discriminant weights, so

that the ratio of the variance within the classes is minimal and the score is maximized [26].

The model for discriminant analysis assumes that the data has a Gaussian mixture distribution and uses a multivariate normal distribution. For linear discriminant analysis, the model has the same covariance matrix for each class and only the means vary. In practical terms, LDA uses the posterior probabilities and the cost of misclassification of the observations to minimize the expected classification cost by:

$$\widehat{y} = \underset{y=1,\ldots,K}{\arg\min} \sum_{k=1}^{K} \widehat{P}(k|x)C(y|k) \tag{2}$$

where $\widehat{y}$ is the predicted classification, K is the number of classes, $\widehat{P}(k|x)$ is the posterior probability of class k for observation x, and $C(y|k)$ is the cost of classifying an observation as y when its true class is k [27].

3.2. Comparison of classifiers

One of the goals of this study is to compare the performance of the ER detection using the three mentioned supervised classifiers. In 2006, Demšar published a review on the statistical tests used for comparisons of algorithms, and then theoretically and empirically examined them [28]. The author recommended the Wilcoxon signed ranks test, for comparison of two classifiers, and the Friedman test with the corresponding post-hoc tests, for comparison of more classifiers over multiple data sets.

The most-common statistical test to compare two data sets is the paired t-test, which checks if the average difference in the data sets is significantly different from zero. However, the t-test suffers from three weaknesses: I) t-test only makes sense when the differences over the data sets are commensurate, II) the paired t-test requires that the differences between the two random variables compared are distributed normally and III) the t-test is affected by outliers which skew the test statistics and decrease the test's power by increasing the estimated standard error. [28]

Nevertheless, t-test is been widely used to compare classifiers and in some cases even with more than two datasets by conducting several paired t-tests, which makes little sense. In those cases, the most-known statistical method, for comparison of multiple datasets, is used: the repeated-measures ANOVA. Unfortunately, ANOVA is based on assumptions which are most probably violated when analyzing the performance of machine learning algorithms. ANOVA assumes that the samples are drawn from normal distributions and satisfied the sphericity property(the random variables have equal variance), which cannot be taken for granted. [28]

For those reasons, Demšar proposed to used the non-parametric equivalent of the repeated-measures ANOVA, the Friedman test, to compare the performance of multiple classifiers. The Friedman test ranks the different algorithms on each dataset according to the average error rates, giving rank 1 to the one with the smallest error. The null-hypothesis is that there is no difference between the algorithms so their average ranks should be the same. The Friedman statistic is calculated according to the following equations [29]:

$$\chi_F^2 = \frac{12N}{K(K+1)}\left[\sum_j R_j^2 - \frac{K(K+1)^2}{4}\right] \tag{3}$$

$$F_F = \frac{(N-1)\chi_F^2}{N(K-1)-\chi_F^2} \tag{4}$$

where N is the number of datasets, K is the number of algorithms and R_j is the rank of algorithm j.

If the null-hypothesis is rejected, a post-hoc test should be performed. Different post-hoc test could be used including the Nemenyi test, the Bonferroni-Dunn test, Hommel's procedure, Hochberg's method, Holm test, Shaffer, among others [28], [30]. The classical methods are Nemenyi's and Holm's procedures, however they are not the more powerful ones [30]. The Bergmann-Hommel procedure is the most powerful one but it requires intensive computation in comparisons involving numerous classifiers [30]. Shaffer's procedure, can be used instead in these cases. Bergmann-Hommel's procedure reject all H_j with j $\notin A$, where the $acceptance\ set$

$$A = \cup\{I : Iexhaustive, \min\{P_i : i \in I\} > \alpha/|I|\} \tag{5}$$

is the index set of null hypotheses which are retained. A valid algorithm obtains all the exhaustive sets of hypotheses (E), using as input a list of classifiers C. Thus, E contains all the possible exhaustive sets of hypotheses for a certain comparison. [30]

3.3. Features dimensionality reduction

In order to optimize the time computation of the supervised classification, dimensionality reduction of the features was considered in this study. Dimensionality reduction of the feature vector is the mapping of the data to a lower dimensional space such that the unnecessary variance of the data is removed [31]. Reduction on the dimension of the features vector has been used to allow data visualization, extract key low-dimensional features, improve the model performance, reduce time/complexity computation, among others. Many methods exist for dimensionality reduction: principal components analysis (PCA), independent component analysis (ICA), canonical correlation analysis (CCA), manifold learning, multidimensional scaling (MDS), etc. [31].

PCA and MDS represent the classical methods guaranteed to discover the true structure of data lying near a linear subspace of the original high-dimensional space [32]. These methods are simple and efficient but can fail to find low-dimensional embedding of some data structures. PCA finds a low-dimensional embedding of the data points that best preserves their variance. Classical MDS finds an embedding that preserves the interpoint distances. However, many data sets contain essential nonlinear structures that are invisible to PCA and MDS [33].

Isometric feature mapping (Isomap) is a nonlinear manifold learning technique that preserves the intrinsic geometry of the data, while reducing the dimensionality of the

feature vector [33]. Isomap uses the geodesic manifold distances between the data points. The algorithm can be divided into three steps [32]:

1. Construction of neighborhood graph. Two data points are considered neighbors if their distance in the original space is shorter than a constant or one of the points belongs to the k nearest neighbors of the other point. A neighborhood weighted graph is constructed with this information, where the weights are equal to the distance of the points in the original space

2. Computation of shortest paths. The shortest path distances in the neighborhood graph are computed to find the geodesic distance on the manifold between all pairs of points.

3. Construction of d-dimensional embedding. The data can be represented with a matrix expressing the geodesic distance of each pair of points on the manifold. Applying classical MDS to this matrix constructs an embedding of the data that best preserves the manifold's estimated intrinsic geometry

One of the main disadvantages of feature transformation is that the physiological meaning of the original feature is typically lost in the transformation [8]. Another disadvantage of using Isomap, is that it does not provide the mapping to embed the original data into the lower dimension, which would be required for classification and visualization of new samples. This problem has been solved using generalized regression neural networks (GRNN) [32].

GRNN were proposed in 1991 by Donald F. Specht as a method to provide estimates of continuous variables. It converges to the underlying (linear or nonlinear) regression surface where assumption of linearity is not justified. GRNN create a memory-based neural network with a highly parallel structure [34]. A GRNN does not require an iterative training procedure as back propagation networks and it approximates any arbitrary function directly from the training data. If the training set size becomes large, the estimation error approaches zero [35].

GRNN falls into the category of probabilistic neural networks so it needs only a fraction of the training samples than a backpropagation neural network would need [34]. Generalized regression means that the output of the system is not assumed to satisfy any particular function of the input, e.g. linear function as in linear regression, but rather it can have any particular functional form and requires no prior knowledge of the appropriate form. This is accomplished by using a probability density function (pdf) that is empirically determined from the observed data using ParZen window estimation [34].

The regression estimation in GRNN is implemented in a parallel, neural-network-like structure that "learns" and begin to generalize immediately, without the need of iterations [34]. The pdf function used in GRNN is the Normal Distribution and each training sample is used as the mean of a Normal Distribution [36]. GRNN stands in the following equations:

$$Y(x) = \frac{\sum Y_i \exp\left(\frac{-d_i^2}{2\sigma^2}\right)}{\sum \exp\left(\frac{-d_i^2}{2\sigma^2}\right)} \tag{6}$$

$$d_i^2 = (X - X_i)^T \cdot (X - X_i) \tag{7}$$

where x is the input sample, X_i the training sample and Y_i the prediction of the training sample. The Euclidean distance, d_i^2, between the training sample and the prediction is used to measure how well each training sample can represent the position of prediction [36]. $\exp(-d_i^2/2\sigma^2)$, is the activation function or weighted function for the training sample.

The network contains four layers: input layer, pattern layer, summation layer and output layer (Figure 5). The input layer feeds the samples to the pattern neurons which contain the activation function (6). The summation layer has two different parts: S summation neuron, computes the sum of weighted responses of the pattern layer; and D summation neuron, is used to calculate un-weighted outputs of pattern neurons. The summation and output layer together perform a normalization of output set. The output layer predicts the output Y by dividing the output of each S-summation neuron by that of each D-summation neuron.[35]

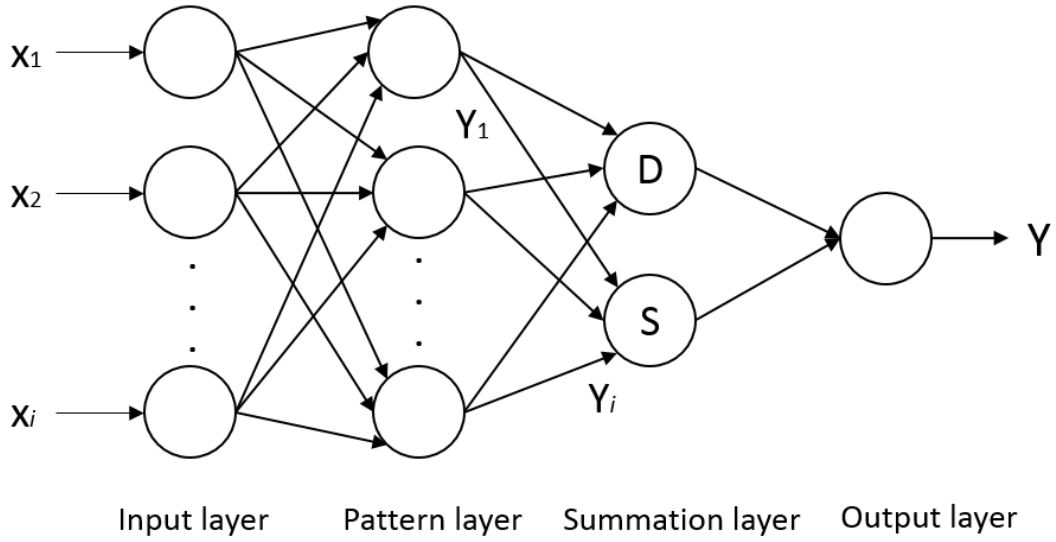

Figure 5. GRNN block diagram

4. AUTOMATIC DETECTION OF ER

In this chapter, the automatic system for the detection of ER is described. Figure 6 represents the general block diagram of the algorithm. The first step is to preprocess the ECG signals to remove noise and unwanted wave-shapes. Then, the waveform prototypes are extracted and used for the lead-based classification of ER. Lastly, the results on the classification of each lead are combined to obtain the final decision on the ER detection and validate the classification. The final decision was evaluated to find subjects that were difficult to classify, accordingly to their distance to the decision boundary of the models. Furthermore, dimensionality reduction of the features was considered and it is discussed at the end of the chapter.

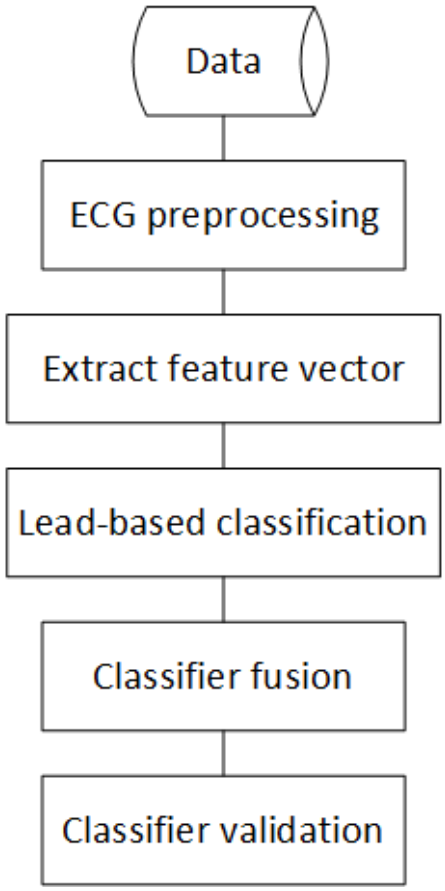

Figure 6. Schematics of the automatic ER detection system.

4.1. ECG preprocessing

ECG signals are often contaminated by noise and artifacts like baseline drift, power line interference, electromyographic noise and motion artifacts [10]. Therefore, it is necessary to preprocess first the signals in order to remove the unwanted noise that affects the quality of the features extracted for the supervised classification. Preprocessing of the data included resampling, baseline offset removal, R peak detection and S-waves and low-SNR filter; which will be discussed in the following sections. One

of the crucial steps in the ECG preprocessing is the removal of the baseline offset because it determines the amplitude of the JP in the notch/slur ER, which is one of the thresholds and criteria to decide whether a patient has the pattern or not.

4.1.1. ECG resampling

The first step was to preprocess the data by resampling the ECG recordings from 256 to 512 Hz. Resampling assumes interpolation of available samples to obtain or remove samples in between the given ones [37]. Upsampling (add data samples) was performed in order to increase the frequency rate of the signals and better locate the ECG morphologies for the extraction of the features. Linear interpolation was used for the resampling of the data, where a linear curve is fitted to obtain the new data points, and a least-squares linear-phase FIR filter is implemented to avoid aliasing (Figure 7).

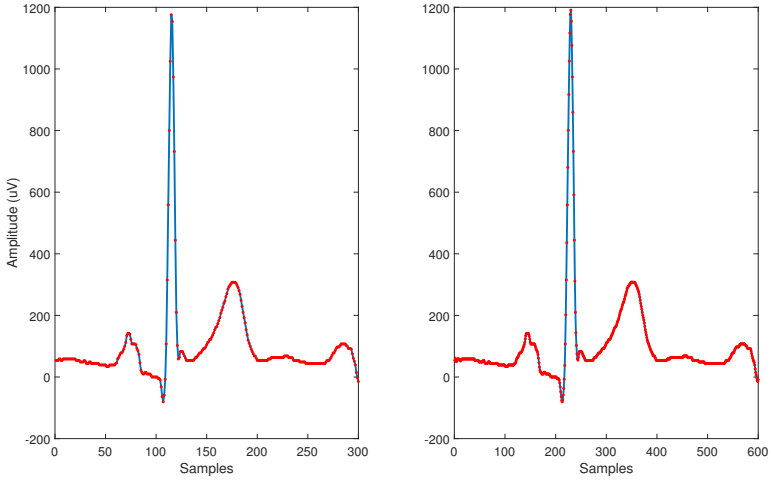

Figure 7. ECG signal resampled from 256 Hz to 512 Hz.

4.1.2. Baseline offset removal

The isoelectric line in the ECG signal used as a baseline for the ER measurements is the PR segment. Therefore, proper automatic detection of this segment was necessary in order to be able to remove the offset, by setting the amplitude of this segment to zero. The procedure used for the removal followed the logic implemented in [4], where the onset of the QRS complex is located and a time window is defined, to calculate the mean amplitude of the PR segment, which is then subtracted from the whole signal.

Onset of the QRS complex was detected by delimiting the interval where it should be located and computing the angles between each sample in the interval, in order to find the minimum value that represents the QRS onset[38]. Once the QRS onset (QRSon) is detected, an average value of the PR segment is calculated by computing the mean value of the data in a 30 ms window between QRSon-30ms and QRSon. In

this way, we can subtract the baseline offset and guarantee that the amplitudes of the signals correspond to the amplitudes measured from the PR segment (Figure 8).

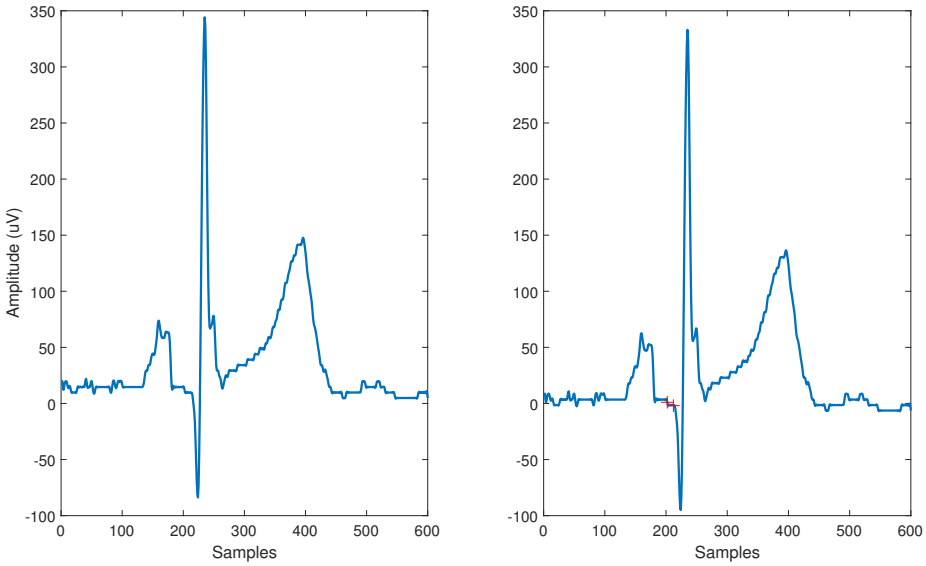

Figure 8. ECG signal with baseline offset (left) and with the removal of the offset based in the PR segment marked with the red crosses (right)

4.1.3. R-peak detection

R peak detection is performed in order to later locate the 50% amplitude of the downslope of an R-wave, which determines the starting point of the ER pattern [2]. The most-known algorithm for R peak detection is the Pan-Tompkins QRS detector [39]. The idea behind this algorithm is that the QRS complex is characterized by a rapid change in the signal amplitude, which can be detected as a higher frequency component in the wave. A band-pass filter is first implemented to attenuate noise, followed by a differentiator to suppress the low frequency P and T components, then squaring the signal emphasizes the large differences in the slope, and finally a moving window integration produces a signal that includes the information about the slope and width of the QRS complex [39]. Using this signal, we define a QRS search window, determined by an amplitude threshold level, and detect the R-peak as the maximum amplitude value in the band-passed filtered ECG signal (Figure9). In this way, the positions of the R-peaks were detected in the ECG signals and stored for the feature extraction.

4.1.4. S-waves and low-SNR filter

A rule-based filter is implemented to filter out low signal-to-noise ratio (SNR) recordings and ECG signals with S waves that are not allowed in the ER pattern definition [2],[3] and would affect the performance of the system.

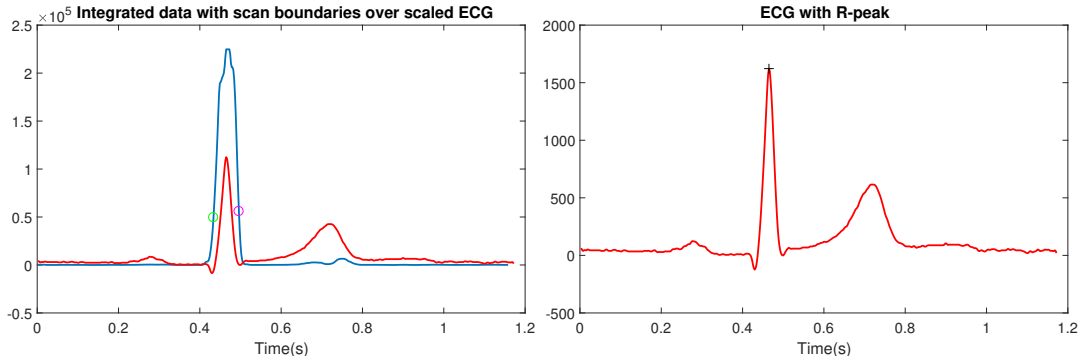

Figure 9. R peak detection with the integrated ECG data to emphasize the QRS changes. The figure in the left shows the scaled ECG data as well as the integrated signal after the moving window integration with the boundaries where the scanning for the QRS should begin (green circle) and end (purple circle). The figure in the right shows the original ECG signal with the R peak detected.

The ECG can contain so many different types of noises and artifacts that occur simultaneously, are often transient, and largely unpredictable in terms of their onset and duration [10]. The SNR is a measurement to compare the level of a desired signal to the level of background noise [40]. In SNR a value greater than 0dB indicates there is more signal than noise. The SNR is computed between the power of the desired signal and the power of the noise. The power of the signal depends on its amplitude, so bigger amplitude signals represents bigger powers. Therefore, R-peak amplitudes were used to remove low signal-to-noise ratio ECG recordings (Figure 10). QRS averaged complexes with R-peak<150 uV, were removed from data.

Likewise, ECG signals with S waves were filtered from the data because they are not included in the ER pattern definition [2],[3] and should not be represented in the feature vector. S wave detection was accomplished by looking for the inflection points and zero-crossings of the ECG band-passed and differentiated signal in a 20 ms time-window determined by the position of the R-peak. An amplitude threshold level in -80 uV was also used to avoid false S waves generated by small noise detection (Figure 10).

4.2. Waveform prototype feature extraction

Extraction of a feature vector consists on the acquisition of a set of prototypes that reflect the distribution of the data to be classified. The features should be able to represent the class they belong, in this case the positive and negative ER waveforms, and train the classifier to later predict the label of new data. Clearly, the more discriminative power have the chosen features, the more accurate will be the whole system [8].

The features usually consist in measurements from the signals (e.g. heart rate, standard deviation of R peak intervals, slopes, frequency analysis, features derived from the shape of the ECG, etc.), information related to the patient (e.g. age, gender, weight, etc.) or both. Other descriptors like wavelet-based features, parametric modelling ap-

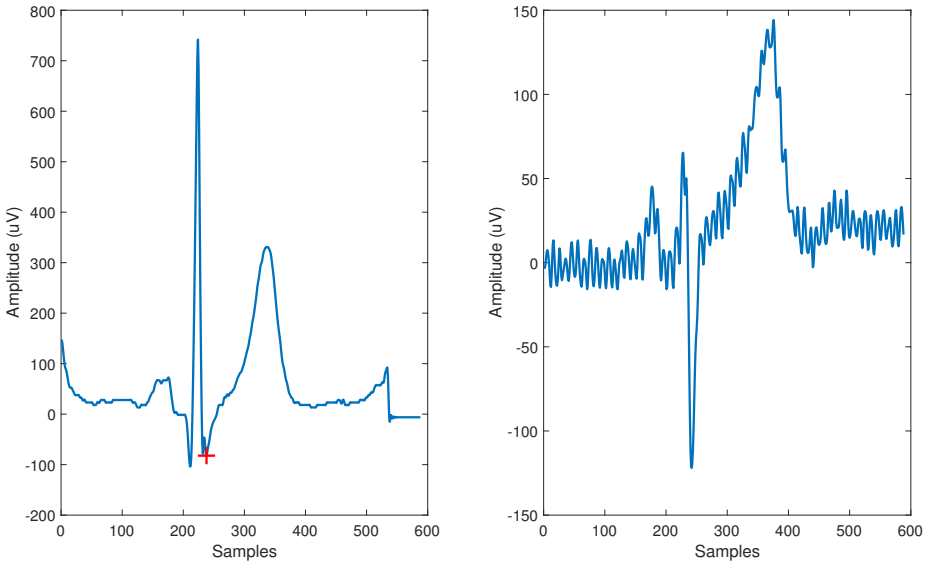

Figure 10. S waves detection (left) and low SNR signal (right).

proaches and features based on non-linear dynamics had been used for ECG signals and depend on the application (e.g. heart rate variability analysis, arrhythmia detection, human identification)[8].

As it was explained in Chapter 2, current methods for the detection of the ER use features for characterization of specific morphological details of the down-slope shape; like the angle of the QRS down-slope, the yield point or point of inflection, times and amplitudes of the QRS complex and the area under the curve [4],[5],[6]. Our proposal is to used waveform prototype-based features for ER classification. Research in cognition has shown that humans categorize objects based on hierarchical comparison using a generic prototype as a point of reference to quantitatively measure the difference of new data with the sample prototype [7]. Prototype-based features had been used for image classification where a set of multiple generic images are constructed and used for classification e.g. face images for face recognition, image-based medical diagnostic tools, etc. [7],[41].

A prototypical case must not necessarily be comprised of the whole set of features describing the cases, it can be represented by only a subset of the most important features [41]. Therefore, the prototype-based feature vector used is a fragment of the ECG signals where the ER pattern is located. The start point is the 50% amplitude of the R-peak in the QRS down-slope and the end point is fixed in a 25 samples window (Figure 11). Cubic spline interpolation was used to reduce the feature vector to 15 dimensions.

In general, the features depend on the specific application and are selected according to it. However, the features for classification should be [42]:

1. Robust. Invariant to changes related to external factors different than the pattern itself. For example, electrode movements, other ECG related changes in the waveshape morphology, noise, etc.

2. Discriminating. Subjects in different classes should have different values.

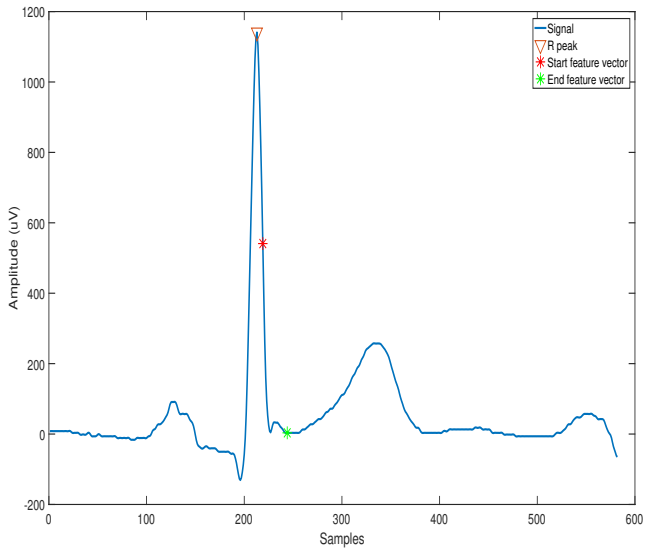

Figure 11. Boundaries for the prototype-based feature extraction.

3. Reliable. Subjects of the same class should have similar values.

4. Independent. Uncorrelated. Different subjects and ECG recordings.

In order to be robust, the cases with S waves and low SNR were removed, because those signals have different waveshapes that could influence the predicted labelling. Similarly, the baseline offset is removed to guarantee that all of the signals are in the same amplitude level and can be compared directly. Then, the amplitude values of the features can represent changes in the down-slope shape related to the ER pattern and not to the baseline offset.

Therefore, the prototype features are robust and the amplitudes are reliable and discriminative for ER. The last condition of the prototypes is that they should be independent. ECG leads for the same subject are not independent. They can have changes in amplitude and morphology due to the "view" of the heart they represent, but they are collecting at the same time and can even be contiguous. Consequently, the features were extracted per each ECG lead independently, and each lead should have an independent feature vector and model for the classification.

4.3. Lead-based classification

The automatic detection of ER was performed, by using the features extracted to train the three different supervised classifiers described in Chapter 3, in each of the eight ECG leads and compare its performance. The classifiers included were KNN algorithm, SVM and LDA. KNN classification model was trained with cityblock distance metric and using 9 neighbors to predict the labels for each lead. Gaussian kernel was used in the SVM classification for the detection of the ER. Results on the classification labels were compared using Friedman test and the Bergmann-Hommel post-hoc,

described in Chapter 3, since it is the most powerful approach and the number of classifiers to compare is small.

4.4. Classifier fusion

Mixture of classifiers (MOC) allows to take advantage of the performance of various classifiers based on different features[8]. There are two techniques for MOC: classifier selection and classifier fusion. In classifier selection, each classifier is trained in an area of the feature space and a weight is given to the feature vector according to the distance to the region [8]. In classifier fusion, each classifier is trained over the whole feature space and the label is assigned by combining their individual opinions to derive a consensus decision [8],[43]. This can be achieved by using different feature sets as well as by different training sets [43]. The combination is obtained with several methods such as majority voting, product rule, sum rule, maximum posterior probabilities, weighted outputs, etc. [8][43].

In this research, eight classifiers were trained independently with different feature sets. Thus, it is no longer possible to consider the computed a posteriori probabilities, as the classification systems operate in different measurement spaces [43]. Following the guidelines of ER pattern detection published in [2], the voting fusion was selected because it corresponds to the standard clinical practice, where at least 2 contiguous leads should have the pattern in order to be considered positive ER. The filtered leads by the S-waves and low-SNR filter were labelled as negative and fused with the lead-based classifier labels. Leads were divided in inferior and lateral and the labels were counted in each region. If more than two votes were found in the region, the subject was classified as positive ER for that territory.

4.5. Classifier validation

Schematics of the classifier validation are shown in Figure 12. The data is first split into testing and training. Then, the rule-based filter removes low SNR ans S waves data from the training feature vectors and assign negative labels to leads in the testing feature vectors with same characteristics. Additionally, for the classifier validation, the data is balanced and the models are trained and tested with the balanced data.

4.5.1. Data splitting for cross validation

In k-fold cross-validation, the dataset is randomly split into k-folds of equal size and the model is tested with one of the folds and trained with the remaining ones. In this way, the performance of the model is evaluated k times and the result is the average value. K-fold cross validation allows to use all of the data for testing and training at the same time, which is ideal when the data sample is small. Thus, the data was divided into 5 folds and 5-fold cross validation was performed.

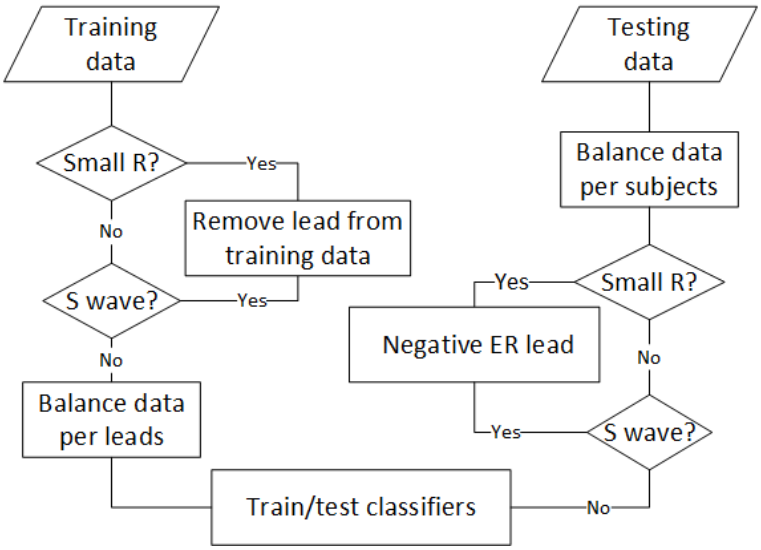

Figure 12. Schematics of the first stage classifier.

4.5.2. Balancing of the data

In an unbalanced dataset, the majority of the data represents one of the classes or labels, while the other one corresponds to a minority of the dataset. This is specially common in medical applications where the prevalence of a disease or a pattern is small compared to the total population. Published observational studies suggest that the prevalence of ERP ranges between 1% and 18% [3]. Studies have shown that imbalanced datasets can result in poor performance from standard classification algorithms [44], especially k-nearest neighbor where the data is classified according to the labels of the closest k training samples.

Thus, the data needed to be balanced to have the same number of instances among the two classes (ER positive and ER negative). In the training data, the same number of cases for each class was selected per each lead independently using the lead based labels after the removal of the wave-shapes with small R-peaks and S waves (Figure 12). For the testing set, the data was balanced per subjects (Figure 12), meaning that the same number of positive and negative ER subjects (cases with either lateral, inferior or both ER) were selected. Balancing of the data for testing and training was done by selecting all of the data with ER positive labels and randomly selecting the same number of ER negative cases.

4.6. Borderline cases assessment

SVM and LDA classifiers define a decision boundary between classes that it is used later for the prediction of new labels. In SVM the boundary is defined by the support vectors, and in LDA using the whole dataset. If the classes are separable, the hyper-plane allows to correctly classify all of the data samples. However, if the distributions of the data are overlapping the classifier will fail in some cases (Figure 13).

The most common solution for non separable data using SVM is to map the data into the high dimensional space, as it was explained before in Chapter 3. Nevertheless, due to the fact that the SVM algorithm uses a linear hyperplane and a soft margin in order to converge and reduce the risk to over-fitting, the high dimensional feature space is not always perfectly separable. On the other hand, the non-separable problem in LDA is commonly approached using quadratic surfaces but with the drawbacks that it increases the computational cost, the risk of over-fitting and it doesn't guarantee either a perfect separation between the classes.

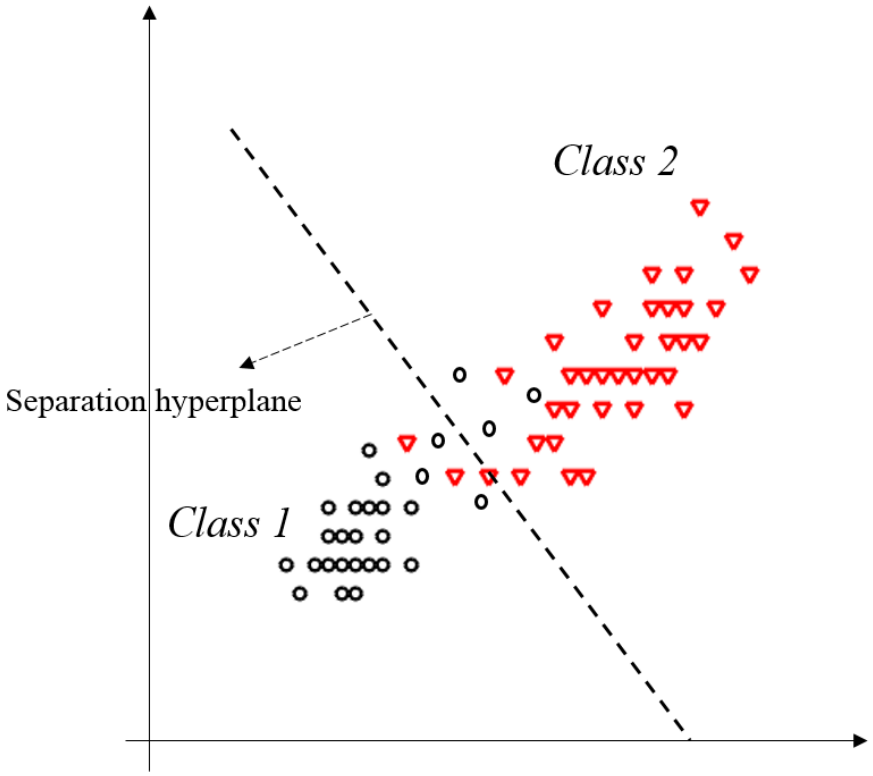

Figure 13. The problem of non-separable data due to overlapping distributions.

As it can be observed in Figure 13, the distributions of the data overlapping are closer to the boundary defined by the classifier. Therefore, we could measure the distance of the data samples to the separating hyperplane, in order to assess the confidence on the decision of the classifier, and to identify borderline cases that are difficult to classify. These cases could be characterized by Jp amplitudes close to the threshold of 0.1 mV or slightly changes in the QRS downslope.

The equation (1) for SVM provides an estimate of the signed distance of x to the separation hyperplane which can be then used to analyze how close are the samples from the decision boundary. Similarly, we can analyze the distance of the testing data from the hyperplanes provided by the LDA classifier. The equation of the signed distance of a point to a plane in the hyperspace is:

$$D = \frac{\sum_{i=1}^{n} a_i x_i + c}{\sqrt{\sum_{i=1}^{n} a_i^2}} \tag{8}$$

where $(a_1, \ldots, a_n)$ represents the coefficients of the hyperplane, c is the constant and $(x_1, \ldots, x_n)$ are the coordinates of the point.

Using the equations (1) and (8), the distances of each of the data samples to the classifier boundary region were calculated for each lead independently. However, we wanted to analyze the subjects that are close to the border and not the leads separately. Thus, the distances of contiguous leads were combined by averaging the normalized distances to the separating hyperplane. Here, we make the assumption that, contiguous leads should have an average distance closer to the classifier separation hyperplane in order to be consider as a borderline subject for manual review by medical experts.

The idea behind calculating the distance of the testing samples to the classifier boundary, is to provide an additional parameter to cardiologists and specialists or to advise them about cases that are difficult to classify, by the system and even by doctors, whether the patient has the ER pattern or not. In this way, medical personnel can manually evaluate the cases where the classifier is not that confident on the decision and reduce the number of false positives and false negatives missed by the automatic detection system of ER.

4.7. Features dimensionality reduction

Dimensionality reduction of the features was considered, to reduce time/complexity computation and allow data visualization of the distribution of the features in the hyperspace, for the automatic detection of the ER pattern. As it was explained in Chapter 3, a combination of Isomap and GRNN allows to preserve the intrinsic properties of the data, embed new samples into the lower dimensional space and reconstruct the original signal. Three-dimensional feature space representation was selected based on the residual variance reduction using the Isomap algorithm (Figure 14). Even though the residual variance of the data is further reduced with embedding the data to higher dimensions than 3-dimensional, it would not allow to easily visualize the data.

Figure 15 shows the schematics of the feature dimensionality reduction algorithm. Isomap is used to embed the original 15-dimensional training data into the 3-dimensional feature space. Dimensionality reduction of the testing data is accomplished using the GRNN. Thus, the GRNN are trained with the 15-dimensional and 3-dimensional training data of each lead, and used to map the testing data into 3-dimensional space .

Additionally, in order to maintain the physiological meaning of the original feature space, GRNN were used to reconstruct the embedding data into the original 15-dimensional feature space (Figure 15). GRNN were constructed with the same training data as before, but in this case they were trained to reconstructed the testing data to 15-dimensional feature space.

After the dimensionality reduction, the classification is performed. The algorithm was thus utilized as a preprocessing step: the data dimensionality is reduced to avoid the classification bias resulting from the curse of dimensionality [32], allow data visualization and reduce time complexity computation. In this way, the efficiency of the nonlinear mapping to maintain the discriminant features of the original data was tested, and the performance of the reconstruction of the data was evaluated based on the results of the automatic detection of the ER pattern.

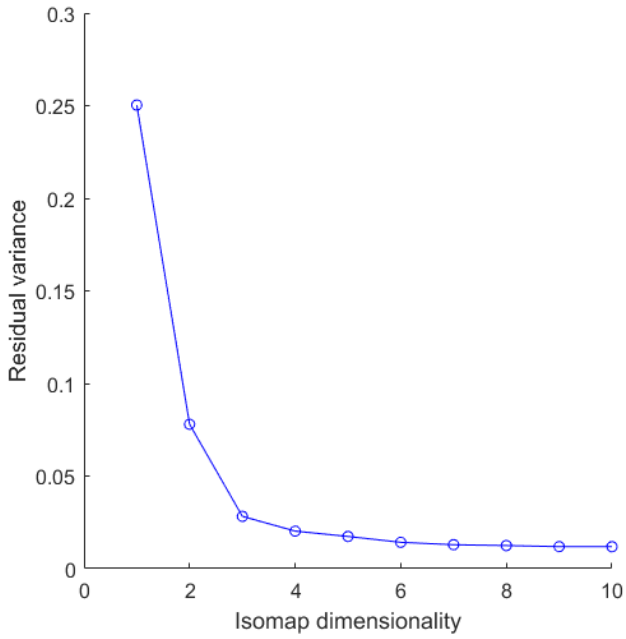

Figure 14. Residual variance of the embedding data using Isomap mapping. X axis represents the dimensionality of the data mapped with Isomap and Y axis the residual variance of it. It can be observed that the residual variance of the data diminish as the dimensionality increases with a break point in 3 dimension

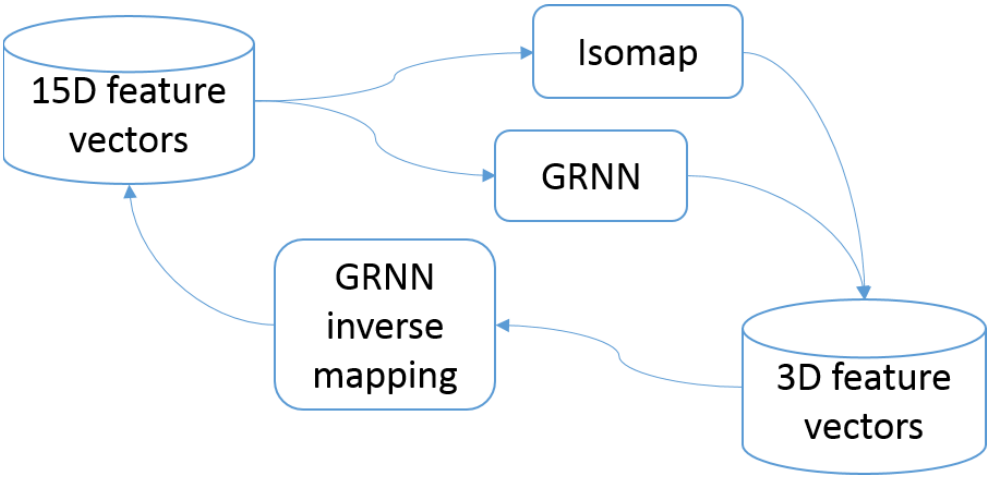

Figure 15. Feature dimensionality reduction and reconstruction.

5. RESULTS

5.1. Study data

The number of study subjects was 8028 adults, between 30 and 80 years. Subjects were selected from the Health 2000 Survey, a general population-based survey in Finland between 2000 and 2001, representative of the Finnish adult population at the time. The study followed all of the principles outlined in the Declaration of Helsinki and was approved by the Institutional ethics committee of Helsinki and Uusimaa hospital district.

Digital 12-lead ECG, were recorded on 6354 subjects using the Marquette MAC 5000 electrocardiograph (GE Marquette Medical Systems, Milwaukee, WI). Averaged representative beats were produced for each lead from the 10-second recording with QT Guard software (v. 1.3, GE Marquette Medical Systems). Subjects with prolonged QRS duration (>120 ms), pre-excitation syndrome, non-sinus rhythm and low-quality ECG were not included in the dataset. After exclusions, a total of 5676 subjects remained in the cohort for the analyses.

Manual grading of ER was performed by an experienced analyst, blinded to the results of the automated method, who collaborated with the study. He graded the ER following the recent consensus statement [2] for each inferior (II, III, aVF) and lateral (I, aVL, V4-V6) lead. At least two inferior or two lateral leads needed to have the ER pattern to consider a subject ER positive. Summary of the manual labels of the data are presented in Table 1. A total of 45408 ECG leads were labeled by the experienced analyst, finding the ER pattern in 3128 of those signals. Inferior ER was more common than lateral ER, even though the number of leads in the inferior region is smaller. From the 5676 subjects, 844 were classified as ER positive either in the inferior, lateral or both territories (57 cases). An example of an infero-lateral ER positive recording is shown in Figure 16.

Table 1. Summary of manual labelling of the data

Subjects	II	III	aVF	I	aVL	V4	V5	V6
ER positive	518	681	349	498	581	41	129	331
Region positive	Inferior: 506			Lateral: 395				
Total positive	844							

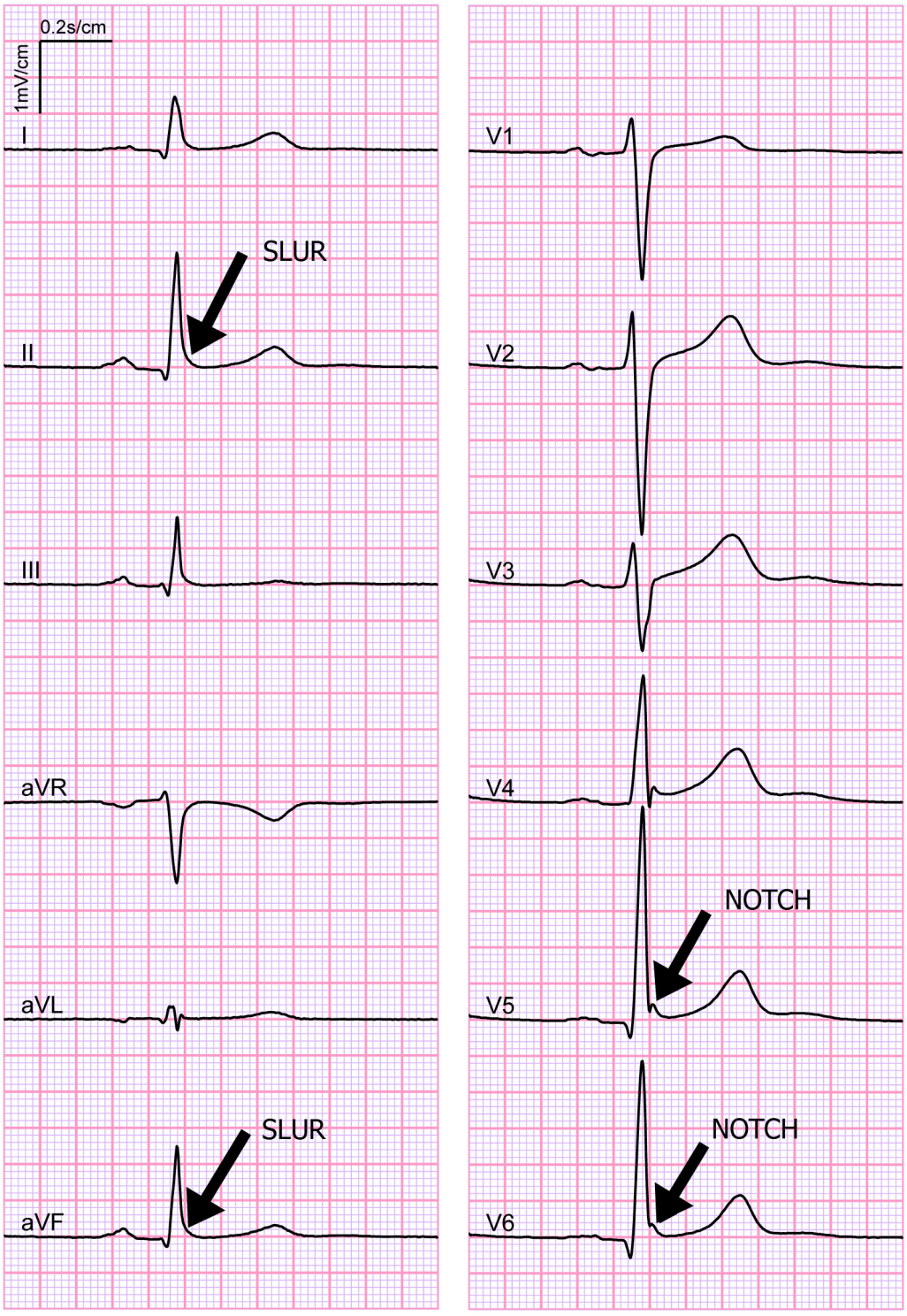

Figure 16. An example of a subject with infero-lateral ER pattern. Leads II and aVF show slurred ER and leads V5 and V6 show notched pattern. Paper speed is 50 mm/s.

5.2. Acquisition of the features

Table 2 summarizes the preprocessing of the data with the low SNR cases removal and ECG leads with S waves. A total of 1229 leads (2.71%) were found by the algorithm to have small R amplitudes and S wave was detected in 23440 leads (51.62%).

Table 2. Summary of preprocessing of the data

Subjects	I	II	III	aVL	aVF	V4	V5	V6
Low SNR	9	14	596	377	227	6	0	0
S waves	1808	2354	2486	2181	2348	5390	4498	2375
Remaining	3830	3279	2565	3089	3072	251	1149	3272

Additional to the low SNR and cases with S waves, 29 subjects were removed from the cohort because the R peak or QRSon detection failed in one or more leads (see example in Figure 17).

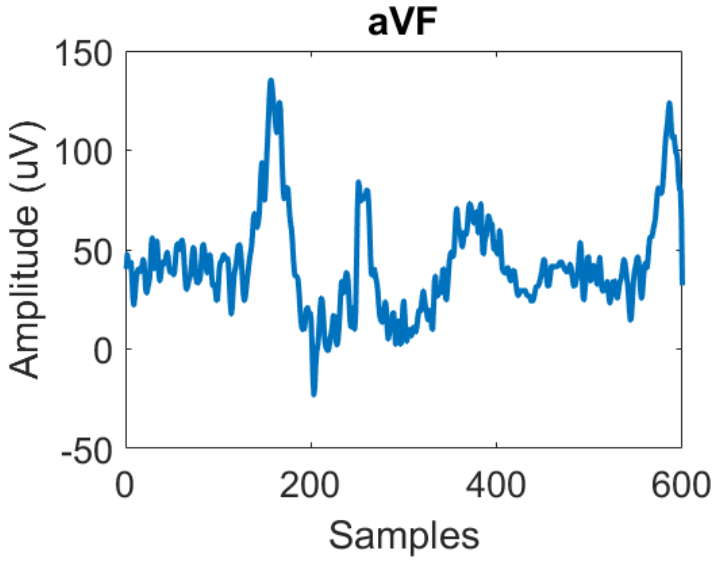

Figure 17. An example of a lead where the detection of the QRS onset failed and the case was removed from the analysis.

Features were extracted from the remaining cases of the last row of the table. Examples of the features extracted are shown in Figure 18 for inferior leads, and Figure 19 for lateral, where the top rows represent negative ER leads, and the bottom positive cases. Variability of the pattern among the leads, related with changes in amplitude and notching or slurring configuration, can be observed and represent the differences in the prototype-based feature vector extracted between normal ECG and ER leads. In Figure 18, lead II represent ER with slurring configuration and leads III and aVF, ER notch pattern. Among of these, leads II and III are close to the threshold level to be considered as ER positive, due to the amplitude of Jp close to 0.1 mV. In the example lateral features, shown in Figure 19, positive ER signals in leads I and V4 represent notching ER; and aVL, V5 and V6, slur ER. A notch can be observed in the negative ER case for lead I in Figure 19. However, the amplitude of the Jp is less than 0.1 mV

so it is not considered as ER pattern. Similar thing happens with V6, where the lower signal is considered as ER because the amplitude of Jp, and the upper case is considered negative because the apparent slowing of the inscription of the waveform is below the amplitude threshold.

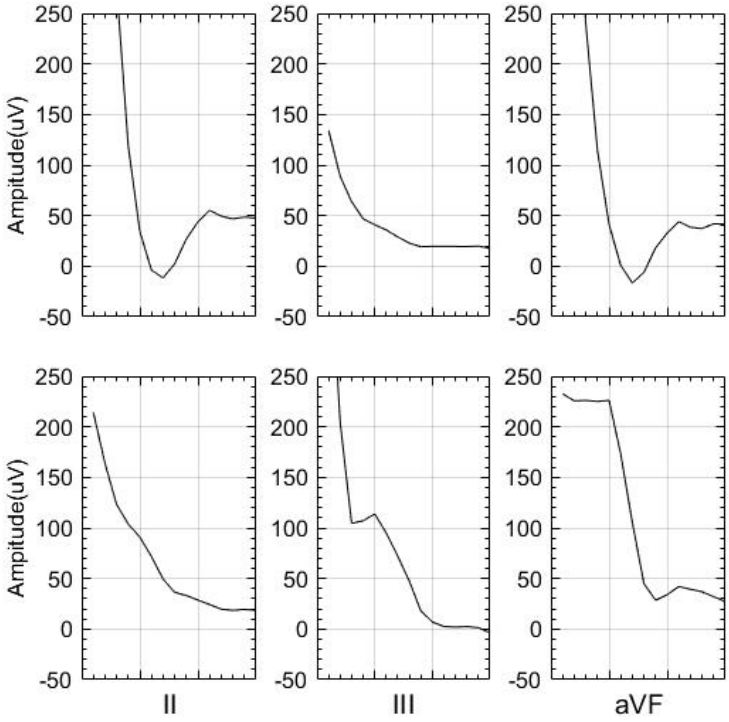

Figure 18. Example of extracted features in the inferior region. Features in the top line represent negative ER cases and in the bottom positive notch/slur ER cases.

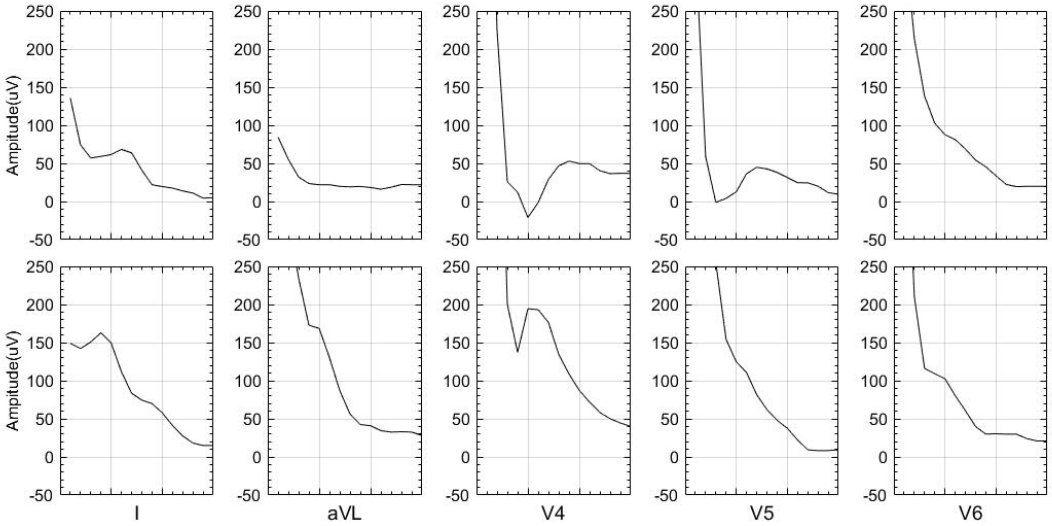

Figure 19. Example of extracted features in the lateral region. Features in the top line represent negative ER cases and in the bottom positive notch/slur ER cases.

5.3. Lead-based classification

As it was explained in Chapter 4, the classification of the features extracted is done per each lead independently. Thus, eight classification models are trained and tested separately. Table 3, contains the accuracy performance, of the 5-fold cross-validation, for each model and classifier. Lowest classification accuracy was accomplished with lead V4. As it was showed in Table 2, the number of cases to train and test the classifier in V4 is less than 251 after the balancing of the data. Consequently, the performance is lower because the dataset is not big enough to train a discriminative model. However, when the results of the classification are combined with the S-waves and low-SNR filtered cases, the accuracy performance of the models increase for all of the classifiers and leads, specially in V4, proving that the rule-based filter is mostly accurate.

Overall, the performance of the lead-based classification is around 90% with the three different classifiers, indicating that the models and the features are discriminative for ER pattern in all of the leads.

Table 3. Accuracy of the lead-based classification using KNN, SVM and LDA; and the combination with the S waves and low SNR filter

Lead accuracy	I	II	III	aVL	aVF	V4	V5	V6
KNN classifier	0.88	0.87	0.80	0.86	0.87	0.79	0.87	0.84
KNN + filter	0.92	0.91	0.90	0.93	0.92	0.98	0.96	0.89
SVM classifier	0.90	0.89	0.83	0.86	0.89	0.78	0.87	0.86
SVM + filter	0.93	0.92	0.91	0.93	0.93	0.98	0.96	0.91
LDA classifier	0.88	0.86	0.81	0.86	0.86	0.72	0.89	0.84
LDA + filter	0.91	0.90	0.90	0.92	0.91	0.98	0.97	0.89

5.4. Classifier fusion and validation

Figure 20 shows an example on the data distribution for one of the folds in lead I. Four parts of the data are selected as the training set (4541 cases). Then the signals with S waves and low SNR are removed. From the remaining cases (3032), the balancing of the data is performed by selecting 844 cases that correspond to 422 positive ER leads and 422 negative. These cases are the ones used to train the classifier. On the other side, the remaining fifth part of the data is used for testing. The balancing is done first by selecting the same number of positive and negative ER subjects, independently the lead labels. Then the S waves and low SNR recordings in that particular lead are directly classified as negative ER. The remaining cases in the testing dataset (257) are used to evaluate the classifier and predict the classes.

The data distribution between the leads vary. The training and testing subjects remain the same, but the cases with S waves and low SNR are different for each lead. Therefore, the number of cases in the feature vector to train the lead-based classifier varies between leads as well as the number of labels predicted. For example, considering the results in Table 2, the number of S waves detected in lead V4 is bigger than in lead I. Thus, the number of cases to train and test the classifier for this particular lead would be smaller than for lead I.

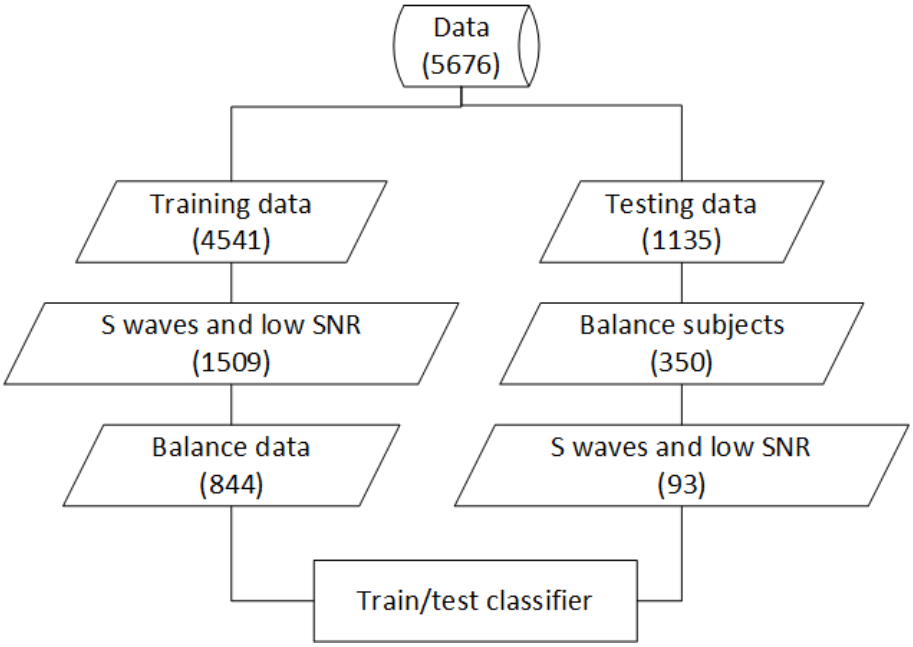

Figure 20. Balancing of the data for lead I in one of the five folds.

The performance of the classifiers was evaluated with the 5-fold cross validation over 10 runs using the same training and testing data among all of the classifiers. The reason to perform 10 runs combined with the cross-validation, is that the whole system includes also random selection of negative cases in order to balance the datasets and S waves and low-SNR filter.

Figure 21, shows the box-plot accuracy of the three classifiers. The accuracy of the classification over the 10 iterations of 5-fold cross validation was between 87.58 and 95.93%. Mean values for inferior ER were 92.48% using KNN, 92.74% for SVM and 91.29% for LDA. Lateral region accuracies showed lower performance compared to inferior detection using KNN and SVM. The mean accuracies for lateral ER detection were 91.57% with KNN, 92.21% with SVM and 91.48% with LDA.

For assessment purposes of the classification, the sensitivity and specificity of the system should also be evaluated. The sensitivity or true positive rate (TPR), is defined as the number of ER cases detected by the algorithm over the total number of positives. The specificity (SPC) or true negative rate, is defined as the number of non ER cases detected over the total number of ER negatives. TPR and SPC are calculated as:

$$TPR = \frac{TP}{TP + FN} \tag{9}$$

$$SPC = \frac{TN}{TN + FP} \tag{10}$$

where True Positives (TP) are the real ER cases detected by the classifier, False Negatives (FN) are the ER events not detected, False Positives (FP) are the non ER cases labelled as ER, and True Negatives (TN) are the correct detection of non ER cases.

SPC and TPR were obtained for the overall performance of the ER detection including the rule-based filter and the predicted labels by the supervised classifiers methods.

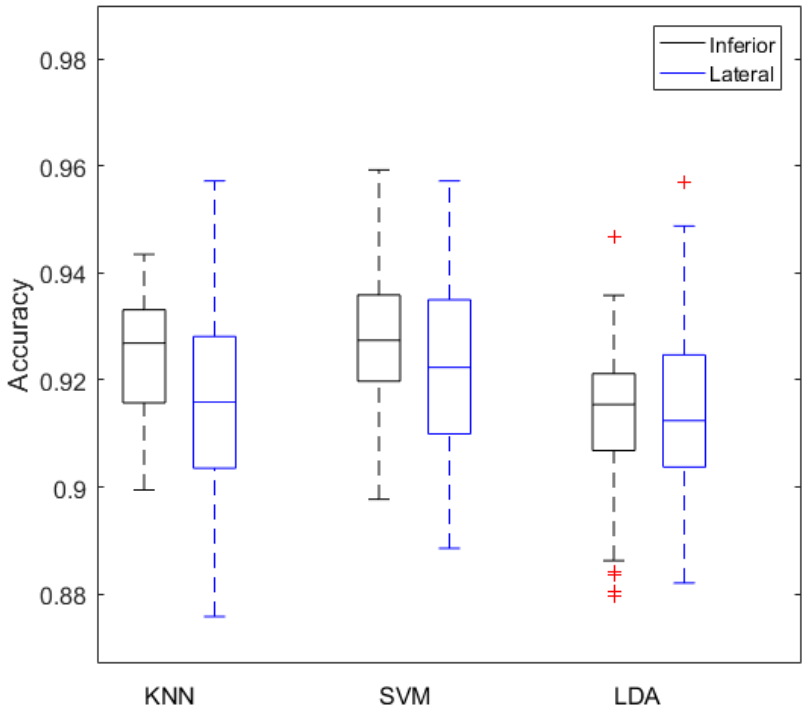

Figure 21. Classification accuracy of ER using the KNN, SVM and LDA algorithms over fivefold cross-validation with 10 repetitions. Each boxplot indicates the median (50th percentile) and interquartile range (25th and 75th percentile) with outliers marked with +.

TPR of the ER detection is shown in Figure 22. Range of TPR was between 100% and 78.84% in the lateral ER detection using LDA. Mean values of TPR in inferior ER were 93.75% using KNN, 91.87% using SVM and 88.31% using SVM. Average TPR in the ER lateral region was highest with KNN (93.36%) algorithm and lowest using SVM (91.74%).

SPC of the ER detection is shown in Figure 23. SPC in the inferior leads showed better results than in the lateral territory. The range of SPC was between 86.94 and 96.15%. Best average SPC was accomplished in the inferior leads (93.13%) and in lateral region (92.34%) using SVM. Lowest average SPC in the ER inferior detection (91.94%) and lateral (91.01%) was obtained using KNN.

5.5. Comparison of the classifiers

As it was explained in Chapter 3 and Chapter 4, the comparison on the performance of the three classifiers is done with Friedman test. The accuracy values of the classifiers, obtained from the 5-fold cross validations over 10 runs, were compared with the null-hypothesis that there is no difference between the algorithms. However, the null-hypothesis was rejected and statistical significant differences were found between the accuracies of the classifiers ($p < .05$). Thus, Bergman-Hommel's post-hoc test was

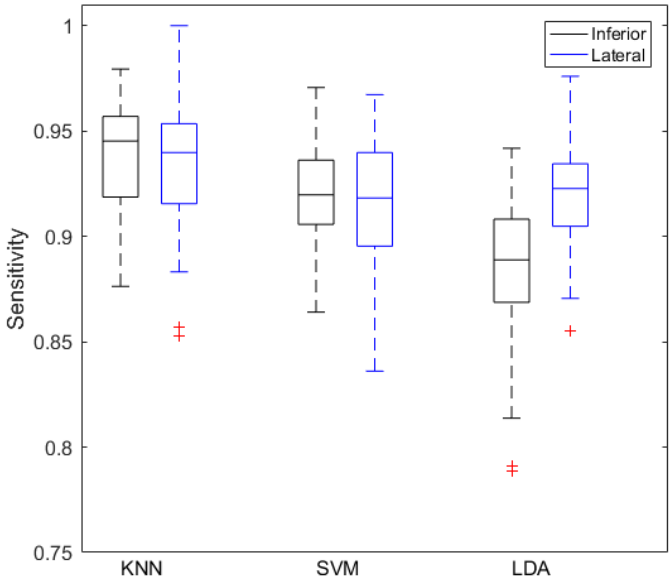

Figure 22. Classification sensitivity of ER using the KNN, SVM and LDA algorithms over fivefold cross-validation with 10 repetitions. Each boxplot indicates the median (50th percentile) and interquartile range (25th and 75th percentile) with outliers marked with +.

performed. Table 4 shows the results of the Bergman-Hommel's post-hoc test with ones indicating significant differences and zeros otherwise; ranking SVM as the best performing algorithm followed by KNN and LDA.

Table 4. Results of Bergman-Hommel's post-hoc test to indicate whether there are statistical significant differences (ones) between the classifiers or not (zeros).

Classifier	KNN	SVM	LDA
KNN	0	1	1
SVM	1	0	1
LDA	1	1	0
Average ranks	1.7	1.3	3.0
Final rank	2	1	3

SVM has also the advantage over the other classifiers that only the support vectors are used to compute the model which reduces its complexity and the amount of the data required. Compared with KNN, it provides the option to measure the distance to the borderline which is used to assess the confidence on the detection of the ER. LDA is simple and works well when the data is linearly separable. However, when there is overlapping of the data, it address the problem by creating a quadratic or polinomial surface in the hyperspace that can be over-fitted. SVM transforms the data to the hyperspace but with a soft margin that address the over-fitting problem. Thus, SVM classifier was found to be the best approach for ER detection in terms of the

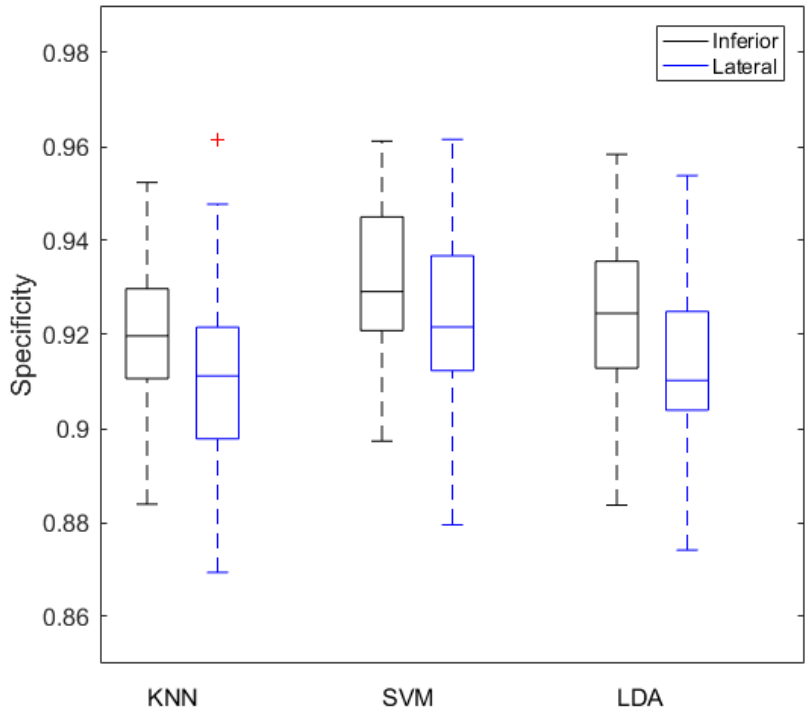

Figure 23. Classification SPC of ER using the KNN, SVM and LDA algorithms over fivefold cross-validation with 10 repetitions. Each boxplot indicates the median (50th percentile) and interquartile range (25th and 75th percentile) with outliers marked with +.

performance, complexity and model used to define the decision boundary between the classes.

5.6. Borderline cases assessment

Borderline cases were defined in Chapter 4, as the subjects that are close to the boundary that separates the hyperspace in classes. Assuming that the data is overlapping in the hyperspace, which is very likely to occur and explains why the classification performance is not 100% correct, the hypothesis is that the cases that are being misclassified are closer to the decision boundary surface. Those subjects are called critical cases because the algorithm is more likely to fail in the ER detection and should be manually labelled by doctors.

The distances were measured as it was explained in Chapter 4, using the testing data in each of the 5 folds, and stored for further analysis. Cases with S waves and low SNR are not included in the analysis because the features were not extracted, so the distances cannot be calculated. Figure 24 shows the histogram of the distances to the SVM classifier decision surface, and it can be seen in the inferior region, that incorrectly classified subjects are closer to the boundary, proving the hypothesis. Therefore,

we can automatically provide those critical cases closer to the boundary for manual assesment of the ER by medical specialists.

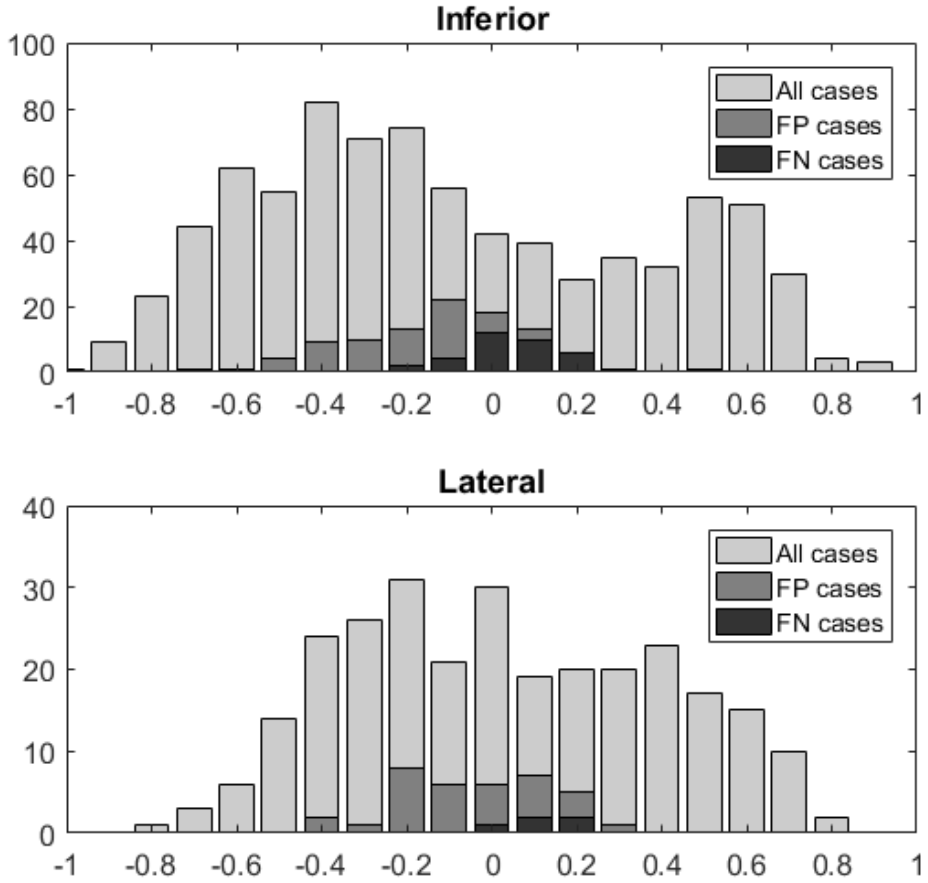

Figure 24. Histogram of the signed averaged normalized distances to the SVM classifier borderline. Cases with S wave and R peaks<0.15 mV are not included.

FN are critical in medical applications because if the patients are not correctly identified to have a medical condition, they would not receive the appropriate treatment or care. On the other side, FP can be easily identified by doctors when looking through the medical tests, in this case the ECG signals, in order to correct the diagnose and avoid unnecessary treatment. In the context of ER, failing to detect a positive case means that the person can be at risk of SCD and treated as without any risk which is very critical. Therefore, the borderline cases assessment was focused on looking for the FN.

Considering the distances of Figure 24, most of the FN cases are between 0 and 0.2. Therefore, if we extract the cases in the inferior region with distances between 0 and 0.2, 81 subjects from 1692 in total are selected for manual revision by experts, from where 22 were FN and 9 were FP. In this way, the TPR of the detection could increase, depending on the doctors verification, up to 96.05% and the SPC up to 94.18% in the inferior territory. However, in the lateral region, is more difficult to find the critical cases since the number of leads is bigger and the presence of S waves is more com-

mon, so we cannot include them in the analysis of the distances. From 49 cases with distances between 0 and 0.2 in the lateral leads, 3 were FN identified by the algorithm and 10 FP leading to an improvement on the TPR of up to 0.01% and on the SPC of up to 0.77%.

Similar results were obtained using the distances to the hyperplane constructed by LDA and with different data distributions. The difference in the data distribution is related to the balancing of the data that randomly selects the negative cases. Therefore, we wanted to check, if from different samples and data, we could still get good results for the analysis of the critical cases. From this dataset, 206 cases were identified by the algorithm as critical in the inferior territory with 43 FN and 17 FP (Figure 25). In the lateral region, 84 cases were detected as critical with 7 FN and 13 FP. In Figure 25, it can be seen that the distribution of the data follows more a normal distribution than in Figure 24, suggesting that the classes for this dataset are more overlapped. However, the algorithm in both cases is able to detect misclassifications based on the distances because the FP and FN are closer to the boundary (zero distance). The difference between the number of cases in Figure 24 and Figure 25 is a product of the S waves and low SNR recordings that are not included in the borderline cases analysis and differ in this situation.

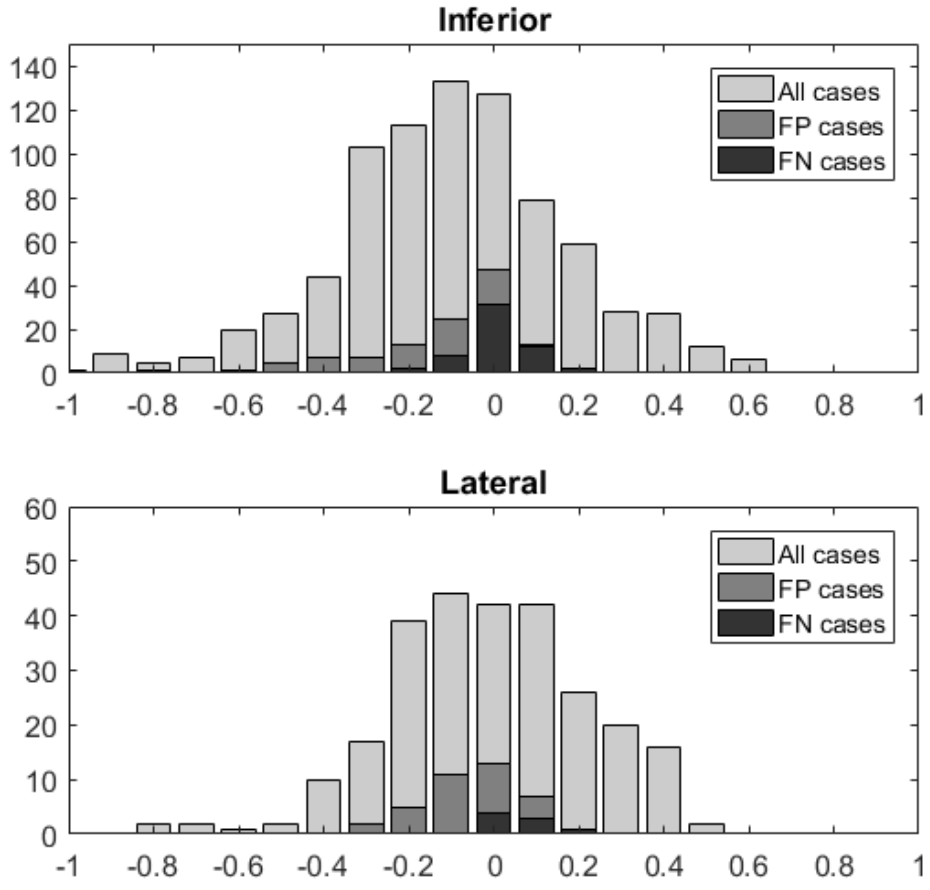

Figure 25. Histogram of the signed averaged normalized distances to the LDA classifier borderline. Cases with S wave and R peaks<0.15 mV are not included.

5.7. Dimensionality reduction

As it was explained in Chapter 4, the feature vectors were reduced from 15-dimensional to 3-dimensional feature space using GRNN and Isomap. The main advantage of this dimensionality reduction is to allow data visualization and reduce time complexity of the classification. Visualization of the 3-dimensional embedding data using Isomap and GRNN for lead I is shown in Figure 26. It can be observed that the two classes are overlapping, which demonstrates that the assumption of overlapping data is valid. It can also be noticed that the misclassified cases are closer to the boundary, specially the positive ER subjects, as it was suggested before and shown in Figures 24 and 25.

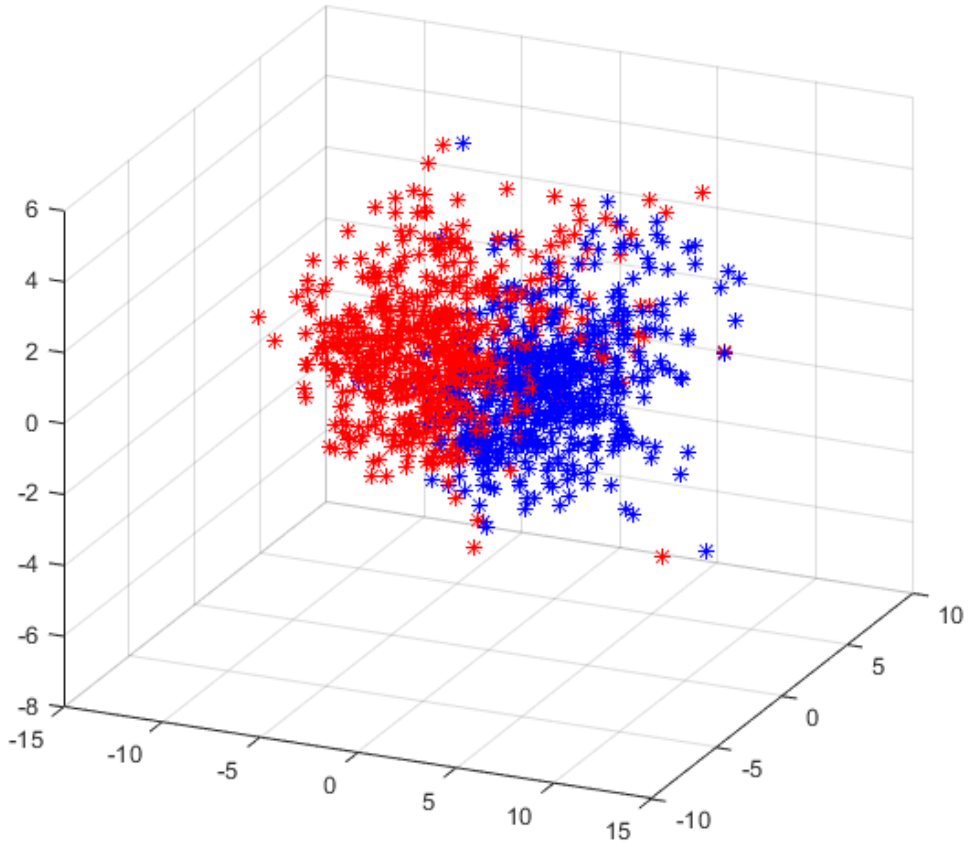

Figure 26. Visualization of the embedding feature vector for Lead II using Isomap. Blue asterisks (*) represent the ER negative cases and red asterisks (*) the positive ER.

One way to prove the embedded feature vector, is to train and perform the classification using the dimensionality reduced data. Therefore, the 3-dimensional features were used to train and test the ER detection using KNN, SVM and LDA. In addition, as it was mentioned in Chapter 4, the embedded data was then reconstructed to the original 15 dimensional feature space, to test the capability of the Isomap algorithm to maintain the discriminant characteristics of the features extracted and preserve the

physiological meaning of the original prototypes. The performances of the classifiers, using the 3-dimensional embedding data and the reconstructed 15-dimensional feature vector, are shown in Table 5.

Table 5. Performance of the three different classifiers with the 3-dimensional embedding data and the 15-dimensional reconstructed feature vector.

Classifier	Inferior		Lateral	
	TPR (%)	SPC	TPR (%)	SPC(%)
KNN 3D	92.76	89.40	93.08	86.82
SVM 3D	93.10	90.05	92.88	87.56
LDA 3D	89.05	91.08	91.11	87.43
KNN 15D reconstructed	93.18	89.14	93.50	86.59
SVM 15D reconstructed	93.73	89.60	93.71	87.04
LDA 15D reconstructed	88.81	91.08	92.38	87.28

TPR and SPC obtained are very similar to the average values obtained with the original feature vector. Differences between the performance of the classifiers using the embedded and reconstructed feature vector can be product of noisy samples. Thus, the performance of the detection is similar with the original and embedded data and it suggest that the Isomap algorithm preserves the characteristics of the prototypes and it can be reconstructed using GRNN. However, it can be observed in Figure 27, that the reconstructed prototypes are smoother and it is more difficult to evaluate the presence of the ER pattern. Two sets of signals, original prototype and reconstructed, are presented in Figure 27 and represent positive ER cases. The prototypes in the left side of the figure were extracted and reconstructed from a positive notch ER subject. The ones in the right from a slur case.

Even though the reconstructed signals are smoother, the performance of the detection is similar to the results obtained with the original and embedded prototypes. This suggests that the discriminant characteristics of the classes are kept but the physiological meaning is affected, specially for notch cases. Similar effect happens when using the embedded data, the reduced features have good discriminative power but they lack of physiological meaning.

5.8. Comparison with existing automatic method for detection of ER

The performance of the ER detection was also compared with the method proposed in [4] and presented in Chapter 2, which is based on the analysis and measurement of the QRS down-slope. Their algorithm was tested with the data used in this study in order to make the comparison of the performance in the detection of ER. Their algorithm showed a TPR of 63.83% and SPC of 95.67%. Inferior ER TPR was 66.80% and 97.35% SPC. Lateral ER detection was 57.72 TPR and 98.02% SPC. Our average performance using SVM showed results of 91.87% and 91.74% for inferior and lateral TPR respectively. Mean SPC using SVM was 93.13% in inferior territory and 92.34% in lateral leads. From the comparison. their method showed highest SPC. However and as it was explained before, the detection of ER should be focused on the TPR of

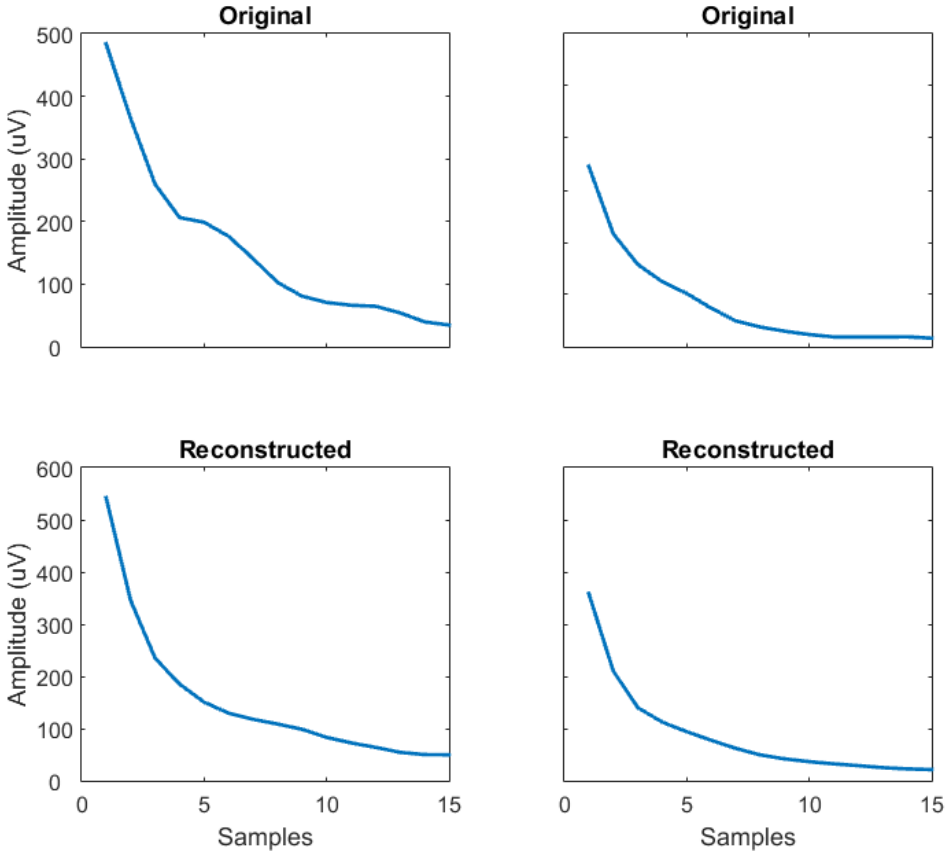

Figure 27. Original and reconstructed prototypes in two different subjects. Left features represent a notch case and right a slur ER.

the algorithm that reflects how well it detects the true positives from the total positive cases.

The advantage of using prototypes to classify ER over the quantification of the slope, is that it trains the model with different types of notches and slurs without performing any signal measurement. Other types of patterns can also be in the model to detect different morphologies e.g. ST segment elevation. Once the models of classification are trained, the only thing that is required for the ER detection is the extraction of the features. In slope analysis methods, the measurement and analysis have to be done which every lead and subject. Another main advantage of using classifier models, is that the models are lead-based, considering the morphology changes that can occur between leads, specially in terms of amplitude. Then, the detection is done per each lead independently.

One disadvantage of using supervised classification over the analysis of the slope, is that the labels of the data are, most of the time, obtained manually. Thus, if the definition of the pattern changes, the labels would need to be changed manually in order to create a reliable system. Similarly, if we would like to separate the actual data in notches and slurs to be able to classify them separately as the method proposed in [4], the manual labelling should be then performed again in this way.

6. DISCUSSION

In this study, a novel technique to detect ER pattern in ECG signals was proposed. The novelty of this method is in the features that are used to perform the detection. Current methods use morphological features measured from the signal e.g. the slope and amplitudes of the notch/slur pattern. In this study, the ECG fragment, where the pattern is located, was extracted and used directly to perform the classification.

Therefore, the main research question for this study was if the ER detection can be based on this raw ECG waveform classification instead of detailed morphological features. To address this question we have first to discuss the method here proposed. As it was presented in Chapter 4, the ECG signal requires some preprocessing first. Thus, the features are not extracted directly from the raw ECG signal but instead from the resampled, baseline removed signal. In this way some measurements like the detection of the R peak and its amplitude, the QRS onset and the PR segment are still needed in order to extract the features. Nevertheless, features related with the ER pattern itself were not required.

The method here, is then based, on prototypes that represent the different shapes of the ER pattern, as well as normal ECG signals. These prototypes require some conditions to guarantee that they can be compared using the classifiers. One of those conditions is that the baseline of all of the prototypes should be set to the same level amplitude. This allows to compare the prototypes in terms of their amplitudes, which is crucial because the ER pattern is defined itself in terms of the Jp voltage. Another requirement is related with the times between the features. In order to have time alignment features, the feature vector was extracted from a fixed window size, for all of the cases and leads. Thus, we can guarantee that we are comparing the prototypes amplitudes that occur at the same time.

A third condition for the prototypes is, that they should be able to represent the different shapes that ECG signals, normal and with the ER pattern, can have. There are two common ER patterns, the notch and the slur. However, each of them can appeared at a different amplitude and with a different Jp level and slope. Thus, the prototypes should be reliable and discriminating. By using a considerable amount of samples, like the ones in this study, we can assume that the prototypes are able to represent a wide variety of ECG down-slope changes and are reliable and discriminating. Also, the ECG signals with S waves and low SNR, that can affect the feature vectors and therefore the classification, which are not of the interest in this approach, were removed so the features can be robust and invariant to these wave-shapes that are not related to the ER pattern.

Finally, to address the question, if these features can be used to detect the ER pattern, the methods used and results obtained should be discussed. The overall accuracies of the different classifiers showed results over 90%, indicating that the prototype-based feature vector is a good representation of the differences between ECG signals with and without the ER pattern. Then, it can be stated that the detection of the ER pattern can be based in ECG signal prototypes with very good performance. Three different classifiers can be used for this purpose; KNN, LDA and SVM, all of them with excellent results. However, Friedman test found statistical significant differences on the accuracy of the classifiers ($p < .05$), and showed that SVM was the one with best performance. Additionally, SVM does not require large storage space, like KNN for

all of the samples, but instead it only uses the support vectors. It defines a borderline between the classes that was used for the analysis on the confidence on the detection. Different than LDA, it deals with the non-separable problem by mapping the original data, into a hyperspace using the kernel trick, which is simple and computationally efficient but yet providing good results.

The detection of ER is performed using lead-based models considering the slurred morphologies changes, that are common between the inferior and precordial leads, related to the differences in signal amplitudes [4]. Supervised learning models provide a general way to include different waveforms, and depending on the training set and labels, new morphologies can be included. In this method, two classes were considered, positive and negative ER, but other classes could had been involved to classify between slur and notch ER or to include ST-segment elevation.

Our algorithm outperformed the one proposed in [4] and tested with this same data. As it was explained in Chapter 2, their method is based on measurements of the ER pattern by fitting a line on the QRS down-slope. In their original publication, their algorithm showed results of 96.2% TPR and 90.1% specificity. However, since a new consensus definition on the ECG measurements of the ER pattern was published after their publication [2], the labelling of the data, on where a subject was considered with the pattern or not, changed in some cases. Previously, there was not an agreed definition on the Jp threshold level for the peak of the notch and onset of the slur. So cases, where the ER pattern was not that notorious, could have been considered as positive or negative depending on the reader. They used notch amplitude threshold ≥ 0.09 mV and slur ≥ 0.1 mV.

The new definition sets the Jp threshold ≥ 0.1 mV for both, notch and slurs. Therefore, this explains the difference between their results and the performance obtained using the same data and labels of this study. Using this data, their algorithm yielded a TPR of 63.83% and SPC of 95.67%, while the method here presented using SVM, showed overall results of 92.31% TPR and 87.05% SPC. The main interest in this application is to detect the ER cases that are evaluated with the TPR. Our results are the mean values, of a 5 fold cross validation over 10 repetitions, with a balanced data set. Thus, the method here proposed showed better results than the algorithm based on ER measurements. The main advantage of using prototypes over ER measurements is that it does not need algorithms to identify and measure the pattern. Thus, if the definition of the pattern changes, then the labelling of the data needs to be changed but not the method itself.

Detection of inferior ER is more significant because it indicates the greatest risk for arrhythmic events [4]. Inferior ER detection was slightly more accurate than lateral ER. Errors in the classification were related to corrupted signals, false slurred detection and cases of Jp close to the threshold. Patient movement, poor electrode to skin contact and averaging ECG complexes can induce random noise to the signals. Slurred patterns are more complex to detect than notched terminal QRS since they have a wide range of variations [4]. Additionally, human errors can affect the manual labelling of the ER especially in cases close to the Jp threshold and in the lateral region where the signal amplitudes seems to vary more and there is a bigger number of leads. It is often recommended that manual labelling of data is done by more than one experienced reader, in order to obtain validation of the labels. However, considering the amount of

study subjects and level of expertise that is required to assess the ER pattern, it was not feasible in this case to obtain more labels of the data to compare.

Identification of critical cases, based on the distances to the decision boundary, contributes to identify subjects that are difficult to label and require further medical examination. According to [3], overall risk assessment should be performed in cases with ER, to distinguish between patients who have the pattern and are not at risk of SCD and patients at risk of lethal ventricular arrhythmias. Clinical and electrocardiographic characteristics should be considered for the risk assessment; including sex, familial history of SCD, history of Syncope, type and location of ER, amplitude of the notch and morphology of the ST segment [3].

Results showed that the measurement of the distances to the classification boundary allows to identify FN, especially in the inferior region. Figure 28 shows a critical inferior case found by the algorithm based on the distance to the hyperspace separation of the classes. As it can be observed, the Jp level of the subject is close to 0.1 mV threshold. Therefore, this case is difficult to assess even manually because it falls in the limit on whether to consider it as positive ER or not. The method here developed is able to provide these cases to the specialists, who are the ones that should take the decision if the ER pattern is present and the patient is at risk of SCD.

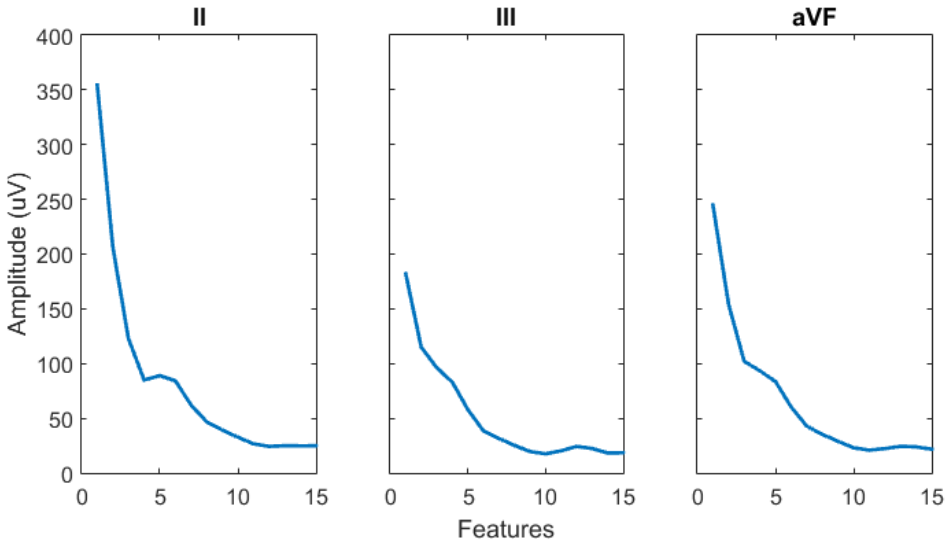

Figure 28. Critical inferior ER case found by the algorithm based on the distance to the borderline method

Dimensionality reduction of the prototype-based feature vector was accomplished with Isomap and GRNN. Automatic detection of the ER pattern showed similar results in the lower dimensonal representation of the prototypes, indicating that nonlinear mapping is an effective way of dimensionality reduction of the features. Nevertheless, when mapping the original features to the 3-dimensional space, the physiological definition and representation of the ECG fragment is lost. The features were reconstructed using GRNN with similar performance (Table 2), indicating that the discriminant characteristics of the data are preserved. However, the reconstructed signal is smoother making difficult to visualize the ER pattern.

Considering the whole system itself and the reasons to reduce the feature vector, visualization and time complexity, dimensionality reduction is not necessary for this approach. The visualization can be replaced with the analysis of the distances to the decision boundary, where the histogram can be used to analyze if the data is normally distributed, how close are the samples and if there is overlapping of the data (considerable number of cases with distances to the boundary close to zero). Time complexity of the classification models is reduced, using data in a reduced feature space, depending on the classifier method. However, dimensionality reduced features did not improved the performance of the detection and the physiological morphology of the prototypes is somehow lost because the reconstructed signals are smoother than the originals.

7. CONCLUSION

A method to automatically detect ER in ECG signals was developed. The algorithm used a prototype-based feature vector extracted from the ECG signal fragment where the ER is located. Results showed that the detection of ER in a dataset of more than five thousands subjects was highly sensitive and specific. Therefore, it proves that the signals can be used directly as the features for the classification with very good results.

The automatic method here presented could be used as a prescreening tool of ER and it provides an additional identification of critical cases, based on the distances to the classifier decision surface. Medical evaluation of positive and critical cases is still recommended in order to corroborate the detection. However, an automatic method like this one, can reduce the workload on the number of subjects to analyze.

Preprocessing of the signal and extraction of the features was critical in order to get good classification results. The features for this type of application should be discriminative between classes, but reliable and similar between the same class. The performance on the classification models would depend on the quality and power of the features. In this study, the waveform prototype-based features proposed showed very good performance with three different supervised classifiers. Thus, the prototypes are clearly discriminating between subjects with and without ER, can train different types of classification models accurately, and can be used for the automatic detection of ER.

An additional objective of this research, was to proposed an evaluation tool of critical cases based on the distances to the classifier boundary region. This evaluation allows to find cases that are close to the Jp threshold and are difficult to detect even manually.

8. REFERENCES

[1] Haïssaguerre M., Derval N., Sacher F., Jesel L., Deisenhofer I., de Roy L., Pasquié J.L., Nogami A., Babuty D., Yli-Mayry S., De Chillou C., Scanu P., Mabo P., Matsuo S., Probst V., Le Scouarnec S., Defaye P., Schlaepfer J., Rostock T., Lacroix D., Lamaison D., Lavergne T., Aizawa Y., Englund A., Anselme F., ONeill M., Hocini M., Lim K.T., Knecht S., Veenhuyzen G.D., Bordachar P., Chauvin M., Jais P., Coureau G., Chene G., Klein G.J. & Clémenty J. (2008) Sudden cardiac arrest associated with early repolarization. The new England journal of medicine 358, pp. 2016–2023.

[2] Macfarlane P.W., Antzelevitch C., Haissaguerre M., Huikuri H.V., Potse M., Rosso R., Sacher F., Tikkanen J.T., Wellens H. & Yan G.X. (2015) The early repolarization pattern: a consensus paper. Journal of the American College of Cardiology 66, pp. 470–477.

[3] Patton Kristen K.and Ellinor P.T., Ezekowitz Michael Kowey P., Lubitz S.A., Perez Marco Piccini J., Turakhia Mintu Wang P., Viskin S. & American Heart Association Electrocardiography and Arrhythmias Committee of the Council on Clinical Cardiology and Council on Functional Genomics and Translational Biology (2016) Electrocardiographic early repolarization. a scientific statement from the american heart association. Circulation 133, pp. 1520–1529.

[4] Kenttä T., Porthan K., Tikkanen J.T., Väänänen H., Oikarinen L., Viitasalo M., Karanko H., Laaksonen M. & Huikuri H.V. (2015) Sensitivity and specificity of automated detection of early repolarization in standard 12-lead electrocardiography. Annals of Noninvasive Electrocardiology 20, pp. 355–361.

[5] Macfarlane P.W. & Clark E.N. (2013) Ecg measurements in end qrs notching and slurring. Journal of electrocardiology 46, pp. 385–389.

[6] Clark E.N., Katibi I. & Macfarlane P.W. (2014) Automatic detection of end qrs notching or slurring. Journal of electrocardiology 47, pp. 151–154.

[7] Ma M., Shao M., Zhao X. & Fu Y. (2013) Prototype based feature learning for face image set classification. In: Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on, IEEE, pp. 1–6.

[8] Sansone M., Fusco R., Pepino A. & Sansone C. (2013) Electrocardiogram pattern recognition and analysis based on artificial neural networks and support vector machines: a review. Journal of healthcare engineering 4, pp. 465–504.

[9] Green J. M. .C.A.J. (2014) 12-Lead EKG Confidence, Third Edition : A Step-By-Step Guide. Springer Publishing Company.

[10] McSharry P. A.F..C.G.D. (2006) Advanced Methods and Tools for ECG Data Analysis. Artech House engineering in medicine biology series, Artech House, Inc.

[11] Textbook Equity Edition (2014) Anatomy and Physiology Volume 2 of 3. Textbook Equity.

[12] Rowlands A. .S.A. (2014) The ECG Workbook. 3rd revised edition, M&K Update Ltd.

[13] Baltazar R.F. (2009) Basic and Bedside Electrocardiography. Lippincott Williams Wilkins.

[14] Bayés de Luna A. (2012) Clinical Electrocardiography: A textbook. John Wiley Sons, fourth edition ed. URL: https://books.google.fi/books?id=ttitnBt8IJ0C&.

[15] WikiHow, Cómo calcular la frecuencia cardiaca a partir de un ecg. URL: https://es.wikihow.com/calcular-la-frecuencia-cardiaca-a-partir-de-un-ECG.

[16] Malmivuo J. & Plonsey R. (1995) Bioelectromagnetism - Principles and Applications of Bioelectric and Biomagnetic Fields. Oxford University Press.

[17] Macfarlane P.W., The new definition of early repolarisation. URL: http://www.siacardio.com/wp-content/uploads/2015/01/Repolarizacion-precoz.-English.pdf.

[18] Deo R. & Albert C.M. (2012) Epidemiology and genetics of sudden cardiac death. Circulation 125, pp. 620–637.

[19] Tikkanen J. (2013) Early repolarization in the inferolateral leads of the electrocardiogram. Acta Universitatis Ouluensis. Oulu: Juvenes Print URL: http://jultika.oulu.fi/files/isbn9789526202310.pdf.

[20] Aagaard P., Shulman E., Di Biase L., Fisher J.D., Gross J.N., Kargoli F., Kim S.G., Palma E.C., Ferrick K.J. & Krumerman A. (2014) Prognostic value of automatically detected early repolarization. American Journal of Cardiology 114, pp. 1431–1436.

[21] Kotsiantis S.B., Zaharakis I. & Pintelas P. (2007) Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering 160, pp. 3–24.

[22] Maglogiannis I.G. (2007) Emerging artificial intelligence applications in computer engineering: real word ai systems with applications in ehealth, hci, information retrieval and pervasive technologies, vol. 160. Ios Press.

[23] Prusty M.R., Chakraborty J., Jayanthi T. & Velusamy K. (2014) Performance comparison of supervised machine learning algorithms for multiclass transient classification in a nuclear power plant. In: International Conference on Swarm, Evolutionary, and Memetic Computing, Springer, pp. 111–122.

[24] Xanthopoulos P., Pardalos P.M. & Trafalis T.B. (2013) Linear discriminant analysis. In: Robust data mining, Springer, pp. 27–33.

[25] Fisher R.A. (1936) The use of multiple measurements in taxonomic problems. Annals of human genetics 7, pp. 179–188.

[26] Maroco J., Silva D., Rodrigues A., Guerreiro M., Santana I. & de Mendonça A. (2011) Data mining methods in the prediction of dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. BMC research notes 4, p. 299.

[27] Mathworks (2017), Fit discriminant analysis classifier. URL: `https://se.mathworks.com/help/stats/fitcdiscr.html`.

[28] Demšar J. (2006) Statistical comparisons of classifiers over multiple data sets. Journal of Machine learning research 7, pp. 1–30.

[29] Al-Ani A., Alsukker A. & Khushaba R.N. (2013) Feature subset selection using differential evolution and a wheel based search strategy. Swarm and Evolutionary Computation 9, pp. 15–26.

[30] Garcia S. & Herrera F. (2008) An extension on"statistical comparisons of classifiers over multiple data sets"for all pairwise comparisons. Journal of Machine Learning Research 9, pp. 2677–2694.

[31] Burges C.J. et al. (2010) Dimension reduction: A guided tour. Foundations and Trends® in Machine Learning 2, pp. 275–365.

[32] Kortelainen J., Vayrynen E. & Seppanen T. (2011) Isomap approach to eeg-based assessment of neurophysiological changes during anesthesia. IEEE Transactions on Neural Systems and Rehabilitation Engineering 19, pp. 113–120.

[33] Tenebaum J., De Silva V. & Langford J. (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290, pp. 2319–2323.

[34] Specht D.F. (1991) A general regression neural network. IEEE transactions on neural networks 2, pp. 568–576.

[35] Hannan S.A., Manza R. & Ramteke R. (2010) Generalized regression neural network and radial basis function for heart disease diagnosis. International Journal of Computer Applications 7, pp. 7–13.

[36] Bauer M.M. (1995), General regression neural network for technical use.

[37] Astola J. & Yaroslavsky L. (2007) Advances in signal transforms: theory and applications, vol. 7. Hindawi Publishing Corporation.

[38] Daskalov I. & Christov I. (1999) Electrocardiogram signal preprocessing for automatic detection of qrs boundaries. Medical Engineering and Physics 21, pp. 37–44.

[39] Pan J. & Tompkins W.J. (1985) A real-time qrs detection algorithm. IEEE transactions on biomedical engineering , pp. 230–236.

[40] Welvaert M. & Rosseel Y. (2013) On the definition of signal-to-noise ratio and contrast-to-noise ratio for fmri data. PloS one 8, p. e77089.

[41] Perner P. (2008) Prototype-based classification. Applied Intelligence 28, pp. 238–246.

[42] Dougherty G. (2012) Pattern recognition and classification: an introduction. Springer Science & Business Media.

[43] Kittler J., Hatef M., Duin R.P. & Matas J. (1998) On combining classifiers. IEEE transactions on pattern analysis and machine intelligence 20, pp. 226–239.

[44] He H. & Garcia E.A. (2009) Learning from imbalanced data. IEEE Transactions on knowledge and data engineering 21, pp. 1263–1284.