Extraction of Morphological QRS-based Biomarkers in Hypertrophic Cardiomyopathy for Risk Stratification using L1 Regularized Logistic Regression

A Lyon[1], A Mincholé[1], R Ariga[2], P Laguna[3], S Neubauer[2], H Watkins[2], N de Freitas[1], B Rodríguez[1]

[1]Department of Computer Science, University of Oxford, Oxford, UK
[2]Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK
[3]Biomedical Signal Interpretation & Computational Simulation (BSICoS) Group, University of Zaragoza, CIBER-BBN, Spain

Abstract

Hypertrophic cardiomyopathy (HCM) is an inherited cardiac disease characterized by an unexplained thickening of the heart ventricles. It is the first cause of sudden cardiac death in young adults. No reliable biomarkers for risk assessment have been presented so far, but the electrocardiograms of HCM patients are often abnormal due to structural and electrical abnormalities. The goal of our study was to extract morphological QRS biomarkers in order to discriminate between HCM patients and control patients by analyzing fifty 12-lead Holter recordings (29 HCM – 21 control). Morphological features such as QRS width or slopes from the QRS complex directly and the coefficients of the first four Hermite transform basis were extracted. Classification was then performed using those features in an L1 regularized logistic regression algorithm. Classification between control and HCM patients reached 95.7% of accuracy (sensitivity of 94.96% for HCM and specificity of 96.90%) using only two main features: the percentage of negative regions of the QRS complex with respect to the isoelectric level and the 3rd coefficient of its Hermite fitting showing interesting connections to cardiac electrophysiology.

1. Introduction

Hypertrophic cardiomyopathy (HCM) is a genetic disease characterized by the thickening of the heart ventricles (usually the left one) and is one of the major causes of sudden cardiac death in young adults [1]. It affects 1 in 500 people in the UK but reliable biomarkers from the electrocardiographic signal to distinguish HCM patients from control ones are still to be studied and the current risk assessment factors have a low sensitivity for predicting complications such as heart failure or sudden cardiac death [2] [3]. HCM patients present abnormalities in the myocardial structure such as wall thickening or fibre disarray that may affect the conduction properties of the heart. These abnormalities in the electrical activation pattern are likely to be highlighted by abnormalities in the morphology of the QRS complex in the ECG. Besides QRS width, amplitude and slopes, we also computed features based on the Hermite transform. Indeed, Hermite functions have been widely used in the literature to characterize the QRS complex and denoise it [4] [5]. Therefore, the goal of our work is to analyse QRS biomarkers extracted from Holter recordings in order to perform classification and discover the features that have the most discriminative power between HCM patients and healthy volunteers.

2. Methods

2.1. Database

Fifty 12-lead Holter ECG recordings (29 HCM patients and 21 control patients) with a sampling frequency of 1 kHz were analysed in order to extract QRS morphology biomarkers. The ECG time series were delineated using the wavelet delineator [6] in order to extract QRS peaks and waveforms limits useful to the computation of the biomarkers. After this segmentation step, the signal was filtered in order to get rid of the baseline wandering and powerline inference. Beats with low signal-to-noise ratio were also removed. The remaining delineated beats were aligned with respect to the QRS complex by Woody's method [7] in order to compute a representative mean QRS of length 130ms over 30-minute excerpts of each lead for each patient.

2.2. QRS biomarkers extraction

In order to perform classification, several representative features were extracted from the signal to compose a feature vector.

2.2.1. Morphological biomarkers

As mentioned earlier, we hypothesize that structural changes in the heart of HCM patients induce abnormalities in the activation sequence and in turn, in the QRS morphology. In order to quantify these changes, several morphological biomarkers were computed directly from the ECG signal including:

- the QRS amplitude: computed as the absolute value of the difference between the maximum and minimum points of the QRS for each lead.
- the QRS width: computed as the time interval between the beginning and the end of the QRS.
- the maximum ascending/descending or descending/ascending slopes of the QRS complex: computed as the maximum ascending and descending slopes, depending on the QRS morphology, for all patients in each lead.
- the percentage of the positive and negative regions in the QRS: computed as the percentage of negative time samples with respect to the isoelectric level.

2.2.2 Hermite transform

In the previous section, we presented features computed directly from the signal. Then, we computed the Hermite transform of the ECG signal. It acts as a noise removal method that can strengthen the classification [5], and most importantly, it can compactly describe the morphology of the QRS thanks to its well QRS fitted basis representation. Indeed, it has been shown that only three Hermite functions enable to recover more than 98% of the QRS signal energy [4].
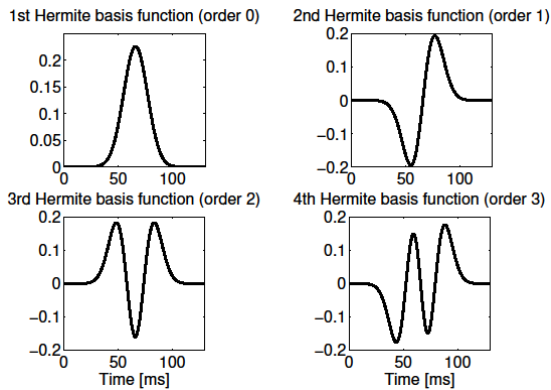

Figure 1. First four Hermite functions (order 0,1,2,3).

In this work, we have used four Hermite functions (figure 1) to reconstruct the QRS signal. Each QRS can then be approximated as a linear combination of the four basis functions.

$$\forall t \in [1,2,\dots,T], \hat{x}(t) = \sum_{n=1}^{4} w_n\, b_n(t)$$

where $\hat{x}(t)$ is the reconstructed QRS signal, $b_n(t)$ the Hermite function of order n, t the time samples from 1 to $T = 130$ms and w_n its coefficients in the linear combination.

One hypothesis motivating the study of the Hermite transform of the signal is that it may capture the degree of abnormality of the QRS morphology. Indeed, we expect the Hermite fitting error (difference between the reconstructed signal and the original one) to be higher in HCM patients due to their more abnormal QRS morphology. Figure 2 presents an example of a reconstructed QRS complex in a control and an HCM patient.
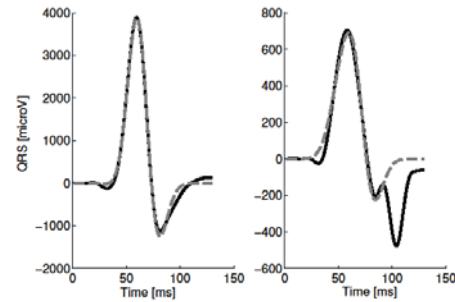

Figure 2. Original QRS complex (plain line) and reconstructed one by Hermite fitting (dash line) in (right) a control patient and (left) a HCM patient, both in lead V3.

The features computed from the Hermite fitting were:
- the coefficients of the first four Hermite basis
- the fitting error between the reconstructed QRS and the original one in terms of the root mean square (RMS) error.

2.3. Machine learning algorithms

Classification was performed using a supervised machine learning algorithm: logistic regression [8] [9]. It is a probabilistic model, maximizing the probability of one instance belonging to one class. In this work, in order to reduce the number of selected features with redundant information, L1 regularization was used.

The feature selection process was then performed using the logistic regression L1 regularized algorithm (Lasso technique [10]). The logistic regression aims at maximizing the conditional log-likelihood of the data $\mathcal{L}$:

$$\mathcal{L}(\boldsymbol{w}) = \sum_n \log p(y_n|\boldsymbol{x_n})$$

where y_n is the output label for the point $\boldsymbol{x_n}$, and the probability of $\boldsymbol{x_n}$ to belong to class $C1$ is $p(C1|\boldsymbol{x_n}) = \sigma(w_n\boldsymbol{x_n})$ where σ is the sigmoid function and w_n the weights of the linear model.

When adding the regularization term, the regularized objective function to minimize becomes:

$$\mathcal{L}(\boldsymbol{w}) = \sum_n \log p(y_n|\boldsymbol{x_n}) + \alpha \sum_i |w_i|$$

where α is the penalty term controlling the trade-off between the classification performance and the size of the weights w_i. With the L1 regularization term, the number of features is reduced creating sparse models, and in turn, it enables interpretability of the results and it reduces overfitting. Indeed, gradually decreasing the penalty term lets appear the most discriminative features of the model.

In our experiments, the dataset was split into a training set (40 patients) for learning and a testing set (10 patients) for performance measurement. The algorithm was run 100 times with randomly selected training and testing sets.

3. Results and discussion

During the classification, standard biomarkers such as QRS width, QRS amplitude or QRS slopes were investigated but they did not add complementary information to the classification. This confirms previous observations on these biomarkers discriminative power ([3] [2]). Table 1 presents the accuracy results obtained for different biomarkers combinations.

Table 1. Accuracy results for different biomarkers using the eight leads.

Biomarkers combination	Accuracy
Amplitude of QRS + QRS width + ascending and descending slope	74%
Error of the Hermite fitting + time location of the maximum error	80.08%
Four Hermite coefficients in each lead	81.25%
Percentage of negative regions in each lead	91%

In Table 2, we included the L1 regularization term for feature selection to obtain a ranking of the features according to their weight in the classification. Table 2 shows the good discriminative power of the Hermite coefficients and the importance of the QRS negative regions in the classification.

Table 2. Ranking of the first four best features.

N.	Feature	Logistic Regression (cumulative accuracy %)
1	% Negative regions in V3	83.51
2	3rd Hermite coeff. in V3	95.78
3	Negative regions in V5	94.05
4	1st Hermite coeff. in V4	94.61

We extracted the first two features:
- the percentage of negative regions in the QRS in lead V3
- the 3rd coefficient of the Hermite fitting in lead V3

Using only these two biomarkers, the classification reached an accuracy of 95.7% with a sensitivity in predicting HCM of 94.96% and a specificity of 96.90%. Adding extra features did not add additional/complementary information to the classifier.

Figure 3 shows the discriminative power in terms of the Student's t-test p-values of six biomarkers (negative regions, 3rd and 1st Hermite coefficients, QRS width, QRS descending slope and RMS error of the fitting) in leads V2, V3, V4 and V5. The QRS width, the descending slope and the RMS did not show any significant difference between HCM and control patients in any lead but V3. On the contrary, the 3rd and 1st Hermite coefficients and the negative regions of the QRS showed p-values$<10^{-6}$ in most of the leads (especially V3 and V4), showing their good discriminative power to classify between HCM and control patients.

These findings connect to cardiac electrophysiology in two ways. First, it is interesting to note that the larger percentage of negative regions in the QRS of HCM patients is really discriminant in all precordial leads. This is in agreement with the ventricular remodeling affecting the septum and the left ventricle in HCM patients [11]. Then, the role played by the coefficients of the Hermite functions highlights the importance of the QRS morphology in HCM abnormalities. Indeed, as shown in Fig.3, the 3rd Hermite coefficient in V3 has a higher absolute weight in HCM patients, while the 1st coefficient is more correlated to the control ones. The wavy and irregular shape of the 3rd coefficient (Fig. 1) is in agreement with the presence of notches and irregularities in the QRS of HCM patients, while the smoother shape of the 1st coefficient agrees with the controls.

Some future work should consider a multi-class classification introducing for example ischemic, with abnormalities in the QRS complex or control athletes, with a thicker heart. More complex classification methods could also be investigated.
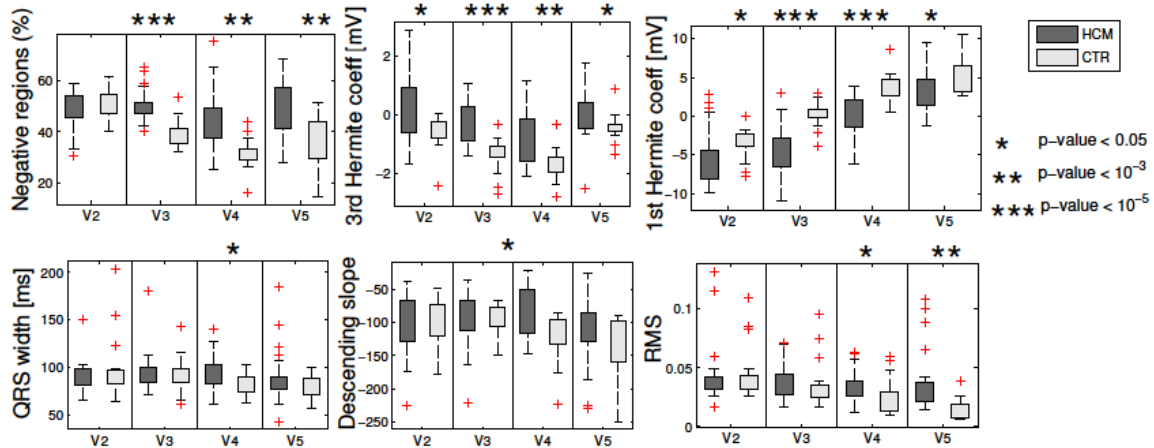
Figure 3. Boxplot of HCM and control patients of QRS width (ms), 3rd Hermite coefficient, negative regions (%), 1st Hermite coefficient, descending slope and RMS, all in leads V2, V3, V4, V5, along with the p-value.

4. Conclusion

Holter recordings from 29 HCM patients and 21 control ones were analyzed to quantify morphological features for HCM classification. Classification results focusing on the morphology of the QRS complex achieved 95.8% of accuracy and led to the extraction of two main features to distinguish between the two groups of patients: the percentage of negative regions of the QRS and the 3rd coefficient of the Hermite fitting, both in lead V3. These findings highlight the importance of the morphology of the QRS complex in understanding the HCM disease.

Acknowledgements

References

[1] B. J. Maron, "Hypertrophic cardiomyopathy: a systematic review," *JAMA*, vol. 287, no. 10, pp. 1308–1320, Mar. 2002.

[2] C. A. Dumont, L. Monserrat, R. Soler, E. Rodríguez, X. Fernandez, J. Peteiro, A. Bouzas, B. Bouzas, and A. Castro-Beiras, "Interpretation of electrocardiographic abnormalities in hypertrophic cardiomyopathy with cardiac magnetic resonance," *Eur. Heart J.*, vol. 27, no. 14, pp. 1725–31, 2006.

[3] R. C. Saumarez, "Electrophysiological investigation of patients with hypertrophic cardiomyopathy," *Br. Heart J.*, vol. 72, no. 6 Suppl, pp. S19–S23, Dec. 1994.

[4] L. Sornmo, P. Ola Borjesson, M.-E. Nygards, and O. Pahlm, "A Method for Evaluation of QRS Shape Features Using a Mathematical Model for the ECG," *IEEE Trans. Biomed. Eng.*, vol. BME-28, no. 10, pp. 713–717, Oct. 1981.

[5] L. L. Campbell, "On the Use of Hermite Expansions in Noise Problems," *J. Soc. Ind. Appl. Math.*, vol. 5, no. 4, pp. 244–249, Dec. 1957.

[6] J. P. Martínez, R. Almeida, S. Olmos, A. P. Rocha, and P. Laguna, "A wavelet-based ECG delineator: evaluation on standard databases," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 4, pp. 570–581, 2004.

[7] L. Sörnmo and P. Laguna, *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Burlington: Academic Press, 2005.

[8] L. Breiman, "Statistical Modeling: The Two Cultures," *Stat. Sci.*, vol. 16, no. 3, pp. 199–231, Aug. 2001.

[9] R. Caruana and A. Niculescu-Mizil, "An Empirical Comparison of Supervised Learning Algorithms Using Different Performance Metrics," in *In Proc. 23 rd Intl. Conf. Machine learning (ICML'06*, 2005, pp. 161–168.

[10] R. Tibshirani, "Regression Shrinkage and Selection Via the Lasso," *J. R. Stat. Soc. Ser. B*, vol. 58, pp. 267–288, 1994.

[11] R. A. Nishimura, S. R. Ommen, and A. J. Tajik, "Hypertrophic Cardiomyopathy A Patient Perspective," *Circulation*, vol. 108, no. 19, pp. e133–e135, Nov. 2003.

Address for correspondence.

Aurore Lyon
Department of Computer Science, University of Oxford, Wolfson Building
aurore.lyon@cs.ox.ac.uk