

Ensemble Classifier based on Linear Discriminant Analysis for distinguishing Brugada Syndrome Patients according to Symptomatology

Daniel Romero¹, Mireia Calvo¹, Nathalie Béhar², Philippe Mabo², Alfredo Hernández³

¹ Laboratoire Traitement du Signal et de l'Image (LTSI). Université de Rennes 1. France

² Cardiology and Vascular Disease Division, Rennes University Health Centre, France

³ INSERM U-01099, Rennes, France

Abstract

Identifying high-risk patients requiring an ICD among asymptomatic Brugada patients is nowadays a bit challenging. In this study, 62 patients suffering from Brugada syndrome (14 symptomatic) were studied by analyzing the 12-lead ECG recording acquired during a physical exercise test. For each patient, conventional HRV indices from time-frequency analysis and heart rate recovery (HRV features), as well as several morphological depolarization indices (QRS features), were evaluated at relevant periods of the test. Most discriminant features from both the HRV and QRS sets were selected using a two-stage feature selection algorithm and used for model classification building. For the detection step, an ensemble classifier using stacking approach plus a fixed combiner was designed, using linear discriminant analysis as the base classification algorithm. Best features from each model were then used for building the final individual and combined classification models. Detection performance using the symptomatic group as the target class, was as follows: HRV-based model: $Se=1$, $Sp=0.67$, $AUC=0.87$; QRS-based model: $Se=75$, $Sp=0.67$, $AUC=0.73$. When joining best features of both models (HRV-QRS-based model), the performance increased up to $Se=1$, $Sp=0.83$, $AUC=0.90$. The study showed that by combining both HRV and depolarization analysis, a better risk stratification can be performed. This could be useful for the identification of Brugada patients with previous symptoms, and it may help to the decision making process of asymptomatic patients needing an ICD.

1. Introduction

Brugada syndrome (BrS) is a genetic pathology, firstly described more than 20 years ago, and associated with a high risk for sudden cardiac death (SCD) in patients with an apparently normal structural heart. The coved type-1 ECG pattern (ST elevation ≥ 2 mm, negative symmetric T wave, etc) in ≥ 1 lead of the right precordial leads V1-

V3 is the only type that can had a BrS major diagnostic marker. [1].

The main challenges regarding BrS are related to patient risk stratification and the definition of the best treatment approach. The cardioverter defibrillator (ICD) implantation is the only validated treatment recommended for symptomatic patients. However, this decision is more complex on asymptomatic patients, who represent between 60% of the BrS population. The difficulty is thus to identify which of the asymptomatic patients may benefit from an ICD implantation.

Depolarization disorders have been one of the two main hypotheses underlying the pathophysiology of BrS. They are associated with slowing conduction within the right ventricular outflow tract (RVOT) [2]. Moreover, the role of the autonomic control, seems to have an important diagnostic component in this pathology.

In this study we aimed at designing a classifier able to distinguish Brugada syndrome patients according to symptomatology. To do this, we extracted a set of classical and unconventional QRS morphological features, as well as features derived from HRV analysis. These two sets of features were used for building two different Brugada models. Overall performances were assessed individually and compared with the model obtained from merging the two alternative approaches.

2. Materials and methods

2.1. Population

The study population comprises 62 patients suffering from Brugada syndrome enrolled in the multicenter PHRC BRUGADA study, led by the department of Cardiologie et Maladies Cardiovasculaires at CHU Rennes. From the total population, 14 patients have had previous events such as syncope, VF or sudden cardiac death (SCD) and were classified as symptomatic patients. The remaining 48 patients were thus asymptomatic. During a physical exercise test, the standard 12-lead ECG recording sampling at 1000

Hz was acquired for each patient. The clinical protocol used for this particular test is described below:

- Warm-up phase: 2 minutes of initial workload by pedaling at 50 W (30 W for women).
- Exercise phase: consecutive periods of incremental effort of 30 W (20 W for women) every 2 minutes until reaching at least 80% of the maximal theoretical heart rate, defined by $HR_{max} = 220 - age$.
- Active recovery phase (ARP) : 3 minutes pedaling with the initial workload of 50 W.
- Passive recovery phase (PRP): 3 minutes at rest.

2.2. Preprocessing

All ECG signals were preprocessed before the automatic extraction of the analyzed indices. This included automatic QRS complexes detection and subsequent visual inspection, baseline drift attenuation via cubic spline interpolation, 4-th order Butterworth low pass filtering at 45 Hz to remove muscular noise and wave delineation using an evolutionary optimization approach [3].

2.3. Depolarization parameters

Several ECG features including classical QRS parameters as well as a set of unconventional QRS parameters, were evaluated through the entire exercise test and for each standard lead in a beat-to-beat basis. In the classical group we have included the R and S waves amplitude (R_a and S_a), the QRS duration (QRS_d) and the maximum vector magnitude (VM). In the unconventional group, we have evaluated the main three QRS slopes, defined as the up-stroke (R_U) and down-stroke (R_D) of the R wave and the up-stroke of the S wave (S_U) [4], as well as the angles formed by projecting the lines associated with the above-mentioned slopes, denoted as the angles (ϕ_R) and (ϕ_S) [5].

After computing the complete set of QRS features, the delta change, $\Delta\mathcal{Y}_{EX}$, achieved between the time of maximum effort peak (T2) and the beginning of the exercise (T1) was evaluated, being $\mathcal{Y} = \{QRS_d, VM, R_a, S_a, R_U, R_D, S_U, \phi_R, \phi_S\}$. Likewise, the delta change, $\Delta\mathcal{Y}_{RE}$, achieved between the end of recovery (T3) and T2 was also evaluated to assess the amount of change occurring during the complete recovery phase. Moreover, both $\Delta\mathcal{Y}_{EX}$ and $\Delta\mathcal{Y}_{RE}$ were corrected by the heart rate changes occurring during the same periods, ΔHR_{EX} and ΔHR_{RE} , and denoted as $\hat{\Delta}\mathcal{Y}_{EX}$ and $\hat{\Delta}\mathcal{Y}_{RE}$, respectively. Relative changes, expressed as $\mathcal{R}_{\mathcal{Y}} = \frac{(\mathcal{Y}_t - \mathcal{Y}_{t-1})}{\mathcal{Y}_{t-1}}$, with $t = \{T1, T2, T3\}$, were also calculated.

2.4. HRV parameters

Time-frequency approach: The HRV analysis was conducted using a time-frequency (TF) approach due to the non-stationary nature of the available signals, acquired

during a physical stress test. First, a cubic splines interpolation at 4 Hz was applied to have RR series regularly sampled, denoted as $x_{RR}(t)$. Then, a bidirectional, 4th order high-pass Butterworth filter was applied with a cut-off frequency of 0.03 Hz to remove the very low frequency components.

The smoothed pseudo Wigner-Ville distribution (SPWVD) transform, available from the Time-Frequency toolbox [6], was subsequently used since it has proved to be useful for cardiovascular signals analysis. This quadratic TF approach is defined as the Fourier transform of the instantaneous autocorrelation function. However, because of the fact that it is affected by significant interference terms, the SPWVD introduces a smoothing kernel function $\Psi(\tau, v)$ that attenuates such interferences terms while maintaining a suitable time-frequency resolution. Let $A_{RR}(\tau, v)$ be the ambiguity function of the signal x_{RR} , the SPWVD is defined as:

$$A_{RR}(\tau, v) = \int_{-\infty}^{\infty} x_{RR}\left(t + \frac{\tau}{2}\right) \left(t - \frac{\tau}{2}\right) e^{-j2\pi vt} dt \quad (1)$$

and

$$\Psi(\tau, v) = exp \left\{ -\pi \left[\left(\frac{v}{v_o} \right)^2 + \left(\frac{\tau}{\tau_o} \right)^2 \right]^{2\lambda} \right\} \quad (2)$$

then, the Cohen's class distribution is defined by:

$$C_{RR}(t, f) = \int \int \Psi(\tau, v) A_{RR}(\tau, v) e^{j2\pi(tv - \tau f)} dv d\tau \quad (3)$$

HRV was assessed as the total power in LF (0.04 - 0.15 Hz) and HF (0.15 - 0.4 Hz) bands obtained from the SPWVD, where the Wigner-Ville distribution was computed from the RR series.

$$LF(t) = \int_{LF} C_{RR}(t, f) df \quad (4)$$

$$HF(t) = \int_{HF} C_{RR}(t, f) df. \quad (5)$$

SPWVD leads to time-frequency HRV estimates representing time series that vary during the execution of the stress test. These features, accounting for the sympathetic and parasympathetic activity of the ANS, were normalized by the total power (TP), defined as the sum of both spectral bands (TP = LF + HF), leading to the time series $LF_{nu}(t)$ and $HF_{nu}(t)$. The ratio between LF and HF in the time-frequency domain, $\frac{LF}{HF}(t)$, was also obtained so as to assess the global sympatho-vagal balance.

Finally, the ANS-related parameters $LF_{nu}(t)$, $HF_{nu}(t)$ and $\frac{LF}{HF}(t)$, were averaged using non-overlapped windows

of 1-min duration for each individual patient, leading to $LF_{nu}(i)$, $HF_{nu}(i)$ and $\frac{LF}{HF}(i)$, which stand for the intrapatient mean evaluated at minute i .

Heart rate recovery: The heart rate recovery (HRR), usually defined as the heart rate decay (beats/min) during the first minute immediately after the exercise cessation, was also calculated for all patients. Moreover, the HRR at minute 2nd and 3rd, were also evaluated in this study to further assess both the parasympathetic reactivation and sympathetic withdrawal during the recovery phase.

2.5. Features selection procedure

Once all features were extracted for each patient and group, an hybrid feature selection approach was applied to select the most relevant features for classifying BrS patients. This hybrid approach consists of two steps. The first one is a simple filter algorithm, applied to quickly remove those features of less importance by ranking them according to the RELIEF algorithm [7]. The second step, applied on the reduced set of features obtained in the first step, is a sequential floating feature selection (SFFS) method [8]. SFFS performs a series of successive forward/backward searches gradually improving the best feature subset found. We also used the *wrapper* strategy during SFFS, which uses further classification algorithm as the evaluation function to compare different combinations of features in terms of model accuracy.

2.6. Ensemble classifier (LDA + combiner)

After having obtained the best set of features for each model, we implemented an ensemble or combined classifier based on two different stages in order to achieve the best possible detection performance. The first stage consisted in training several “base” classifiers (with LDA algorithm) using the stacked generalization approach [9], which basically perform a k -fold cross-validation type selection. In brief, the entire training subset is divided into k blocks, and each base classifier C_b^k is first trained on $k-1$ blocks of the training subset Tr_b^k . Then, each base classifier is evaluated on the k -th block Te_b^k not seen during training. The outputs of individual base classifiers O_b^k (soft outputs or posterior probabilities) are then combined using a simple fixed rule (*mean, product, maximum* or *minimum*) in the second stage of the ensemble classifier for reducing classification error and the overall risk of making a particularly poor selection. Figure 1 illustrates the stacking procedure using 3 folds, although we set $k=10$ in this study.

2.7. Performance evaluation

The initial dataset was randomly split into training and testing subsets using 75% and 25% of the data, respec-

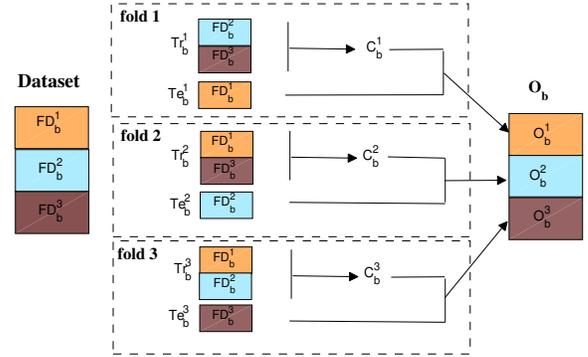


Figure 1. Diagram of the first part (*stacking*) of the ensemble classifier. FD_b^k : features dataset split in k blocks; Tr_b^k , Te_b^k : training and testing sets of fold k ; C_b^k base classifiers; O_b^k soft outputs used as the input data of combiner.

tively. The best set of features obtained for each model (QRS and HRV models) were used individually, and in conjunction, to train the ensemble classifier described in the above section. The classification performance was assessed for each individual trained classifier (the QRS- or HRV-based model) using the test subset, as well as for the global classifier that combines all the best features from both groups (QRS-HRV-based model). As performance metrics, we calculated the area under the ROC curve (AUC), as well as the sensitivity (Se) and specificity (Sp) associated with the optimum operating point in the receiver operating characteristic (ROC) curves.

3. Results

The final two Brugada models derived after applying the feature selection step are displayed in Table 1. For the HRV-based model, 16 features were selected from a total of 73. Two-thirds of the LF/HF and HFnu related features were associated with the recovery (RE) and post-recovery (PR) periods of the exercise test, which is also noted for features related to LFnu. Regarding the QRS-based model, only 11 relevant features were selected from a total of 304. Most of them were related to leads with remarkable positive QRS morphology such as leads II, aVF, V5 and V6. Lead V3 was also relevant for the markers ϕ_S and S_U , which are better defined in QRS complexes with QS and rS morphologies.

The ROC curves associated with the final two Brugada models presented in Table 1 are displayed in figure 2, together with the ROC curve corresponding to the model combining both the QRS and HRV selected features. As it can be observed, the overall performance of the HRV model (AUC=87%, Se=100%, Sp=67%) is superior to that of the QRS model (AUC=73%, Se=75%, Sp=67%). Like-

Table 1. Features obtained from each model

Model	Features
HRV model	$(\frac{LF}{HF}nu)^{EX2} - (\frac{LF}{HF}nu)^{RE6} - (\frac{LF}{HF}nu)^{PR4}$ $(\frac{LF}{HF}nu)^{PR7} - (HFnu)^{EX1} - (HFnu)^{EF3}$ $(HFnu)^{RE1} - (HFnu)^{PR3} - (HFnu)^{PR5}$ $(LFnu)^{EX1} - (LFnu)^{EX2} - (LFnu)^{EF1} -$ $(LFnu)^{RE6} - (LFnu)^{PR2} - (LFnu)^{PR5} -$ $(LFnu)^{PR6}$
QRS model	$(R_a)^{T3} - (R_a)^{T3'} - (\phi_R)^{T1} -$ $(\Delta\phi_R)_{aVF}^{EX} - (\Delta\phi_R)_{aVF}^{EX} - (\Delta S_U)_{V3}^{RE} -$ $(\Delta\phi_S)_{V3}^{EX} - (\Delta\phi_S)_{V3}^{RE} - (\phi_R)_{V6}^{T3'} -$ $(R_{RU})_{V5}^{EX} - (R_{SU})_{V4}^{RE}$

wise, when both models are combined, the overall performance of the new generated model (AUC=90%, Se=100%, Sp=83%) is superior to both individual models. All the above-mentioned results were obtained using the *mean* operator as the fixed combiner.

4. Discussion and conclusions

In this study, an ensemble classifier has been implemented to distinguish Brugada syndrome patients according to symptomatology. Morphological features extracted from the QRS complex and HRV markers obtained from a time-frequency approach were used alone and in combination to design three different Brugada models.

The proposed two-stage feature selection approach significantly reduced the final subsets of features used for the QRS- and HRV-based models. From those relevant features selected in the HRV model, two-thirds were related to recovery and post-recovery periods of the exercise test, suggesting that autonomic control (including sympathetic and parasympathetic divisions) assessment might be more appropriate in these particular periods when studying Brugada patients. Moreover, from table 1 it can be noted that parasympathetic related features predominate as compared to the sympathetic related ones. Regarding the QRS model, none of the classical markers except R_a were relevant for the model finally obtained, thereby emphasizing the potential utility of non-conventional QRS markers. Note also that, for most features, better performing results were found for their absolute or relative changes during the exercise periods.

Finally, the overall performances obtained for the three models, suggested that HRV markers are better suited for classifying Brugada patients than the QRS morphological markers. However, when both models are combined in a new one, the performance increases noticeably.

Acknowledgements

D.R. acknowledges the financial support of the Fondation Lefoulon-Delalande, Institut de France, France.

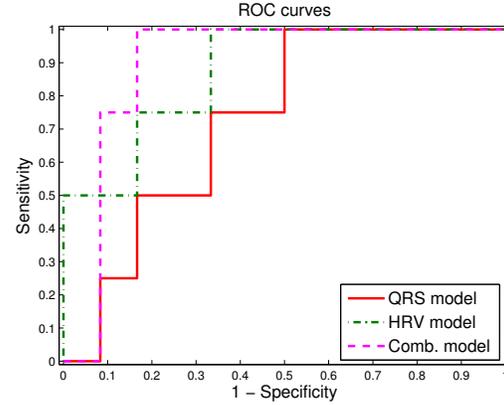


Figure 2. ROC curves for the final HRV- and QRS-based models and for the combination of both.

References

- [1] Sarkozy A, Paparella G, Boussy T, Casado-Arroyo R, Yazaki Y, Chierchia GB, Brugada J. The usefulness of the consensus clinical diagnostic criteria in Brugada syndrome. *International Journal of Cardiology*, 167, 2700-2704.
- [2] Meregalli PG, Wilde A, Tan HL. Pathophysiological mechanisms of Brugada syndrome: Depolarization disorder, repolarization disorder, or more?. *Cardiovasc Res*, 2005; 67:367-378.
- [3] Dumont J, Hernandez A, Carrault G. Improving ECG beats delineation with an evolutionary optimization process. *Biomedical Engineering, IEEE Trans Biomed Eng*, 2010; 57:607-615
- [4] Romero D, Ringborn M, Laguna P, Pahlm O, Pueyo E. Depolarization changes during acute myocardial ischemia by evaluation of QRS slopes: standard lead and vectorial approach. *IEEE Trans Biomed Eng*, 2011, 58:110-120
- [5] Romero D, Ringborn M, Laguna P, Pueyo E. Detection and quantification of acute myocardial ischemia by morphological evaluation of QRS changes by an angle-based method. *J Electrocardiol*, 2013; 46:204-214.
- [6] Auger F, Flandrin P, Gonalvs P, Lemoine O. Time-frequency toolbox. CNRS France-Rice University, 46, 1996.
- [7] Sanchez N, Alonso A, Tombilla M. Filter methods for feature selectiona comparative study. In *International Conference on Intelligent Data Engineering and Automated Learning*. Springer Berlin Heidelberg, 2007; 178-187
- [8] Pudil P, Novoviov J Kittler J. Floating search methods in feature selection. *Pattern recognition letters*, 1994; 15: 1119-1125.
- [9] Kuncheva LI. *Combining pattern classifiers: methods and algorithms*. John Wiley Sons; 2004

Address for correspondence:

Daniel Romero Pérez
 LTSI. Université de Rennes 1. 35042. Rennes
 daniel.romero@univ-rennes1.fr