A Machine Learning Approach for Detection and Quantification of QRS Fragmentation

Griet Goovaerts†, Sibasankar Padhy†, Bert Vandenberk, Carolina Varon, Rik Willems and Sabine Van Huffel

***Abstract*— *Objective:* Fragmented QRS (fQRS) is an accessible biomarker and indication of myocardial scarring that can be detected from the electrocardiogram (ECG). Nowadays, fQRS scoring is done on a visual basis, which is time-consuming and leads to subjective results. This study proposes an automated method to detect and quantify fQRS in a continuous way using features extracted from Variational Mode Decomposition (VMD) and Phase-Rectified Signal Averaging (PRSA). *Methods:* In the proposed framework, QRS complexes in the ECG signals were first segmented using VMD. Then, 10 VMD- and PRSA-based features were computed and fed into well-known classifiers such as Support Vector Machine (SVM), K-nearest neighbors (KNN), Naive Bayesian (NB), and TreeBagger (TB) in order to compare their performance. The proposed method was evaluated with 12-lead ECG data of 616 patients from the University Hospitals Leuven. The presence of fQRS in each ECG lead was scored by five raters. Both detection and quantification of fQRS could be achieved in this way. *Results:* The experimental results indicated that the proposed method achieved AUC values of 0.95, 0.94, 0.90, and 0.89 using SVM, KNN, NB and TB classifiers respectively for detecting QRS fragmentation. Assessment of quantification performance was done by comparing the fQRS score with the total score, obtained by summing the scores from the individual raters. Results showed that the fQRS score clearly correlated with this estimate of fQRS certainty. *Conclusion:* The proposed method obtained good results in both fQRS detection and quantification, and is a novel way of assessing the certainty of QRS fragmentation in the ECG signal.**

***Index Terms*— ECG signal processing, Phase rectified signal averaging, QRS fragmentation, variational mode decomposition, machine learning, support vector machine**

I. Introduction

The electrocardiogram (ECG) is a non-invasive diagnostic tool that is used for diagnosis of various heart conditions. While ECG signals can be single-lead or multilead, in clinical practice 12-lead ECGs are commonly used in order to assess the spatio-temporal dynamics of the heart. There are a number of biological markers that have been used for diagnosis of different cardiac diseases that can be derived from the ECG signal. One of the promising markers in the 12-lead ECG is QRS fragmentation (fQRS). It is defined by Das et al. [1] as the presence of an additional R wave (R') or a notch in the nadir of the S wave in at least two contiguous leads on the 12-lead ECG. Das et al. showed that fQRS can be a convenient marker of myocardial scar, which may lead to high-risk cardiac events like heart failure, need for revascularization, or sudden cardiac death [1], [2]. Other studies have linked the presence of QRS fragmentation to adverse outcome in patients with Brugada syndrome [3] and defibrillator shocks [4].

While the presence of fQRS was first noticed in 1969 by Flowers et al. [5], to the best of our knowledge, there have been few studies that focus on automated fQRS detection [6]–[8]. Until now, visual assessment of each lead individually is considered to be the gold standard in clinical practice. Interpreting ECG signals is however labor-intensive, time-consuming, expensive, and most importantly requires adequate training of clinicians in order to get reliable results. Recently, we have shown that visual fQRS assessment is not ideal as the inter-observer variability differs significantly depending on the expertise level of the observers [9]. Therefore, automated methods for fQRS detection should be considered and can serve as a complementary tool for the clinician.

Currently, fQRS detection is done where each lead is given a score of 0 or 1 depending on the absence or presence of fragmentation. Fragmentation can however take many forms as the number and location of deflections varies. Binary scoring might therefore not be optimal since it fails to capture differences between fQRS subtypes. It is furthermore expected that the spatial and temporal characteristics of the deflections can be important prognostic factors in determining patient outcome [10]. This paper

† These two authors contributed equally to the work.

We thank Bert Vandenberk, Tomas Robyns, Mathias Claeys, Frederik Helsen and Sofie Van Soest for their work on annotating the database. This work was supported by imec funds 2017. European Research Council: The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Advanced Grant: BIOTENSORS (no 339804). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information. G. Goovaerts is a PhD fellow of IWT Vlaanderen. C. Varon is a postdoctoral fellow of the Research Foundation-Flanders (FWO). R. Willems has received research funding from Biotronik, Boston Scientific and Medtronic and speakers and consultancy fees from Biotronik, Boston Scientific, Medtronic, St Jude Medical and Sorin. R. Willems is supported as postdoctoral clinical researcher by the Fund for Scientific Research Flanders.

G. Goovaerts, S. Padhy, C. Varon and S. Van Huffel are with the Department of Electrical Engineering (ESAT), STADIUS, KU Leuven, Belgium, and imec, Leuven, Belgium (Email: {griet.goovaerts, sibasankar.padhy, carolina.varon, sabine.vanhuffel}@esat.kuleuven.be).

B. Vandenberk and R. Willems are with the Department of Cardiovascular Diseases, Experimental Cardiology, KU Leuven, Belgium (Email: bert.vandenberk, rik.willems@uzleuven.be).

therefore proposes an automatic and objective method to determine a fQRS score that quantifies the certainty of the presence of QRS fragmentation in each lead.

For this, a necessary step is an accurate segmentation of the QRS complex. QRS segmentation is a mature technology and nowadays a number of methods have been developed, including time-domain, frequency-domain, time-frequency analysis or transform-domain methods. A broad methodological review on QRS segmentation can be found in [11]. In the past two decades, wavelet transform based methods (such as [12]) have been widely adopted. Recently, empirical mode decomposition (EMD) has also been used to segment the QRS complex [13], [14]. The EMD is an empirical algorithm with lack of a theoretical basis and is sensitive to noise. Additionally, it has been shown that EMD introduces distortions in the beginning and end of the QRS complex, which may cause erroneous results in this specific application [15]. The majority of methods including the wavelet- and EMD-based approaches are non-adaptive and hence cannot be used when signal characteristics change extensively. To overcome these limitations, Dragomiretskiy and Zosso proposed variational mode decomposition (VMD) [16]. VMD is based on the framework of variational theory and adaptively determines the relevant frequency band. In the last four years, it has been successfully used in different fields including ECG applications like arrhythmia characterization [17], [18] or detection of shockable ventricular arrhythmia [19]. In this work, we apply VMD to segment the QRS complex and extract features that help quantify the certainty of the fQRS.

In this paper, machine learning is used to detect and quantify fQRS. In our previous study [20], phase-rectified signal averaging (PRSA) was used to extract features that characterize QRS fragmentation. Here, additional features are computed from the results of VMD and combined in a classifier. The main goal of this paper is to assess whether the combination of this novel feature set with different machine learning approaches succeeds in both detecting and quantifying fQRS in a continuous way. The comprehensive approach conducted in this study has the following main contributions:

- Accurate segmentation of the QRS complex using VMD
- Extracting VMD- and PRSA-based features that characterize the certainty of fragmentation in a lead
- Comparison of different machine learning methods for fQRS detection and quantification
- The assessment of the certainty of the presence of QRS fragmentation as opposed to a binary classification, by assigning a score between 0 and 1.

The remainder of the article is organized as follows. The database used in this work is described in Section II. The different steps required for calculating the fQRS score are described in Section III. Results of the proposed method (including accuracy of the QRS segmentation) with comparison of evaluation performance among different classifiers are presented in Section IV. The experimental results are discussed in V. Finally, conclusions of the proposed work are given in Section VI.

II. Database

For this study, a dataset of 12-lead ECG signals recorded in 723 patients before the implantation of an cardiac defibrillator (ICD) in the prevention of sudden cardiac death due to cardiac arrhythmia was used. All signals were recorded in the University Hospitals Leuven, Belgium. The ECG signals are digitized with a sampling frequency of 250 Hz and have a duration of 10 seconds. ECGs from patients with ventricular pacing or cardiac arrhythmia like atrial fibrillation were excluded since they have altered ECG morphologies. While these patients might exhibit QRS fragmentation, they were insufficiently represented in the dataset in order for the machine learning algorithms to learn their specific ECG characteristics. We therefore decided to remove them. A total of 616 records were included in the final dataset. Clinically, lead aVR is not used for fQRS analysis [4], [9]. This lead was therefore excluded from all records in this analysis. The study was approved by the ethical committee of the University Hospitals Leuven (S56074/ML9965).

The database was fully annotated by five readers on the presence of fQRS in each lead: they individually gave a score of 1 if fQRS was present and 0 otherwise. The persons who scored the signals were all clinicians: four cardiologists in training and one cardiology laboratory technician. Two of the readers were considered experienced observers in fQRS analysis, due to their involvement in previous studies and the three remaining ones were novice in fQRS analysis prior to this project but received training prior to scoring the signals. All five readers are experienced in research. The scores of all raters can be combined by summing them, resulting in a total score. If all readers agree that a lead does not show fQRS, the total score is 0; similarly, the total score is 5 if all readers agree on the presence of fQRS in a lead. Leads where some raters disagree have intermediate values. The total score given by all raters is thus related to the uncertainty of the raters about the presence of fQRS. In other words, a higher score means more raters have scored the signal as fragmented and the fragmentation is assumed to be more prominent in that lead. Signals with certain absence of fragmentation will have lower scores. The frequency of occurrence of the scores (0 to 5) in the dataset is shown in Table I. A full description of the scoring process, including results on inter- and intra-rater variability can be found in [9].

III. Methods

The block diagram of the proposed method for fQRS quantification in ECG signals is depicted in Figure 1. It consists of 4 main steps: preprocessing, QRS segmentation using VMD, feature extraction and classification. Each step is described in detail in the following subsections.

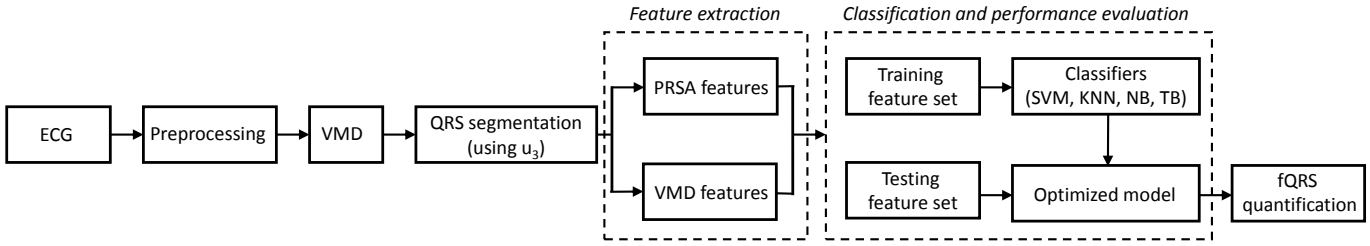

Fig. 1: Block diagram of the proposed method for fQRS detection and quantification. After preprocessing and QRS segmentation using VMD, features were extracted using PRSA and VMD after which a classifier was trained and evaluated. The final fQRS score represents the certainty of QRS fragmentation in an ECG lead.

TABLE I: Frequency of occurrence of different scores, obtained by summing the annotations from all five readers, in the database.

Total score	Counts	Percent
0	2775	40.95
1	894	13.19
2	490	7.23
3	390	5.76
4	553	8.16
5	1674	24.7

A. Preprocessing

In the preprocessing step, baseline wander and high-frequency noise were removed from ECG signals. Baseline wander was removed by passing each lead through a digital fourth-order Butterworth high-pass filter with cut-off frequency of 0.5 Hz and high-frequency noise components using a fourth-order Butterworth low-pass filter with cut-off frequency of 70 Hz. The filters were applied both in forward and backward direction to get zero-phase distortion. Then, the signal was normalized by calculating the z-score of each lead in order to remove the amplitude differences between different recordings.

B. VMD-based QRS segmentation

1) Variational Mode Decomposition: In this study, variational mode decomposition was applied to segment the QRS complexes. VMD adaptively decomposes a real-valued multi-component signal $x(\mathrm{t})$ into K discrete number of modes or components $u_k(t)$ with $k = 1, ..., K$. All components have certain sparsity properties, and the bandwidth in spectral domain is considered as the sparsity prior of each mode. All modes are mostly compact around their center frequencies ω_k.

The bandwidth of a mode is evaluated in three steps: (i) the analytic signal of the real-valued signal is computed using the Hilbert transform such that the frequency spectrum becomes unilateral, (ii) the frequency spectrum of the analytic signal corresponding to each mode is shifted to baseband regions by multiplying it with the factor $e^{-i\omega_k t}$, (iii) then, the bandwidth is estimated through the squared L^2-norm of the gradient of the demodulated signal. Mathematically, this constrained variational problem can be expressed as follows

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad s.t. \sum_k u_k(t) = x(t) \tag{1}$$

where δ is the Dirac-Delta function, $*$ is the convolution operator, and $\{u_k\}$ and $\{\omega_k\}$ represent the set of modes and center frequencies, respectively.

The constrained problem in Eq. (1) can be solved by converting it into an unconstrained problem using the balancing parameter α and the Lagrangian multiplication parameter $\lambda(t)$. The modified equation with the augmented Lagrangian multiplier is expressed as

$$\mathcal{L}(\{u_k\},\{\omega_k\},\lambda) := \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| x(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), x(t) - \sum_k u_k(t) \right\rangle. \tag{2}$$

The above Eq. (2) is solved using the alternate direction method of multipliers [21]. We have skipped the intermediate steps to solve this equation and some other related algorithms as these are beyond the scope of this paper. The readers are encouraged to refer to the original article for the full implementation of the VMD method [16]. The solutions to Eq. (2) in the Fourier domain represent all modes and their central frequencies where each mode and its center frequency are updated iteratively as

$$\hat{U}_k^{n+1}(\omega) = \frac{\hat{X}(\omega) - \sum_{i \neq k} \hat{U}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \tag{3}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left| \hat{U}_k(\omega) \right|^2 d\omega}{\int_0^\infty \left| \hat{U}_k(\omega) \right|^2 d\omega} \tag{4}$$

where $\hat{X}(\omega)$, $\hat{U}_i(\omega)$, $\hat{\lambda}(\omega)$, and $\hat{U}_k^{n+1}(\omega)$ are the Fourier transforms of the respective time domain representations. The Wiener filter structure in Eq. (3) makes the VMD method more robust to noise and sampling [22]. Finally, the real part of the inverse Fourier transform of $\hat{U}_k^{n+1}(\omega)$ gives the time domain representation of the modes. The center frequency in each iteration in Eq. (4) is the center of

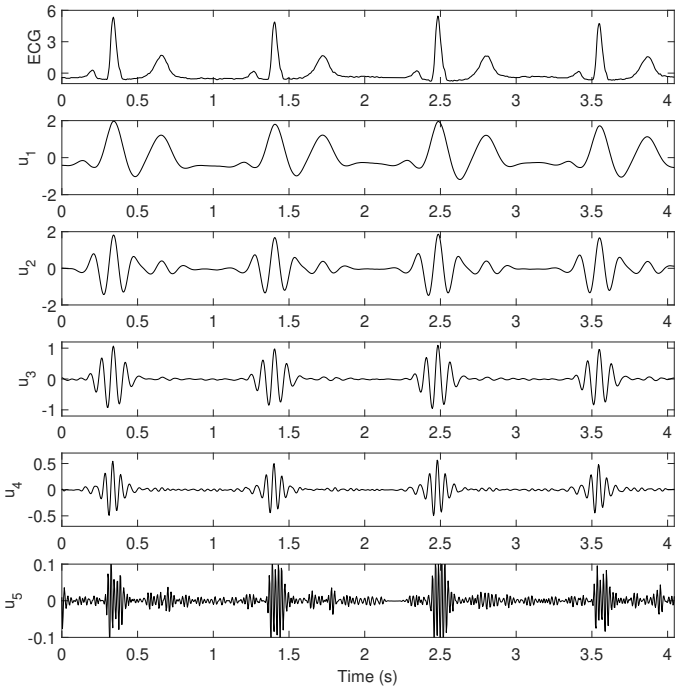

Fig. 2: Example of ECG signal with the corresponding modes of the output of Variational Mode Decomposition with $k = 5$ and $\alpha = 100$.

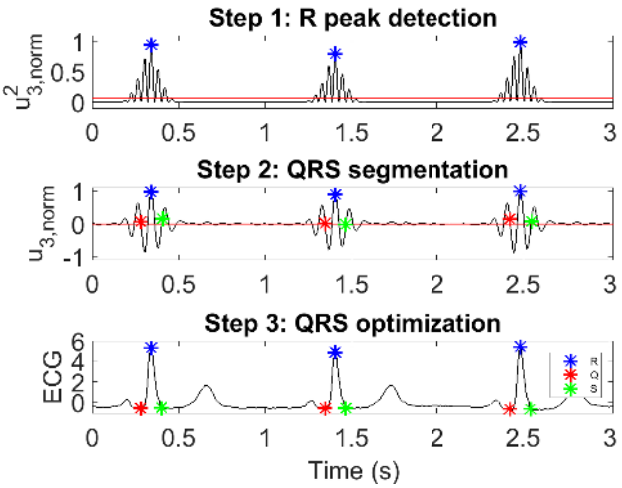

Fig. 3: Illustration of the three main steps for QRS segmentation, applied to the signal shown in Figure 2. First the R peaks are detected in $u^2_{3,norm}$. Then the QRS complex is segmented in $u_{3,norm}$. Finally the QRS locations are optimized in the original ECG signal. The R peaks are depicted in blue, the beginning and end of the QRS complex in respectively red and green.

gravity of the corresponding positive part of mode's power spectrum.

2) QRS segmentation: From the VMD algorithm discussed above, it is clear that the VMD is a parameterized signal decomposition method. VMD requires two parameters to be fixed: the number of modes K and the bandwidth control parameter α. Initialization of these parameters is a non-trivial problem, since wrong parameter choices may create problems like mode splitting (where one component is shared by several modes) or mode mixing (where multiple components are decomposed into one mode) [22]. To avoid this problem, we adopted the optimization technique developed by Guo et al. [22] to fix the number of decomposition modes K and balancing parameter α prior to applying VMD. After optimization, these values were fixed to $K = 5$ and $\alpha = 100$. The other input VMD parameters, namely the time step of the dual ascent and the tolerance of the convergence criterion were set to standard values 0 and $1e^{-6}$ [16]. All central frequencies ω_k were uniformly distributed at initialization.

VMD was applied to each ECG lead $x(t)$ resulting in a decomposition into 5 modes:

$$x(t) = \sum_{k=1}^{5} u_k(t) \tag{5}$$

where $u_k(t)$ is the kth mode. Figure 2 shows an example of an ECG segment after preprocessing with the corresponding modes u_1–u_5 in different rows. The modes are sorted by their central frequency in ascending order. It can be observed that different components (characteristic waves) of the ECG signal are decomposed into different modes. The low-frequency components appear in lower modes and vice versa. The QRS complex is a high frequency wave with sharp amplitude and thus appears in higher modes (u_2-u_4). Based on the fact that the modes are mostly compact around their center frequencies, the QRS complex and hence the Q and S points can best be detected using u_3, which is amplitude-normalized to obtain $u_{3,norm}$. The complete QRS segmentation algorithm consists of three main steps, which are illustrated in the different rows of Figure 3:

1) *R peak detection*
   Square $u_{3,norm}$ in order to obtain $u^2_{3,norm}$ to enhance the main peaks in $u_{3,norm}$ (corresponding to QRS complexes) and minimize noise peaks. Detect the R peaks in $u^2_{3,norm}$ by finding local maxima with an amplitude higher than 0.075. The minimal distance between two consecutive peaks must be larger than 0.5 seconds. This corresponds to a maximal heart rate of 120 beats per minute which is realistic for ECG signals measured in rest. For signals with potentially increased heart rates, this parameter can easily be changed to accomodate them.
2) *QRS segmentation*
   Transfer the R peaks to $u_{3,norm}$. The beginning of the QRS complex corresponds to the first zero-crossing *before* the first local minimum *preceding* the R-peak in $u_{3,norm}$. Similarly, the end of the QRS complex can be found by detecting the first zero-crossing *after* the first local minimum *after* the R peak.
3) *QRS optimization*
   Optimize the locations of the Q and S points by finding a local minimum in the ECG signal in the neighbourhood ($\pm$ 5 samples) of the points detected

in $u_{3,norm}$.

QRS segmentation was done in each lead separately. In a final stage, segmentations in all leads were combined in an automated way to remove false detections. False positive detections can occur when the ECG signal contains noise with characteristics similar to QRS complexes. They were removed by first calculating in how many leads a QRS complex was detected at a certain time instance and subsequently removing QRS complexes that were detected in less than half of the ECG leads. False negative detections were solved similarly by automatically adding QRS complexes in leads where a complex was missed by the algorithm. The start- and end points for these additional complexes were selected as the mean start- and end points of the complexes detected by the algorithm in the other leads of the signal.

Finally all points between the end of the QRS complex and the beginning of the next complex are set to zero since they will not be analyzed further.

C. Feature extraction

In this subsection, feature extraction using PRSA and VMD is discussed. PRSA feature extraction was used in our earlier study [20], and the same approach has been adopted in this work. All features were extracted from the QRS complexes segmented using the method described in the previous subsection.

1) PRSA-based features: Phase-rectified signal averaging aims to detect and quantify quasi-periodic oscillations masked by multi-component non-stationary signals. The method consists of three steps:

(i) *Selection of anchor points*
First, two sets of anchor points q_i were selected according to certain properties in the QRS complex. The anchor points can be selected in different ways [23]; here, a first set of anchor points consists of all points located on the increasing part of the QRS complex (q_i^+), a second set contains all points on decreasing parts of the QRS complex(q_i^-).

(ii) *Window selection*
Then, temporal windows of length $2L$ were selected around each anchor point. The anchor points which were too close ($< L$ ms) to the beginning or end of the complex were discarded since for these points no such window can be selected. The choice of L is dependent on the characteristics of the signal: it must be larger than the period of the slowest oscillation that should be detected. Here L is fixed to 20 ms, similar to [20].

(iii) *Averaging*
In the final step, the PRSA curve $\tilde{q}$ was obtained by aligning all windows and averaging them. The windows corresponding to decreasing anchor points q_i^- were inverted before calculating the average, e.g. $window(q_i^-) = -window(q_i^-)$. Inverting the windows is necessary to ensure that all windows have similar slopes.

The different steps to calculate the PRSA curve for a normal and a fragmented beat are shown in Figure 4. Two positive and two negative anchor points are highlighted in respectively red, cyan and green, blue to highlight the changes in the signals. For a normal beat (depicted in the top row), all increasing anchor points are located in the first part of the QRS complex and all decreasing points in the second part due to the very simple morphology of the wave. Most windows are thus similar in shape and the PRSA curve can be approximated by a straight line. When a signal shows fragmentation (shown in the bottom row), the anchor points are not contained in one part of the QRS complex. The surrounding windows thus exhibit more differences which result in a PRSA curve with a smaller inclination, which can be seen by comparing the rightmost panels of Figure 4.

To extract features from the PRSA curve, it was approximated with a linear fit. Three parameters were extracted to quantify the curve:

(i) Mean derivative of the PRSA curve (sl): $\frac{1}{2L}\sum_{t=0}^{2L}\frac{d\tilde{q}(t)}{dt}$
(ii) Slope of the linear fit (m)
(iii) Intercept of the linear fit with the y-axis (c).

2) VMD-based features: Fragmentation introduces additional high-frequency components in the ECG signal, which in turn causes extra oscillations in the VMD modes that contain QRS information. Therefore the central frequencies of these modes also increase. As discussed in Subsection III-B.1, modes u_1 and u_2 contain the low-frequency ECG components such as P- and T-waves, and other modes mostly carry the high-frequency components such as the QRS complex. Due to the presence of notches or additional waves, it is expected that the number of local optima per QRS complex will increase with respect to the certainty of fQRS. Hence, two different VMD-based features were extracted from modes u_3, u_4 and u_5: the average number of local optima per QRS complex (pks_3, pks_4, pks_5) and the central frequency of the QRS complexes in the considered modes ($\omega_3^{QRS}, \omega_4^{QRS}, \omega_5^{QRS}$), leading to a total of 6 VMD-based features.

One additional feature was extracted directly from the QRS complex: the number of local peaks in the ECG signal per QRS complex, pks_{ECG}. This feature was combined with the three PRSA features and six VMD features to obtain a total of 10 features for each lead.

D. Classification

The extracted features were used as input to a classifier in order to detect and quantify QRS fragmentation. It has been shown that support vector machines (SVMs) are in many types of problems superior to 16 other popular classifiers [24]. They are also known to have good generalization properties in supervised classification problems. The basic idea of an SVM is to find a hyperplane which separates two classes by maximizing the margin between the classes. This can be done in a linear or non-linear way,

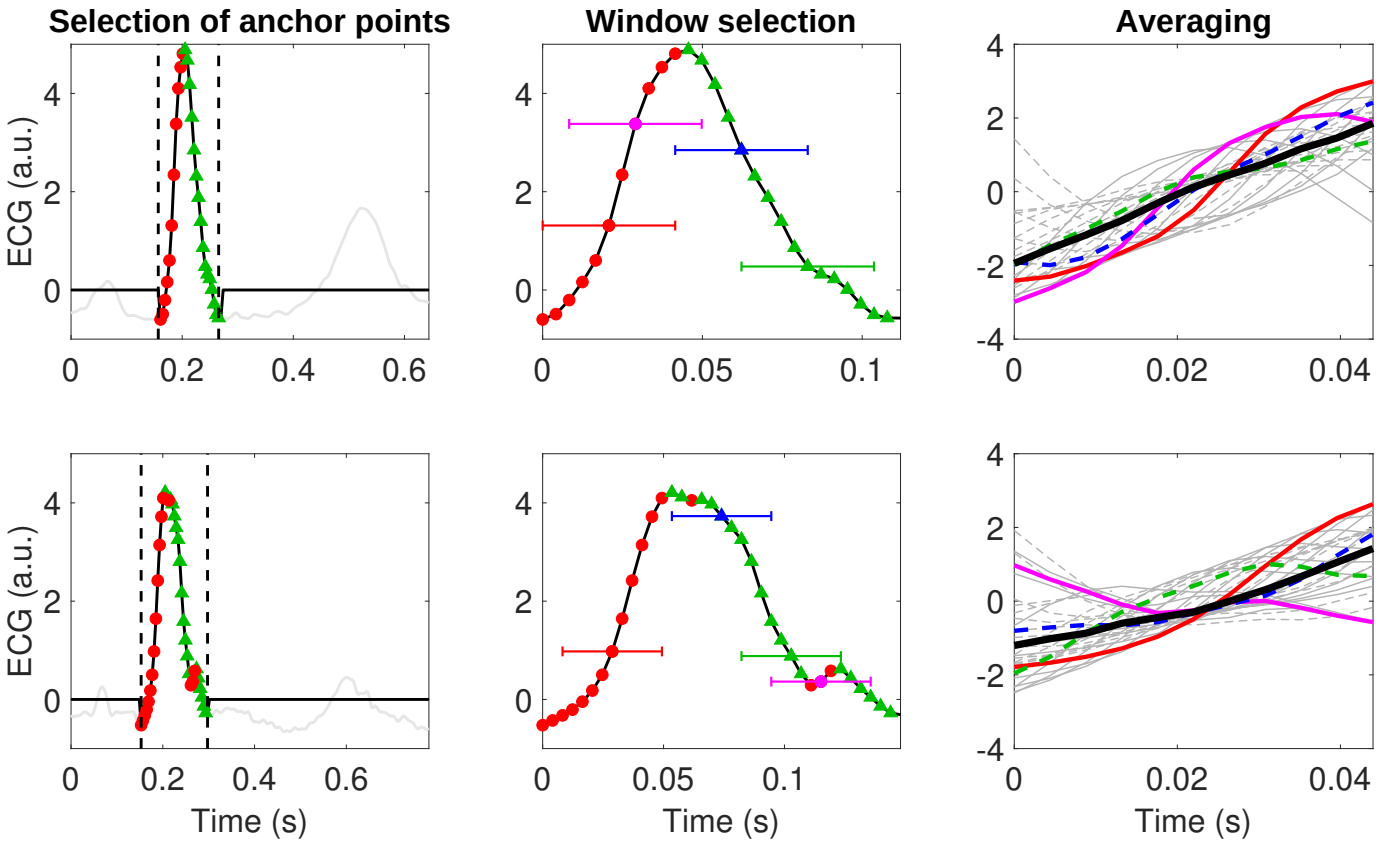

Fig. 4: Illustration of the three steps to create the PRSA curve for a normal heartbeat (top) and a heartbeat with fragmentation (bottom). In the middle panel, two increasing and two decreasing anchor points are highlighted in respectively red, cyan and green, blue. Their corresponding windows are also indicated in the right panel.

by introducing kernel functions that transform the feature space. When perfect separation cannot be achieved, misclassifications are allowed by introducing a soft margin constant that penalizes points that lie on the wrong side of the margin. The performance of SVMs with a linear, polynomial and radial basis function kernel was compared to select the kernel which obtains the best performance. In order to define the optimal SVM parameters, automatic Bayesian optimization of hyperparameters (both the soft margin constant and kernel parameters) using 10-fold cross-validation was performed. Additionally, to illustrate the effectiveness of the proposed method, the performance of three different classifiers (K-nearest neighbors (KNN), Naive Bayesian (NB), and TreeBagger (TB)) was also evaluated and compared with the results obtained by SVMs. The output of the classifiers (e.g. the score belonging to the positive class) was transformed to a score between 0 and 1 through the use of Platt scaling, which fits a logistic regression model to the output [25]. The obtained score is taken as fQRS score and is expected to reflect the certainty of fQRS.

The dataset used in this study contained 6776 signals in total (616 records with 11 leads per record). These signals were divided into one training and two test sets. To train the classifier, only signals where all experts agreed on the presence of QRS fragmentation were used. There were 2775 normal signals and 1674 fragmented signals without disagreement (see Table I). These were randomly split in 80% training set and 20% test set. Both sets had equal ratios of normal and fragmented signals. A second test set contained all signals that were not used in the training stage. The first test set was used to evaluate the performance of the method for fQRS detection: since all raters agreed on the label for all signals in this set, this label was considered to be correct and binary classification could be done. The second test set on the other hand was used to assess whether the developed score reflected the certainty of QRS fragmentation in an ECG signal.

The training set and the first test set are fully independent: signals from the same patient are only present in one of both sets. Since the second test set contains all ECG channels where no perfect agreement was reached, it also contains channels from patients in the training set, and both sets are not fully independent.

E. Performance evaluation metrics

Different measures were used to evaluate the performance of the developed method. Statistical analyses of parameter values were done with one-way Analysis of Variance (ANOVA) with the F-test [26] followed by a post-hoc test analysis based on Tukey's honest significant difference (hsd) test. These tests were carried out to verify whether means of different categories are significantly different or not. ANOVA assesses the relative size of variance among

category means compared to the average variance within categories and is an appropriate test for evaluating the effect of categorical independent variables on a continuous response variable. The F-ratio, the value obtained from the ratio of the variance between categories and the variance within categories, was used as the parameter to decide the statistical significant performance. The F-test represents the determination of the significance of the F-ratio by comparing it to a critical value derived from the probability distribution. The null hypothesis that the means of all categories are equivalent is rejected if the F-ratio is greater than the critical value. If the F-test is statistically significant, then there is, in principle, at least one significant difference in means. Then, post-hoc test using Tukey's hsd test was executed to perform specific comparisons for the purpose of discovering the origin(s) of the difference. In all tests, a p-value < 0.01 was considered statistically significant.

Performance of the different classifiers was evaluated by constructing Receiver Operator Curves (ROC) and calculating the corresponding Area Under The Curve (AUC). The ROC is generated by varying the decision threshold and calculating the senstivity and specificity results for each threshold. The results of fQRS detection were further quantified by calculating the number of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) detections and computing the sensitivity (sens), specificity (spec) and accuracy (acc) as follows:

$$\text{sens} = \frac{TP}{TP+FN}$$
$$\text{spec} = \frac{TN}{TN+FP}$$
$$\text{acc} = \frac{TP+TN}{TP+FN+TN+FP}$$

IV. Results

The full dataset was analyzed using the proposed method depicted in Figure 1. In this section, we present the results of the different analyses. All experiments in this work were performed using MATLAB R2017b.

A. QRS segmentation

To evaluate the results of the proposed approach to QRS segmentation, the publicly available QT database was used, which contains 105 ECG signals of 15 minutes with manually determined wave boundaries [27]. The database has been used extensively as benchmark tool for ECG delineation algorithms. The performance is quantified by calculating the mean deviation and standard deviation between the VMD-based segmentation and the manual annotations.

The proposed R peak detection algorithm detected 98.11% of all R peaks correctly. This is equivalent to the results found in literature [28]–[30]. The accuracies of the delineation of the onset and offset of the QRS complex are summarized in Table II. The results for segmentation of the QRS onset are comparable to the state-of-the art algorithms. The standard deviation of the QRS offset detection is slightly worse than the results reported by Madeiro et al. [29] and Akhbari et al. [30], but the difference is limited to five ms, which is equal to one sample for signals sampled at 250 Hz.

TABLE II: Accuracy results of QRS segmentation of the proposed method and three state-of-the art alternative approaches on the QT database. Mean and standard deviation of the difference between the provided annotations and the segmentation obtained by the algorithm are given in ms.

	QRS_{on}		QRS_{off}	
Method	mean	std	mean	std
Proposed method	-3.6	11.16	6.67	17.28
Martínez et al. [28]	4.6	7.7	0.8	8.7
Madeiro et al. [29]	-3.4	11.6	-6.5	12.3
Akhbari et al. [30]	-5	10	1.5	11.5

B. Analysis of feature values

Figure 5 shows box plots for all ten features. Each set of box plots is calculated by computing the particular feature value for all leads of all signals and grouping them per total score (e.g. the sum of the scores by all raters). As discussed in Section II, the total score is a substitute for the certainty of fQRS in an ECG lead. We observe that for all features the feature values gradually change with changing total score. The average slope of the PRSA curve and the slope of the linear fit are both inversely related to fQRS certainty, which is expected based on the examples in Figure 4. Similarly, the intercept of the linear fit increases with increasing certainty. The number of peaks per QRS complex in the VMD modes regularly increases with the level of fragmentation. The increase in the median values in modes u_4 and u_5 is more prominent compared to the medians in u_3. It is interesting to note that the central frequency of the QRS complexes in these modes also follows the same trend.

ANOVA is used to statistically compare the means of the values in different score groups. Here, the null hypothesis is that the mean of a feature is identical for all six categories. The F-ratio results reveal that for **all** features, there is at least one mean that is significantly different from one or more categories. As this happens for all features and due to page limits, we only present the post-hoc analysis.

Table III shows the post-hoc analysis based on the Tukey's hsd test that shows which categories differ from each other by comparing the means between consecutive categories. Significant p-values ($p < 0.01$) are indicated with 'Y'.

C. Classifier performance

In this subsection, we present the performance of different classifiers for fQRS detection and quantification. As mentioned in Section III-D, the classifiers were trained

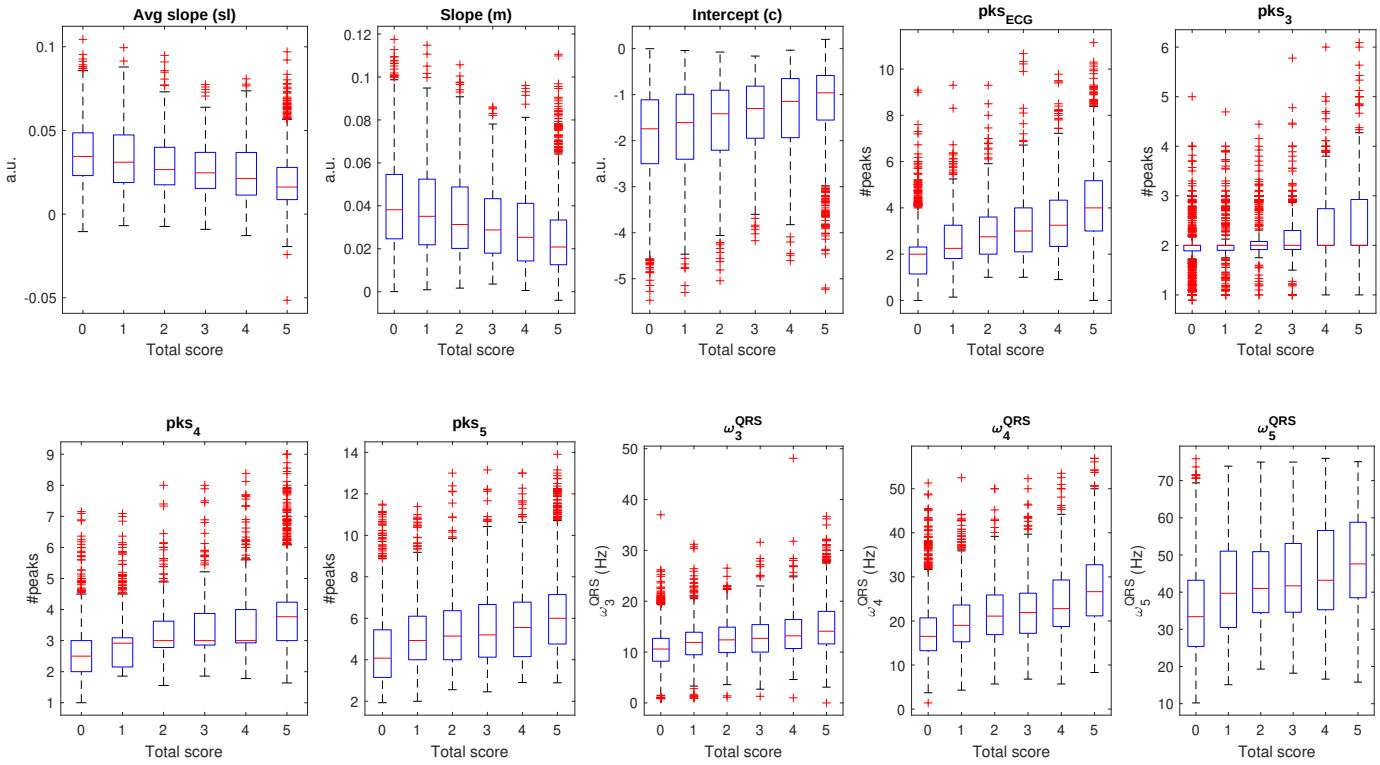

Fig. 5: Box plots showing all feature values grouped by the total score given by 5 experts. The box plots represent the median value in a group together with the interquartile range. Outliers are shown in red.

TABLE III: Significance results of the post-hoc analysis for the comparison of the VMD- and PRSA-based features using Tukey's hsd test. Results are shown only between the consecutive categories. 'Y' stands for statistically significant with p <0.01. p-values are mentioned for all non-significant results.

Compared categories:	0-1	1-2	2-3	3-4	4-5
PRSA-based features					
Avg. slope (sl)	Y	Y	0.140	0.070	Y
Slope (m)	Y	0.108	0.075	0.236	Y
Intercept (c)	Y	0.102	0.069	0.289	Y
VMD-based features					
ω_3^{QRS}	Y	Y	0.018	Y	Y
ω_4^{QRS}	Y	Y	0.699	0.038	Y
ω_5^{QRS}	Y	Y	0.466	0.062	Y
pks_3	Y	0.019	0.639	0.828	Y
pks_4	Y	0.050	0.410	Y	Y
pks_5	Y	Y	0.459	Y	Y
pks_{ECG}	Y	0.077	0.865	0.300	Y

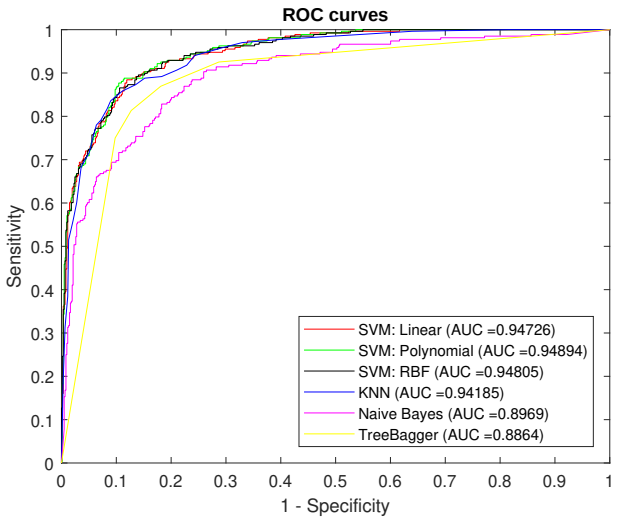

Fig. 6: ROC curves for fQRS detection with all classifiers together with the corresponding AUC values.

on 80% of the signals without disagreement between the raters. The other signals were used for evaluation during the testing stage.

Figure 6 shows the ROC curves with corresponding AUC values for all classifiers. Here, the test set consists of the 20% remaining signals with perfect agreement. The true class of these signals is considered to be certain so classification can be performed. The differences between the SVMs with different kernels are neglegible (AUC = 0.947 (lin), 0.948 (pol) and 0.948 (RBF)). The KNN classifier (AUC = 0.941) performs slightly worse than the SVMs while the performance of the NB (AUC = 0.896) and TB (AUC = 0.886) classifiers are comparatively worse.

Table IV finally summarizes the sensitivity, specificity

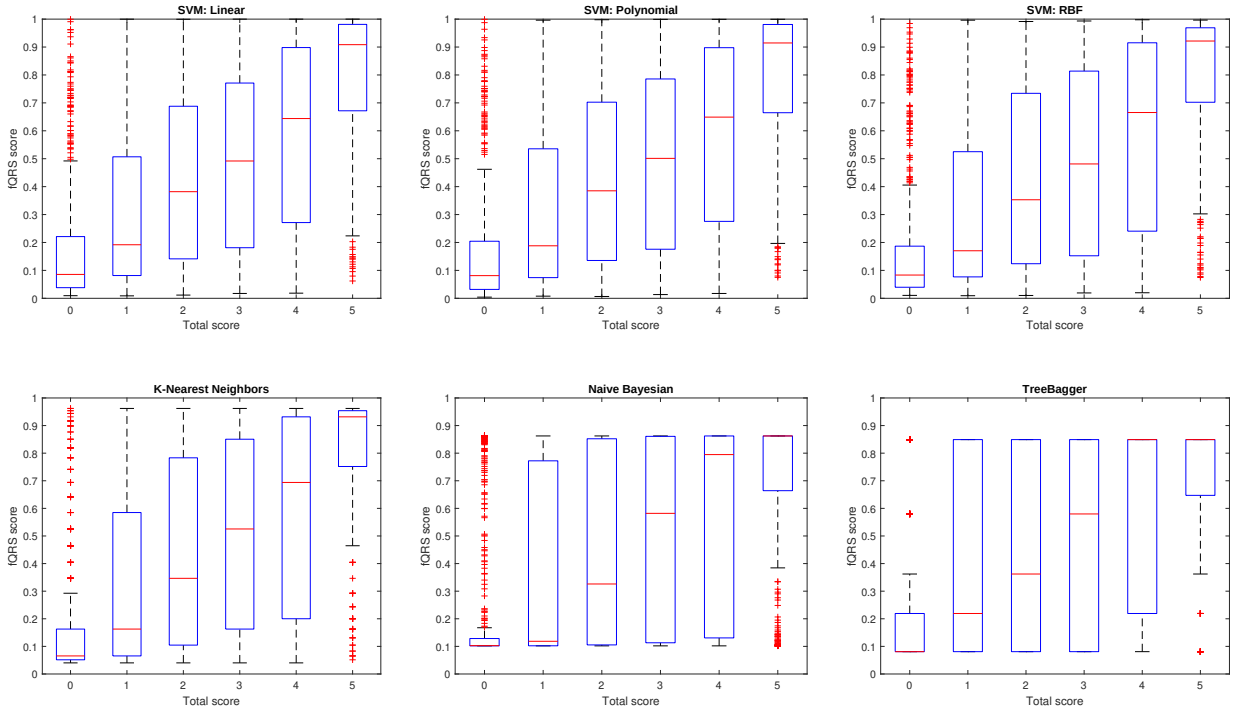

Fig. 7: Illustrating the fQRS quantification score for the second group (Section III-D) using SVM (Linear, Polynomial and RBF kernels), KNN, NB and TB classifiers.

TABLE IV: Comparison of the fQRS classification results from the proposed method with methods from literature. The obtained sensitivity, specificity and accuracy on the first test set are mentioned.

Method	Sens (%)	Spec (%)	Acc (%)
Proposed method	86%	89%	88%
Jin et al. [7]	72%	78%	75%
Maheshwari et al. [6]	84%	87%	85%
Bono et al. [8]	75%	78%	77%

and accuracy results obtained by applying the proposed method (with RBF kernel) and the three alternative algorithms [6]–[8] on the first test set. The other approaches use the wavelet transform [6], [8] or intrinsic time-scale decomposition [7] to decompose the ECG signal and detect fQRS from features obtained from the decomposed signal.

The box plots in Figure 7 show the fQRS quantification results. Here, the test dataset contains all instances that are not considered in the training stage. The box plot for each category is computed by combining the fQRS scores for all signals with the same total score. The output scores from most classifiers follow an increasing trend with respect to the different categories. The results for the NB and TB classifiers are considerably worse than the other classifiers. Results for SVM and KNN classifiers are again comparable, but the interquartile range for the different SVM box plots is smaller than for KNN. Furthermore the range of SVM scores spans the full range (0–1) while the range for the other classifiers is limited to a smaller interval.

V. Discussion

This study proposes an innovative method to detect and quantify fQRS. In the first step, the QRS complexes are segmented using a novel VMD-based algorithm. According to Table II, the approach gives results that are comparable to many state-of-the-art algorithms. While the widely-used wavelet-based method from Martínez et al. [28] obtains superior results, the differences are limited to a few samples.

Figure 5 and Table III summarize results for the feature values in different groups. All values change gradually with increasing raters' scores, indicating that the reasoning in Section III-C is correct: fragmentation has a significant influence on the PRSA curve and introduces extra high frequency components in the results of VMD, which in turn give rise to extra peaks. Table III shows that all features are significantly different between the most extreme groups (0-1 and 4-5). The median differences between other groups are not always statistically significant, but there is a clear trend which can also be noticed in Figure 5. The feature values for signals with a total score of 2 and 3 are similar, with p-values larger than 0.35. This is not unexpected since the signals in these groups are comparable: they represent the signals which are scored as fragmented by approximately half of the experts and normal by the other half.

Once the PRSA- and the VMD-based features have been extracted from the QRS complexes, they were fed into

different classifiers in order to evaluate and compare their performance. Experimental results presented in Figure 6 show that all classifiers give good results for detecting signals with clear QRS fragmentation. SVMs are slightly superior to other classifier types. The results in Table IV show that our approach outperforms the other methods, although the results obtained by [6] are in the same range. A second test set was used to evaluate the performance for fQRS quantification. Here the goal was to verify whether the developed score is representative for the certainty of the presence of fQRS in an ECG lead. Platt scaling transforms the classifier output to a score between 0 and 1, which corresponds to the posterior probability. This results in the box plots shown in Figure 7. For the different types of SVMs, the fQRS score varies between 0 and 1 and linearly changes with the total score from all experts. As explained in Section II, the total score is related to the certainty of fQRS in a lead. It is therefore expected that leads with a questionable fQRS presence (represented by a total score of 1–2, where only one or two raters observe fQRS in that lead) have lower fQRS scores than leads with a probable (total score of 3–4) or certain presence (total score = 5). Figure 7 confirms this hypothesis. If the boxplots of the different scores are compared, the boxplot heights of values 0 and 5 are significantly smaller than the heights of the intermediate values. This confirms that the classifier output for values 1–4 is indeed more uncertain than for the extreme values. This is also expected, since the true label contains more uncertainty (for the intermediate values, the presence or absence of QRS fragmentation is after all not certainly known).

While the number of outliers in the boxplot corresponding to a total score of 0 seems large at first sight, they represent less than 2% of all signals in the second test set. These outliers can have different explanations: For some signals, the QRS segmentation is not 100% accurate, which can have an influence on the feature values. For other signals, an increased noise level can change the ECG characteristics and have an influence on the final classification. The fact that the number of outliers is small and the good AUC scores obtained on the first test set however confirms that this is only the case in a small minority of the data.

KNN also results in a stepwise increasing score, while the results of NB and TB are considerably worse. The scores obtained by KNN, however, do not span the full range between 0 and 1 but are limited from 0.05 to 0.95. It has been shown before that Platt calibration of KNN classifiers can lead to diminished results while the results of SVM classifiers are significantly boosted [31], [32]. Since SVMs also result in superior detection results (shown on Figure 6), SVM classifiers are preferred over other classification methods.

The performances of three different SVM kernels (linear, polynomial and RBF kernels) were compared. They all have similar results both in fQRS detection and quantification. Since the numerical results are almost identical, the choice of kernel does not appear to have a significant effect in this application. In general, however, the RBF kernel is preferred in such cases since it is known to be a universal approximator.

VI. Conclusion

This paper presents a novel and precise way to detect and quantify QRS fragmentation in ECG signals using machine learning techniques. The fQRS score represents the certainty of QRS fragmentation in a continuous way based on fQRS annotations from 5 experts. The features extracted using VMD and PRSA show values that change gradually with increasing fQRS certainty. The fQRS score obtained using SVM classifiers is closely related to the total score given by all raters which is representative for the level of fragmentation in a lead.

The results demonstrate that the proposed fQRS score is an effective way of detecting and quantifying QRS fragmentation. Comparison with existing techniques shows that the proposed method outperforms other methods found in literature. The extensive scoring done by five independent raters is a strong indicator that the results are robust to inter-rater variability and can be generalized to other datasets. This should, however, be validated with other datasets scored in a similar way. The availability of scores by different raters also allows us to evaluate fQRS quantification, which is a novel approach to examine the biomarker.

Clinically, QRS fragmentation is scored on a per-lead basis, after which the detections in separate leads are combined per cardiac region. We have shown in previous work that assessing fQRS per region might be of more clinical value than using an overall fQRS interpretation per ECG lead [4]. Future work therefore includes combining the fQRS scores in a similar way to achieve a score per cardiac region which can be used for clinical validation. Additionally, the scores from all leads could also be combined in order to obtain a per-patient decision on the presence of fQRS. The final goal is to verify whether a continuous way of scoring QRS fragmentation also leads to superior results in clinical studies that focus on analyzing patient outcomes.

References

[1] M. K. Das, C. Saha, H. E. Masry, J. Peng, G. Dandamudi, J. Mahenthiran, P. McHenry, and D. P. Zipes, "Fragmented qrs on a 12-lead ecg: A predictor of mortality and cardiac events in patients with coronary artery disease," *Heart Rhythm*, vol. 4, no. 11, pp. 1385 – 1392, 2007.

[2] G. Pietrasik and W. Zaręba, "Qrs fragmentation: diagnostic and prognostic significance." *Cardiology Journal*, vol. 19, no. 2, pp. 114–121, 2012.

[3] H. Morita, K. F. Kusano, D. Miura, S. Nagase, K. Nakamura, S. T. Morita, T. Ohe, D. P. Zipes, and J. Wu, "Fragmented qrs as a marker of conduction abnormality and a predictor of prognosis of brugada syndrome," *Circulation*, vol. 118, no. 17, pp. 1697–1704, 2008.

[4] B. Vandenberk, T. Robyns, G. Goovaerts, S. Van Soest, V. Floré, C. Garweg, S. Van Huffel, J. Ector, and R. Willems, "Inferior and anterior qrs fragmentation have different prognostic value in patients who received an implantable defibrillator in primary prevention of sudden cardiac death," *International journal of cardiology*, vol. 243, pp. 223–228, 2017.

[5] N. C. Flowers, L. G. Horan, J. R. Thomas, and W. J. Tolleson, "The anatomic basis for high-frequency components in the electrocardiogram," *Circulation*, vol. 39, no. 4, pp. 531–539, 1969.

[6] S. Maheshwari, A. Acharyya, P. E. Puddu, E. B. Mazomenos, G. Leekha, K. Maharatna, and M. Schiariti, "An automated algorithm for online detection of fragmented qrs and identification of its various morphologies," *Journal of The Royal Society Interface*, vol. 10, no. 89, p. 20130761, 2013.

[7] F. Jin, L. Sugavaneswaran, S. Krishnan, and V. S. Chauhan, "Quantification of fragmented qrs complex using intrinsic time-scale decomposition," *Biomedical Signal Processing and Control*, vol. 31, pp. 513–523, 2017.

[8] V. Bono, E. B. Mazomenos, T. Chen, J. A. Rosengarten, A. Acharyya, K. Maharatna, J. M. Morgan, and N. Curzen, "Development of an automated updated selvester qrs scoring system using swt-based qrs fractionation detection and classification," *IEEE journal of biomedical and health informatics*, vol. 18, no. 1, pp. 193–204, 2014.

[9] B. Vandenberk, T. Robyns, G. Goovaerts, M. Claeys, F. Helsen, S. V. Soest, C. Garweg, J. Ector, S. V. Huffel, and R. Willems, "Inter- and intra-observer variability of visual fragmented qrs scoring in ischemic and non-ischemic cardiomyopathy," *Journal of Electrocardiology*, 2017.

[10] M. A. E. Haukilahti, A. Eranti, T. Kenttä, and H. V. Huikuri, "Qrs fragmentation patterns representing myocardial scar need to be separated from benign normal variants: hypotheses and proposal for morphology based classification," *Frontiers in physiology*, vol. 7, p. 653, 2016.

[11] R. J. Martis, U. R. Acharya, and H. Adeli, "Current methods in electrocardiogram characterization," *Computers in Biology and Medicine*, vol. 48, pp. 133 – 149, 2014.

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

[13] M. A. Arafat and M. K. Hasan, "Automatic detection of ecg wave boundaries using empirical mode decomposition," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, April 2009, pp. 461–464.

[14] S. Pal and M. Mitra, "Empirical mode decomposition based ecg enhancement and qrs detection," *Computers in Biology and Medicine*, vol. 42, no. 1, pp. 83 – 92, 2012.

[15] U. Satija, B. Ramkumar, and M. S. Manikandan, "A unified sparse signal decomposition and reconstruction framework for elimination of muscle artifacts from ecg signal," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 779–783.

[16] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, Feb 2014.

[17] A. Mert, "Ecg feature extraction based on the bandwidth properties of variational mode decomposition," *Physiological Measurement*, vol. 37, no. 4, p. 530, 2016.

[23] A. Bauer, J. W. Kantelhardt, A. Bunde, P. Barthel, R. Schneider, M. Malik, and G. Schmidt, "Phase-rectified signal averaging detects quasi-periodicities in non-stationary data," *Physica A: Statistical Mechanics and its Applications*, vol. 364, pp. 423 – 434, 2006.

[18] U. Maji, M. Mitra, and S. Pal, "Characterization of cardiac arrhythmias by variational mode decomposition technique," *Biocybernetics and Biomedical Engineering*, vol. 37, no. 3, pp. 578 – 589, 2017.

[19] R. K. Tripathy, L. N. Sharma, and S. Dandapat, "Detection of shockable ventricular arrhythmia using variational mode decomposition," *Journal of Medical Systems*, vol. 40, no. 4, p. 79, Jan 2016.

[20] G. Goovaerts, B. Vandenberk, C. Varon, R. Willems, and S. Van Huffel, "Phase-rectified signal averaging for automatic detection of qrs fragmentation," in *2016 Computing in Cardiology Conference (CinC)*, Sept 2016, pp. 637–640.

[21] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[22] Y. Guo, Z. Zhang, J. Cao, T. Gong, and W. Yang, "An optimized variational mode decomposition for extracting weak feature of viscoelastic sandwich cylindrical structures," *Measurement Science and Technology*, vol. 29, no. 3, p. 035006, 2018.

[24] D. Meyer, F. Leisch, and K. Hornik, "The support vector machine under test," *Neurocomputing*, vol. 55, no. 1, pp. 169 – 186, 2003, support Vector Machines.

[25] J. Platt *et al.*, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Advances in large margin classifiers*, vol. 10, no. 3, pp. 61–74, 1999.

[26] B. Rosner, *Fundamentals of biostatistics*. Belmont: Duxbury Press, 2006.

[27] P. Laguna, R. G. Mark, A. Goldberg, and G. B. Moody, "A database for evaluation of algorithms for measurement of qt and other waveform intervals in the ecg," in *Computers in cardiology 1997*. IEEE, 1997, pp. 673–676.

[28] J. P. Martínez, R. Almeida, S. Olmos, A. P. Rocha, and P. Laguna, "A wavelet-based ecg delineator: evaluation on standard databases," *IEEE Transactions on biomedical engineering*, vol. 51, no. 4, pp. 570–581, 2004.

[29] J. P. Madeiro, P. C. Cortez, F. I. Oliveira, and R. S. Siqueira, "A new approach to qrs segmentation based on wavelet bases and adaptive threshold technique," *Medical engineering & physics*, vol. 29, no. 1, pp. 26–37, 2007.

[30] M. Akhbari, M. B. Shamsollahi, O. Sayadi, A. A. Armoundas, and C. Jutten, "Ecg segmentation and fiducial point extraction using multi hidden markov model," *Computers in biology and medicine*, vol. 79, pp. 21–29, 2016.

[31] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006, pp. 161–168.

[32] A. Niculescu-Mizil and R. Caruana, "Predicting good probabilities with supervised learning," in *Proceedings of the 22nd international conference on Machine learning*. ACM, 2005, pp. 625–632.