A machine learning algorithm for electrocardiographic fQRS quantification validated on multi-center data

Amalia Villa[1,*], Bert Vandenberk[2,3], Tuomas Kenttä[4], Sebastian Ingelaere[2], Heikki V Huikuri[4], Markus Zabel[5], Tim Friede[6,7], Christian Sticherling[8], Anton Tuinenburg[9], Marek Malik[10,11], Sabine Van Huffel[1], Rik Willems[2], and Carolina Varon[1,12]

[1]Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Leuven, Belgium
[2]Department of Cardiovascular Diseases, Experimental Cardiology, KU Leuven, Leuven, Belgium
[3]Department of Cardiac Sciences, Libin Cardiovascular Institute, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada
[4]Research Unit of Internal Medicine, Medical Research Center Oulu, University of Oulu and Oulu University Hospital, Oulu, Finland
[5]Department of Cardiology and Pneumology, Heart Center, University of Göttingen Medical Center, Göttingen, Germany
[6]Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany
[7]DZHK (German Center of Cardiovascular Research), partner site Göttingen, Göttingen, Germany
[8]Division of Cardiology, University of Basel Hospital, Basel, Switzerland
[9]Department of Cardiology, University Medical Center Utrecht, Utrecht, Netherlands
[10]National Heart and Lung Institute, Imperial College, London, UK
[11]Department of Internal Medicine and Cardiology, Masaryk University, Brno, Czech Republic
[12]Service de Chimie-Physique E.P., Université Libre de Bruxelles, Brussels, Belgium
[*]amalia.villagomez@kuleuven.be

ABSTRACT

Fragmented QRS (fQRS) is an electrocardiographic (ECG) marker of myocardial conduction abnormality, characterized by additional notches in the QRS complex. The presence of fQRS has been associated with an increased risk of all-cause mortality and arrhythmia in patients with cardiovascular disease. However, current binary visual analysis is prone to intra- and inter-observer variability and different definitions are problematic in clinical practice. Therefore, objective quantification of fQRS is needed and could further improve risk stratification of these patients.
We present an automated method for fQRS detection and quantification. First, a novel robust QRS complex segmentation strategy is proposed, which combines multi-lead information and excludes abnormal heartbeats automatically. Afterwards extracted features, based on variational mode decomposition (VMD), phase-rectified signal averaging (PRSA) and the number of baseline-crossings of the ECG, were used to train a machine learning classifier (Support Vector Machine) to discriminate fragmented from non-fragmented ECG-traces using multi-center data and combining different fQRS criteria used in clinical settings.
The best model was trained on the combination of two independent previously annotated datasets and , compared to these visual fQRS annotations, achieved Kappa scores of 0.68 and 0.44, respectively. We also show that the algorithm might be used in both regular sinus rhythm and irregular beats during atrial fibrillation. These results demonstrate that the proposed approach could be relevant for clinical practice by objectively assessing and quantifying fQRS. The study sets the path for further clinical application of the developed automated fQRS algorithm.

1 Introduction

The presence of fragmentation in the QRS complex (fQRS) on a 12-lead electrocardiogram (ECG) has been reported to be an indicator of myocardial scarring or fibrosis[1]. fQRS has been shown to be related to arrhythmic and mortality risk in patients who received an implantable cardioverter-defibrillator (ICD) in primary prevention of sudden cardiac death[2]. Therefore, fQRS is a biomarker of interest in risk stratification of cardiac patients. Currently, the presence of fQRS is diagnosed by visual interpretation of the 12-lead ECG to identify irregularities within the QRS complex. This requires specific training to distinguish

fragmentation from other patterns, such as early repolarization or non-specific conduction delays. Different definitions of fragmentation, which vary slightly in some of the morphologies considered, have been described. The most frequently used criteria are these proposed by Das et al.[1,3], which define morphologies which should be identified as fQRS, both for narrow (120 ms) and broad (>120 ms) QRS complexes. However, the application of these criteria might be ambiguous and might lead to intra- and inter-observer variability[4]. Therefore, other criteria have been proposed to overcome these uncertainties, such as those described by Torigoe et al., 2012, where risk stratification was based on the number of fragmented leads[5], or by Maheshwari et al., 2013, where an automatic approach is proposed to detect different patterns of fQRS[6]. However, these methods have neither been validated nor do they express the severity of fQRS. More recently, Haukilahti et al. defined benign and malignant fQRS variants in an attempt to improve the specificity of fQRS as a biomarker[7].

As the application of different fQRS criteria and inter-observer variability limit the reproducibility of clinical studies in different centers[8], the detection of fQRS may benefit from an automatic signal-based, algorithmic approach, to provide decision support. However, there is no true "golden standard" and automatic algorithms trained using predefined fQRS criteria might reproduce the same ambiguities as human annotations. Some automatic fQRS detection algorithms have previously been described most of which are based on signal decomposition methods, such as wavelet transform or intrinsic time-scale decomposition for the delineation of the QRS complex and detection of fragmentation[6,9,10]. However, all these algorithms were trained based on their own fQRS criteria in single-center data. The most recent automatic algorithm published by our group (Goovaerts et al., 2018), was based on the Das criteria for fQRS detection[11]. This model, developed in a single-center cohort of sinus rhythm ECGs including narrow and broad QRS complexes, offers a continuous scoring output between 0 and 1 for fQRS quantification and is based on Variational Mode Decomposition (VMD)[12] and Phase-Rectified Signal Averaging (PRSA)[13]. Although this algorithm does not distinguish different fQRS patterns, the output was associated with agreement between observers and the risk of all-cause mortality[14].

This work proposes a fully automatic algorithm for fQRS detection and quantification aiming to avoid subjective interpretation by the physician. We present a novel and robust QRS complex segmentation strategy which combines multi-lead information and excludes abnormal heartbeats automatically. Correct segmentation of the QRS complex plays a major role in further analysis since too restrictive segmentation might ignore fQRS patterns in the Q or S waves and the inclusion of artefacts or small oscillations might be confused with fragmentation. Further, fQRS criteria vary for different types of heartbeats[3]. Subsequently, extracted features were used to train a classifier (Support Vector Machine) to discriminate fragmented from non-fragmented ECG-traces. Different models were trained and tested on multicenter data using different fQRS criteria comprising narrow and broad QRS complexes and sinus rhythm and atrial fibrillation (AF). These models were tested and their performance is presented compared to visual fQRS detection.

2 Methods

The block diagram in Figure 1 illustrates the methodology proposed. The algorithm uses all leads for segmentation, but feature extraction and classification are applied to each lead independently. As in the Das criteria for region specific fQRS analysis, all leads, except aVR, were used in the analysis[2,3]. Hence, the output presents 11 fQRS scores per patient's recording. In the remainder of this paper, each lead will be referred to as *signal*, and the set of 11-lead ECG will be referred to as *recording*.

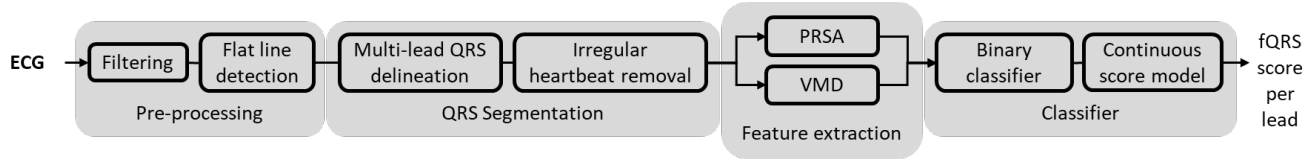

Figure 1. Block diagram of the fQRS scoring algorithm per lead.

2.1 Data

Two independent datasets were used in this study. The first dataset includes 723 digital ECG recordings (10 seconds 12-lead ECG, 250 Hz sampling rate) of patients before ICD implantation in prevention of sudden cardiac death at the University Hospitals Leuven (UZL) in Leuven, Belgium. For this analysis, only recordings in sinus rhythm and AF were included, corresponding to 673 recordings (616 sinus rhythm and 57 AF). The details of these patients are summarized in Table 1.

The presence of fQRS was annotated independently by 5 clinicians[4] (referred here as observers), who assigned a binary label to each of the 11 signals as fragmented (i.e. 1) or non-fragmented (i.e. 0) by using the fQRS criteria for narrow and wide QRS complexes defined by Das et al.[1,3]. Each signal in the database was assigned the sum of the independent labels given by the 5 observers, resulting in a joint label per lead with values ranging between 0 and 5 (Table 2). While labels 0 and 5 represent

full agreement of the clinicians in the absence or presence of fQRS, respectively, the labels in-between represent inter-observer variability. The signals with full agreement were used as ground truth to train the algorithm, as these signals represent the highest level of certainty.

Table 1. Patient characterization for UZL and EU-CERT datasets. Expressed as number of ECG recordings and percentage between brackets, or mean ± standard deviation.

	UZL	**EU-CERT**
Demographics		
Age at implant (y)	62.4 ± 11.6	63.8 ± 10.9
Male gender (%)	570 (84.7)	1006 (79.8)
Ischemic cardiomyopathy (%)	446 (66.3)	869 (68.9)
Left ventricular ejection fraction (%)	32.3 ± 12.3	24.9 ± 6.2
Primary prevention (%)	381 (56.6)	1261 (100)
Sinus rhythm (%)	616 (91.5)	1102 (87.4)
Atrial fibrillation (%)	57 (8.5)	159 (12.6)
Comorbidity		
Diabetes mellitus (%)	119 (17.7)	342 (27.1)
Stroke (%)	67 (10.0)	132 (10.5)
ECG		
Wide QRS (%)	370 (55.0)	496 (39.3)
Heart rate (bpm)	64 ± 13	72 ± 15
Follow-up		
Duration (y)	3.7 ± 3.4	3 ± 2.3
Appropriate Shock (%)	222 (33.0)	165 (13.1)
Death (%)	186 (27.6)	230 (18.2)

The second dataset was provided by the multi-center EUropean Comparative Effectiveness Research to Assess the Use of Primary ProphylacTic Implantable Cardioverter Defibrillators (EU-CERT-ICD) project[15]. This project included a retrospective registry of 5111 patients who received a primary prevention ICD between 2002 and 2014 in one of the 14 participating centers[16]. The cohort presented is based on the detailed ECG analysis reported by Pelli et al., 2020[17]. After further exclusion of rhythms other than sinus and AF, a total of 1560 ECGs were included in the current analysis. This selection contains 402 12-lead recordings from University Hospital Basel, Switzerland (10 seconds, 500Hz, least significant bit resolution of 4V provided by C.S); 29 from Oulu University Hospital, Finland (8 seconds, 1000Hz, 2.5V, by T.K, H.V.H); 551 from University Medical Center Utrecht, Netherlands (10 seconds, 500Hz, 4.88V, by A.T); 277 from University Medical Center Goettingen, Germany (10 seconds, 500Hz, 4V by T.F, M.Z); and 301 from the UZ Leuven, Belgium (10 seconds, 500Hz, 4.88V, by B.V, R.W). Since the ECGs from UZ Leuven are also part of the UZL data, these were excluded from the EU-CERT-ICD database, reducing it to 1259 ECGs. The details of this dataset are described in Table 1.

As described by Pelli et al., 2020[17], the presence of fQRS was annotated independently by two clinicians, who later discussed a final binary label for leads at which they previously disagreed. The narrow QRS complexes were annotated using the criteria defined by Haukilahti et al., 2016[7], while wide QRS complexes were annotated following Das et al., 2008[3]. The binary label 1 of EU-CERT-ICD indicates agreement on the presence of fQRS, and 0 agreement on the absence of fQRS (Table 2). Since this dataset does not contain uncertain labels, the complete set of labels are considered ground truth.

Table 2. Percentage of leads per label in the UZL and EU-CERT datasets, both for sinus and AF files.

		Percentage files per label in UZL					
		0	1	2	3	4	5
UZL	sinus	40.95	13.19	7.23	5.76	8.16	24.70
	AF	41.15	13.56	10.85	5.90	6.06	22.49
EU-CERT	sinus	76.60					23.40
	AF	77.70					22.30

The inter-observer variability between UZL and EU-CERT-ICD labels was assessed by comparing the annotations of the 301 recordings shared by both datasets. This comparison was performed by binarizing the UZL labels for different thresholds from 1 to 5, and deriving the Kappa score κ, for each of these scenarios. The highest agreement ($\kappa = 0.46$) was obtained when

all observers agreed on the presence of fQRS. The relaxation of this threshold by including cases in which fewer observers agreed on fQRS in the positive class implied a linear decrease of κ. Overall, there was only moderate agreement when using the different clinical criteria to assess fQRS. The criteria for fQRS scoring used in the EU-CERT-project were stricter than the Das-criteria used in the UZL-dataset. This could be expected since all signs of fragmentation were scored in UZL, while in EU-CERT-ICD some notches in the QRS complex were not considered as fQRS unless they represented one of the specific patterns described by Haukilahti et al., 2016[7].

2.2 Pre-processing

A two-step pre-processing was applied. First, the signals were band-pass filtered using Butterworth filters of orders 4 and 6, respectively from 0.5 to 70 Hz, to remove baseline wander and high-frequency noise while maintaining the potential fQRS information. Second, each signal was normalized by removing the mean and dividing by the standard deviation of all voltage values.

2.3 QRS segmentation

Segmentation is crucial to properly estimate the level of fragmentation within a QRS complex. While it must be wide enough to capture the information in the Q and S waves, it needs to be robust enough not to confuse these waves with oscillations and noise around the QRS complex.

The first step of a robust QRS segmentation is detecting the R-peaks in the ECG signal, which is followed by identification of the surrounding Q and S waves. In this work, R-peak detection was performed using the algorithm included in the R-DECO software[?]. Since this algorithm treats every lead independently, the multi-lead information is integrated in a consecutive post-processing stage. The starting point is the number of heartbeats n_R in the recording, which should be equal in all signals from a same recording. The value n_R is defined as the most commonly found number of automatically detected R-peaks in all signals of one recording. The aim is to have a set $\mathbf{R}$ of R-peaks with n_R annotations per signal l. If more than n_R R-peaks are detected in a signal, only those n_R complexes closest to the average location of each R_i peak are kept. Conversely, in a given signal l_i in which less than n_R R-peaks are automatically detected, a second search is performed to detect the missed heartbeats by performing a refined maximum search in a window of 40ms in the absolute values of the signal l_i around the average position of the missed R-peak. The pseudo-code for this step is summarized in Algorithm 1.

Algorithm 1 Multi-lead Rpeak detection

Input: Multi-lead ECG recording
Output: Set $\mathbf{R}$ with n_R Rpeaks per lead l_i

1: Detect Rpeaks in each absolute lead, R_{l_i}, using Rpeak detector by Moeyersons et al., 2019[?].
2: Calculate n_R as the most commonly found number of Rpeaks detected in all the leads.
3: **for** the leads in which the size of $R_{l_i} > n_R$ **do**
4: Keep only those n_R closest to the average location of each R_i peak.
5: **end for**
6: **for** the leads in which the size of $R_{l_i} < n_R$ **do**
7: Find missing peaks in absolute signal in a window of 40ms around the average R_i.
8: **end for**

After having identified the set $\mathbf{R}$ with the same number n_R of R-peaks in all leads, the QRS complexes were segmented. In a review by Beraza et al., 2017[19], the segmentation method proposed by Martinez et al., 2004[20], and used in the ECGkit software tool[21], was identified as the most accurate. This method is based on wavelets and was used to detect the onset and offset of the QRS complex for each lead independently. Next, similar as for R-peak detection, the multi-lead information is integrated in a post-processing stage presented in Algorithm 2. Once the QRS complexes are automatically delineated for each signal, the Q and S waves are reallocated per heartbeat to the median of all signals. Thereupon, these locations are adjusted using the information contained in the rest of the heartbeats of the same signal exploiting the periodicity of the signal. Hence, a set of rules is performed per signal. First, the QRS complexes are segmented and normalized to a maximum amplitude of 1. Next, a template is selected per signal, defined as the QRS complex with the highest cumulative correlation to the other QRS complexes within the lead. The rest of the QRS complexes are aligned to this template, and their Q and S locations are adjusted accordingly. Subsequently, a quality-check is performed to remove irregular and erroneously delineated heartbeats. A vector $\rho = [\rho_1, \rho_2, \ldots, \rho_{n_R}]$ is calculated per heartbeat, containing the normalized correlations between it and the remaining heartbeats in the given signal. This vector has n_R elements and compares the given QRS complex to each of the n_R heartbeats in the signal. If more than half of the entries in vector ρ are lower than a quality limit q, where $q \in [0, 1]$, the given heartbeat is removed from further signal analysis. While q can be user selected, it was set here to 0.85 based on visual

inspection by an experienced clinician. An example of the main steps of the segmentation strategy for a signal with irregular heartbeats is illustrated in Figure 2. The QRS segmentation methodology is publicly available together with illustrating demos (https://github.com/avillago/multiLeadSegmentation).

Algorithm 2 Multi-lead QRS segmentation

Input: Multi-lead ECG recording, set **R** with n_R Rpeaks per lead l_i, quality limit q
Output: Final Q,R,S locations

1: Detect Q and S locations in each lead using the algorithm by Martínez et al., 2004[20].
2: Adjust Q and S locations to the median of each point between all leads.
3: **for** each lead **do**
4: Segment each heartbeat and normalize by the maximum.
5: Align all heartbeats using correlation.
6: Select template heartbeat as the one with highest cumulative correlation.
7: Re-align rest of the heartbeats to the template and extract their correlation index ρ_i, l.
8: Remove the heartbeats with $\rho_i < q$ for more than n_R elements.
9: **end for**
10: Extract final Q,R,S points

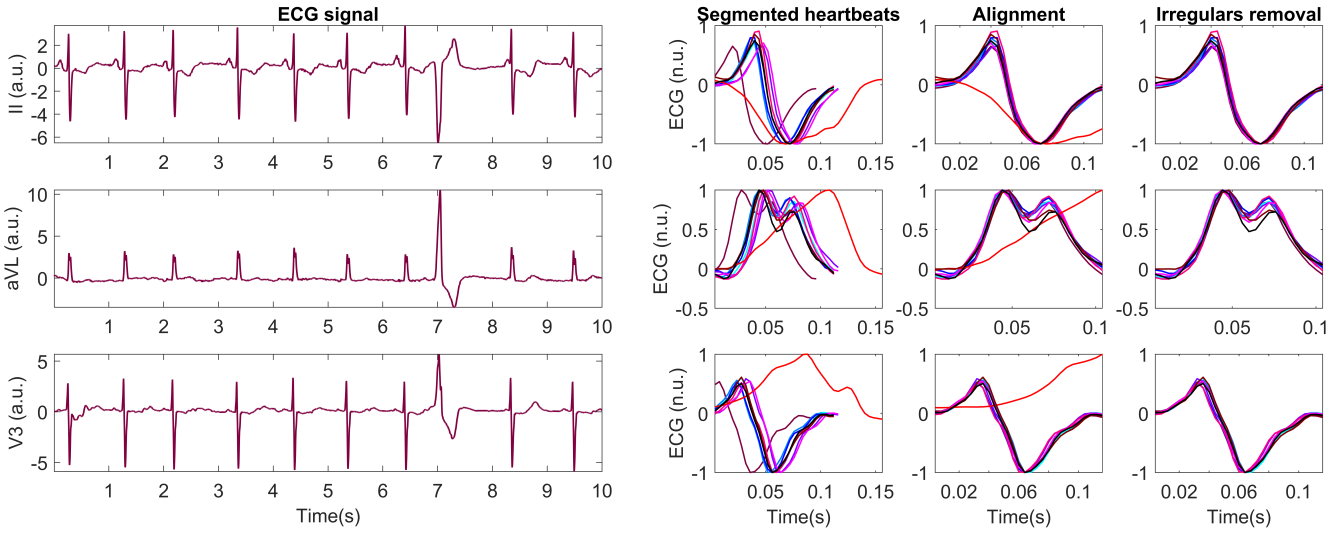

Figure 2. Example QRS segmentation for multi-lead signal with irregular heartbeats, showing the segmentation steps for leads II, aVL and V3. From left to right, the columns show the ECG signal, the segmented heartbeats (step 4 of Algorithm 2), the aligned heartbeats (step 5), and the final segmented heartbeats after removing irregular heartbeats.

2.4 Feature extraction

Feature extraction was previously described by Goovaerts et al., 2018[11], and was performed independently for each signal to capture the oscillations and notches present in the QRS complex. In brief, the feature extraction is based on variational mode decomposition (VMD), phase-rectified signal averaging (PRSA) and the number of zero-crossings of the ECG and is summarized in Table 3. VMD decomposes the signal into its main frequency components according to their level of contribution to the ECG waves and was previously used for heartbeat classification[22] and QRS detection and delineation[11]. Using VMD, each lead of the ECG is decomposed in 5 variational modes which capture most of the signal information. Since fQRS manifests itself as high-frequency notches in the QRS complex, it was shown that the central frequencies and morphologies of the higher-frequency modes were different when comparing fragmented with non-fragmented signals[11]. Therefore, after decomposing the signal into its 5 principal modes and sorting these from lower to high frequency, the central frequencies and number of zero-crossings of the last 3 modes are used as features for classification. PRSA allows detecting and quantifying oscillations in the signal as disturbances of its expected trajectory. A series of PRSA curves is extracted from the QRS complex and is subsequently approximated by a linear fit. Features derived from this linear approximation, such as the mean derivative

of the curve, the slope, and the intercept of the linear fit with the y-axis, have been shown to capture differences between fragmented and non-fragmented signals[23]. The last feature is the average number of zero-crossings per QRS complex in the ECG signal.

Table 3. List of features extracted from each lead.

VMD features	
Central frequencies	Mode 3
	Mode 4
	Mode 5
Number of zero crossings	Mode 3
	Mode 4
	Mode 5
PRSA features	
Slope linear fit	
Y-crossing linear fit	
Mean derivative PRSA	
ECG features	
Number zero crossings	

2.5 Classifier

These 10 features extracted from each signal were used to train a Support Vector Machine (SVM)[24], a machine learning classifier, to discriminate between fragmented and non-fragmented signals. Each signal was introduced as an independent input, with no context of other leads[2]. The binary labels indicating the presence of fQRS, as described in Section 2.1, were used as an established baseline to train the classifier. To account for possible non-linearities in the data, SVM with three different kernels were used to transform the feature space, namely the linear, polynomial and radial basis function (RBF). The parameters required to use them were tuned using Bayesian optimization. The SVM classifiers provide a binary output, which corresponds to fragmented or non-fragmented signals. To quantify fQRS, Platt scaling was used to convert the binary output of the model into a continuous score by fitting a logistic regression to the outputs[25]. By doing so, the model extracts the probability of a signal belonging to the fragmentation class.

2.6 Experiments

The different challenges of this study required independent training strategies and experiments to confirm their added value to the final algorithm. The first goal was to evaluate the impact of the new segmentation algorithm on the quantification of the fQRS. Secondly, we wanted to assess the effect of different fQRS definitions on the training and performance of the classifier. Finally, we evaluated the performance of the model in irregular signals form patients in atrial fibrillation (AF).

Experiment 1: The impact of the new segmentation algorithm

The impact of the proposed QRS segmentation algorithm was assessed by comparing its performance with the method proposed by Goovaerts et al., 2018[11], referred to as the *reference*. Both approaches were trained and tested on the same subset of the UZL dataset. First, training and evaluation was performed using only those signals in which all 5 observers agreed in the presence or absence of fQRS, i.e. labels 0 and 5. Patients were randomly split into training (80%) and test (20%), referred as *test1*. In order to reduce bias, this random division was performed 10 times, leading to training and evaluation with the reference and the proposed methodology in 10 different training-test sets. Next, the best model for each of these 10 repetitions was tested by adding the signals with labels from 1 to 4, both for the reference and proposed algorithm. This group of signals with annotations between 0 and 5 is referred to as *test2*. It allowed a comparison of the continuous score obtained from the algorithm with the certainty of fragmentation, represented by the agreement among the observers.

Experiment 2: The effect of different fQRS annotation criteria

The second experiment aimed at evaluating the effect of different fQRS annotation criteria in the multi-center data on the automated fQRS algorithm. By evaluating the algorithm in both UZL and EU-CERT-ICD datasets, its performance could be assessed from two different perspectives, acknowledging the effect of the different fQRS definitions used. The final goal was to select a model that could be implemented in clinical practice by combining both approaches.

Three different training strategies were used:

1. Training exclusively on UZL. Here, the model selected in Experiment 1 was used, testing it in the corresponding UZL test signals and on the complete EU-CERT-ICD dataset.

2. Training exclusively on 80% of the data from EU-CERT-ICD. Given the lower percentage of fragmented leads in this dataset, a balanced set of signals was used for training the classifier. Therefore, only the same number of fragmented signals present in the training set were randomly selected from all the non-fragmented signals of the training patients. The other 20% of the EU-CERT-ICD patients were used for testing, as well as the complete UZL dataset.

3. The third strategy merged both datasets, using 80% of each dataset for training, while the remaining 20% of each was used as an independent test set. Similar to the previous strategies, out of these 80% of the patients, only the training leads labeled 0 or 5 were used in UZL, and a balanced set of leads was selected for EU-CERT-ICD.

A summary of the data assignment is shown in Figure 3. Given the random selections used to assign signals to the training set, each of these divisions was performed 10 times to reduce bias, and the set leading to the best results was kept in each case.

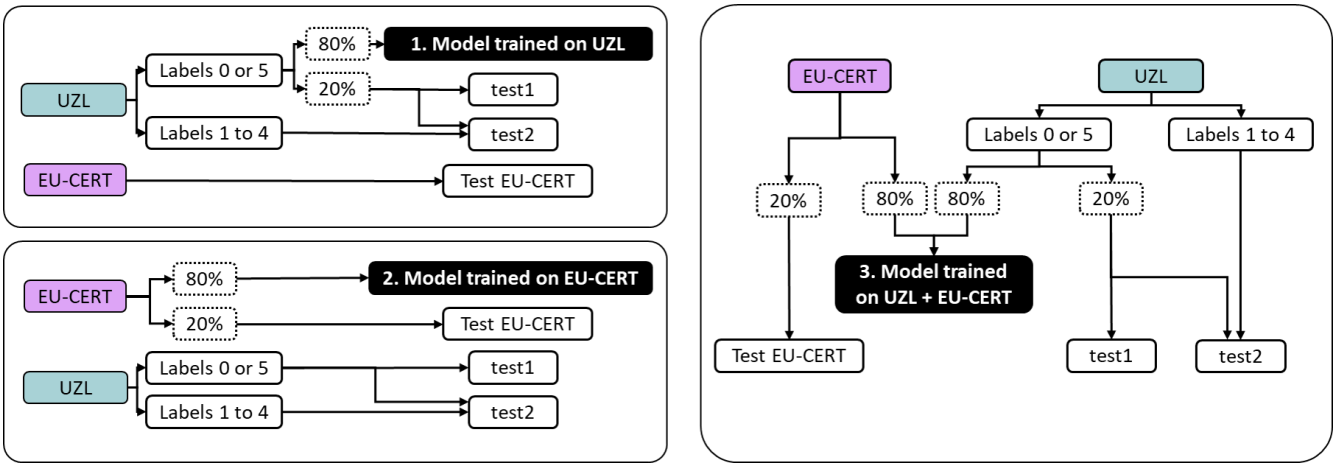

Figure 3. Block diagram describing the training strategies to evaluate the effect of multiple fQRS annotation criteria.

Experiment 3: Using the fQRS algorithm in irregular AF recordings

All experiments previously described were trained and evaluated on ECGs with regular sinus rhythm. Due to the morphology similarities of QRS complexes in sinus and AF signals, the introduced models should also apply to AF signals. In previous works, these signals were excluded from analysis due to their irregularities, since the short interval between heartbeats in AF might pose a problem to this type of algorithms. In the last experiment the best model selected from Experiment 2 was applied to the AF recordings of UZL and EU-CERT-ICD to assess the performance of the algorithm in ECGs with an irregular rhythm.

2.7 Performance metrics

The models were evaluated in test sets with different characteristics. Due to the lack of a "gold standard" for fQRS annotation, the signals scored 0 or 5 from UZL and the binary labels from EU-CERT-ICD based on 2 observers are considered as the established baseline. Therefore, the results were evaluated using sensitivity, specificity, positive predictive value (Sens, Spec and PPV, respectively) and area under the curve (AUC) both for the receiver operator characteristic (ROC) and precision-recall (PR) curves. The latter is particularly relevant in the experiments that involve the EU-CERT-ICD dataset, given that its labels are highly imbalanced with fewer fragmented signals. In such a situation, the PR AUC provides more informative results than the ROC AUC[26]. These metrics were derived by binarizing the algorithm fQRS score using a 0.5 threshold. In the cases in which 10 iterations for each of the 3 kernels proposed were performed, the results of these 6 metrics are provided as the mean and the standard deviation. Additionally, Kappa scores (κ), reflecting the agreement between the algorithm scores and the manual labelling, are presented. The κ scores were calculated several times for each experiment, exploring 100 different thresholds uniformly distributed between 0 and 1. The reported values correspond to the highest score obtained during these iterations. Lastly, the models were also evaluated on the UZL signals where observers did not agree. Since these signals involve uncertainty, the Pearson correlation coefficient between the labels and the continuous output of the models was calculated.

3 Results

3.1 Experiment 1: The impact of the new segmentation algorithm

The average sensitivity, specificity, PPV, ROC AUC and κ score of the 10 training-test combinations for the 3 kernels, both for the proposed and the reference approaches, are shown in Table 4 and Figure 4. The newly proposed approach achieved better results than the reference for all metrics. The difference in results between the 3 kernels were small, indicating that the feature space allowed a linear separation between both classes. Taking this into account, the models selected made use of the linear kernel for simplicity. The mean and standard deviation of the results for each of the 3 kernels are presented in Table S1 of Supplementary Material. Hence, the best training-test set was selected out of the 10 combinations and the result of the classifier using the linear kernel is shown in the second row of each block in Table 4. The PR AUC values were 0.89 and 0.84 for the proposed and reference approaches, respectively. This small difference can be explained by the high percentage of fragmented signals in the UZL test set making it almost balanced.

Table 4. Results of the proposed and reference approach on test1 of UZL including only signals with agreed labels between annotators. The first row of each block shows the average and standard deviation of each metric for all the iterations performed, and the second one shows the results of the selected models.

Algorithm		Sens	Spec	PPV	ROC AUC	κ
Proposed	Average + STD	0.72 ± 0.06	0.92 ± 0.01	0.86 ± 0.02	0.93 ± 0.02	0.68 ± 0.04
	Selected model	0.76	0.92	0.86	0.93	0.71
Reference	Average + STD	0.68 ± 0.02	0.89 ± 0.02	0.8 ± 0.02	0.88 ± 0.01	0.62 ± 0.03
	Selected model	0.68	0.89	0.79	0.88	0.62

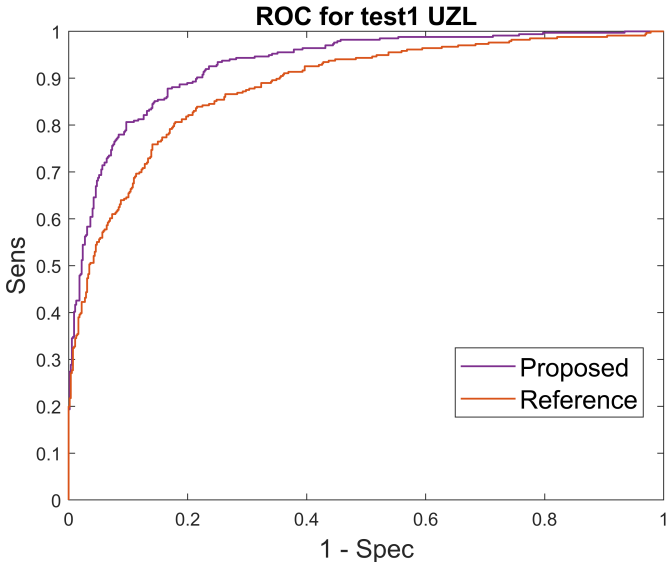

Figure 4. ROC curves for the proposed and reference approaches applied to test1 signals of UZL dataset, exclusively including signals without inter-observer variability.

Next, the continuous outputs were evaluated in the complete UZL-dataset with signals with labels ranging from 0 to 5. The boxplots in Figure 5 show the scores assigned by the proposed (purple) and reference (orange) model to the signals labeled 0 to 5. For both approaches the continuous score increased with the higher labels, illustrating the relation between the automated scores and the inter-observer variability. Pearson correlation coefficients were 0.50 for the proposed approach, and 0.45 for the reference (both $p < 0.001$). Additionally, in the proposed model, the scores assigned to the 0-labelled and 5-labelled signals were more skewed towards 0 and 1 values, respectively, which is reflected in the average scores for each of these labels presented in Table 5. The scores assigned by each of the approaches to these two labels were statistically significantly different both for 0-labelled and 5-labelled. Therefore, we can conclude that compared to the reference algorithm, the proposed approach provides more consistent scores for the signals with inter-observer agreement.

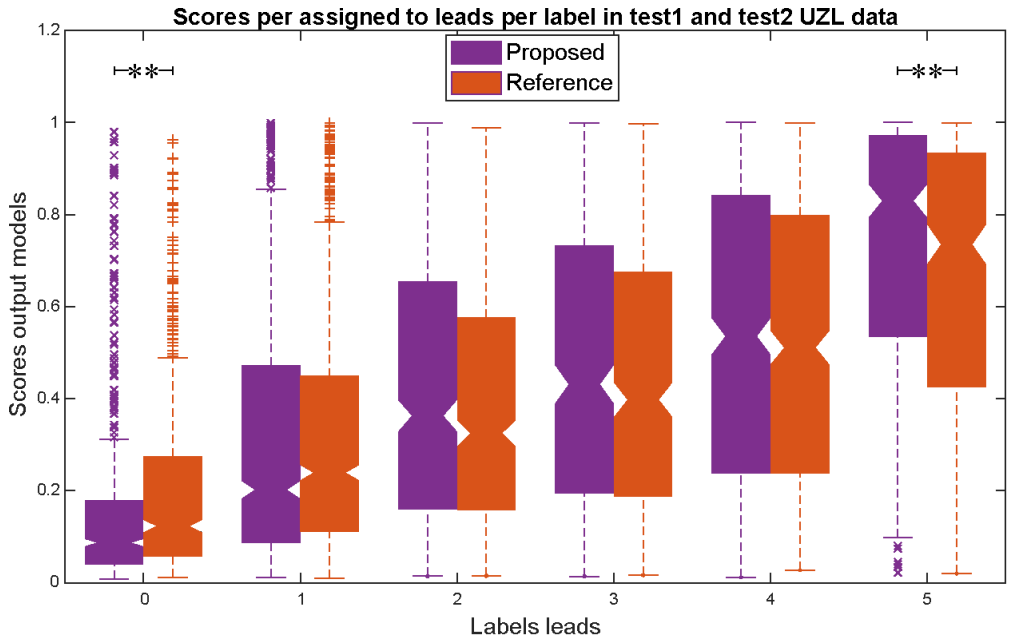

Figure 5. Relation between the output scores of the proposed and reference approaches (in black and grey) with the labels of test2 UZL signals, ranging from 0 to 5. Friedman test was applied to compare the results of both approaches, obtaining p-values below 0.005. These statistically significant differences are described by the sign ** in groups 0 and 5.

Table 5. Summary of the results of the proposed and reference approach in test2. The first two columns show the average ± standard deviation fQRS scores assigned to signals labelled as 0 and 5. The third and forth columns show the correlation coefficients between the fQRS scores and the ground truth labels from 0 to 5, both with p-values below .001.

	0	5	Pearson Corr.	p-value	95% CI
Proposed	0.16 ± 0.2	0.72 ± 0.29	0.50	$\ll .001$	[0.48, 0.53]
Reference	0.2 ± 0.21	0.65 ± 0.3	0.45	$\ll .001$	[0.42, 0.47]

3.2 Experiment 2: The effect of different fQRS annotation criteria

The results of experiment 2 are shown in Table 6 and Figure 6. For UZL test1 signals, the ROC and PR curves were very similar, whereas for EU-CERT-ICD, the PR AUC was lower than the ROC AUC. As expected, the best performance was obtained when the test set belonged to the same center as the data that was used to train the model. The trade-off achieved by the combined model was of interest since the performance for UZL was only mildly reduced whereas the performance in the EU-CERT-ICD test improved when compared to the EU-CERT-ICD trained approach. These variations might be attributed to the lower percentage of fragmented signals and stricter fQRS criteria in EU-CERT-ICD. This was confirmed by the κ scores since the best results on the EU-CERT-ICD data were achieved for thresholds of 0.5 for the model trained on the same dataset, and for thresholds greater than 0.8 for the models that included UZL data in the training stage. On the other hand, the thresholds that led to the best results in UZL were lower than 0.5 (0.42, 0.31 and 0.47 for training strategies 1, 2 and 3, respectively).

Table 6. Results of the 3 training strategies to evaluate the effect of multiple fQRS annotation criteria. From top to bottom, each row shows the results for the approach trained on UZL, on EU-CERT, and in both. The left side of the table shows the results for test1 UZL data, and the one on the right for EU-CERT.

Training strategy	UZL						EU-CERT					
	Sens	Spec	PPV	ROC AUC	PR AUC	κ	Sens	Spec	PPV	ROC AUC	PR AUC	κ
UZL	0.76	0.92	0.86	0.93	0.89	0.71	0.90	0.56	0.37	0.82	0.53	0.41
	0.74 ± 0.02	0.92 ± 0.01	0.84 ± 0.02	0.92 ± 0.01	0.86 ± 0.02	0.69 ± 0.02	0.90 ± 0.00	0.56 ± 0.01	0.37 ± 0.00	0.82 ± 0.00	0.53 ± 0.00	0.41 ± 0.00
EU-CERT	0.59	0.92	0.84	0.88	0.83	0.61	0.77	0.78	0.51	0.85	0.58	0.48
	0.57 ± 0.02	0.92 ± 0.00	0.84 ± 0.01	0.87 ± 0.00	0.83 ± 0.00	0.59 ± 0.02	0.74 ± 0.03	0.77 ± 0.01	0.48 ± 0.02	0.83 ± 0.01	0.55± 0.02	0.46 ± 0.01
UZL + EU-CERT	0.75	0.90	0.83	0.91	0.85	0.68	0.88	0.64	0.42	0.84	0.57	0.44
	0.73 ± 0.02	0.90 ± 0.02	0.83 ± 0.01	0.90 ± 0.01	0.84 ± 0.02	0.66 ± 0.02	0.86 ± 0.02	0.64 ± 0.01	0.42 ± 0.01	0.83 ± 0.01	0.55 ± 0.03	0.44 ± 0.02

Subsequently, the 3 training strategies were evaluated in UZL test2 signals. The continuous scores obtained for each of the

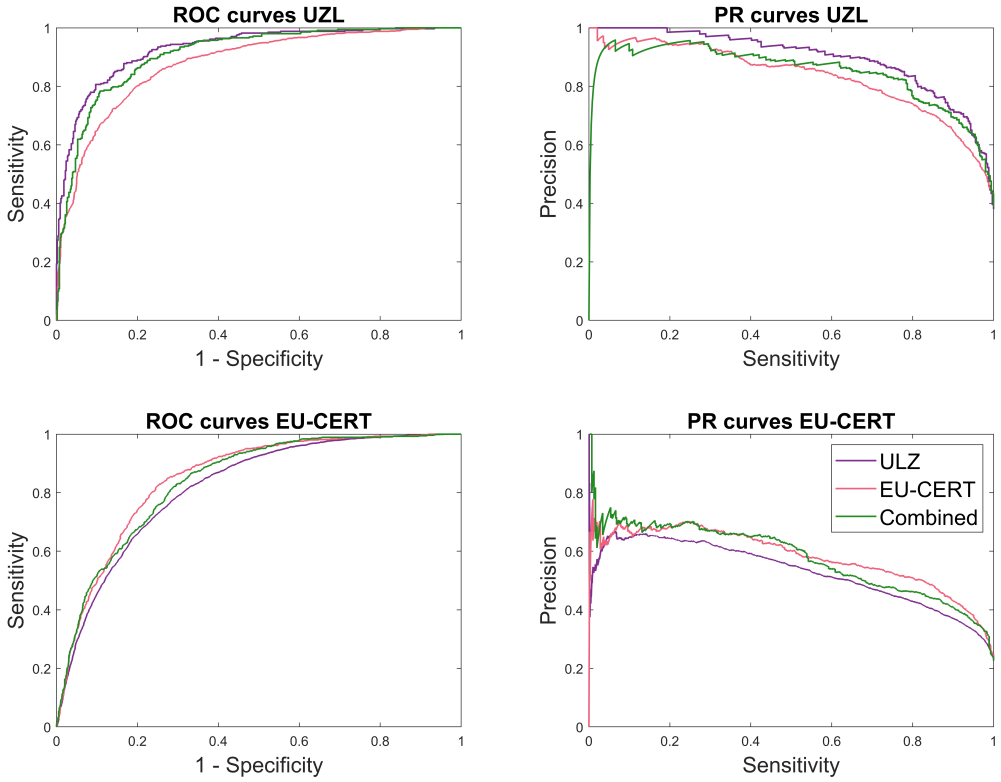

Figure 6. ROC and PR curves for test1 UZL and EU-CERT for the 3 training strategies considered.

3 strategies were compared to the labels using Pearson correlation, obtaining coefficients of 0.5, 0.56 and 0.48 (all $p < 0.001$) for the UZL, EU-CERT-ICD and the combined model, respectively. Table 7 breaks down the results of the combined model by QRS duration both in test1 UZL and EU-CERT-ICD data. There was a lower sensitivity when the algorithm was applied to narrow QRS complexes.

Table 7. Results for the combined model in test1 UZL and EU-CERT broken down by QRS duration.

	UZL						EU-CERT					
QRS duration	Sens	Spec	PPV	ROC AUC	PR AUC	κ	Sens	Spec	PPV	ROC AUC	PR AUC	κ
Narrow	0.69	0.91	0.86	0.91	0.87	0.68	0.86	0.68	0.53	0.85	0.67	0.49
Broad	0.83	0.90	0.80	0.92	0.84	0.73	0.90	0.62	0.33	0.84	0.47	0.44

3.3 Experiment 3: Using the fQRS algorithm in irregular AF recordings

Table 8 shows the results of applying the combined model proposed in experiment 2, to binary-labeled signals of UZL and EU-CERT-ICD with AF where the percentage of fQRS labelled signals is 22% and 33% respectively, compared to previously shown performance in regular sinus rhythm signals. As it can be seen, there was a sensitivity decrease in both datasets, accompanied by a slight increase in specificity and PPV. This indicates that while the detected signal fragmentation was generally correct, some AF signals, that were considered to be fragmented on visual inspection, were missed by the model.

Table 8. Results of the combined model obtained from Experiment 2 in AF signals of both UZL and EU-CERT.

	UZL						EU-CERT					
Test rhythms	Sens	Spec	PPV	ROC AUC	PR AUC	κ	Sens	Spec	PPV	ROC AUC	PR AUC	κ
AF	0.63	0.96	0.89	0.90	0.86	0.63	0.74	0.73	0.43	0.82	0.54	0.43
sinus	0.75	0.90	0.83	0.91	0.85	0.68	0.88	0.64	0.42	0.84	0.57	0.44

4 Discussion

We proposed and validated a new, fully automated machine learning approach for quantification of QRS fragmentation (fQRS) compared to the visual assessment of fQRS by clinical observers. The main contributions of our novel approach are:

- The definition of a new multi-lead algorithm for QRS segmentation, integrating information from different leads for a more robust QRS delineation. This method is publicly available at https://github.com/avillago/multiLeadSegmentation.
- Automatic removal of irregular heartbeats from the fQRS analysis process. This new QRS segmentation strategy includes a method to align the heartbeats in each lead, which allows deriving fQRS results only from normal and atrial premature heartbeats.
- Training and evaluation on a multi-center dataset, including ECG signals from 5 different European hospitals, allowing to assess the impact of inter-observer variability and the use of different fQRS criteria.
- The new algorithm was proven to be usable in both sinus rhythm and atrial fibrillation, as well as in narrow and wide QRS complexes.

The results show that the new QRS segmentation method improved the performance to detect visually confirmed fQRS, when compared to the earlier described reference method. The PPV and specificity indicate that the modified approach reduced the number of false positives. The sensitivity did improve as well. Further, the results indicate a better discrimination of the most evident fQRS cases compared to the reference, since the scores of signals in the UZL-dataset where all observers agreed (i.e. labels 0, as non–fragmented, and 5 as fragmented) were more skewed to the extreme scores 0 and 1[11].

Currently, the only available comparator for our automated approach is the visual assessment of fQRS. This is not only prone to intra- and inter-observer variability[4], but different definitions of fragmentation have also been proposed for clinical practice. The impact of different fQRS criteria on the training and performance of our automated fQRS quantification method are of interest. Previous automated approaches were exclusively trained with a single definition of fQRS and involved fewer observers. In our study, the models trained exclusively on one dataset achieved low results when tested on the other one, due to their differences in fQRS criteria. The fQRS criteria used to score the EU-CERT-ICD dataset considered the magnitude and location of the fragmentation. Annotating only specific fQRS patterns and allowing some benign patterns to be labelled as non-fragmented, result in a more restrictive annotation and a lower prevalence of fQRS. The use of these criteria poses a challenge to the proposed algorithm, since the features used do not include information about the location of the notches detected. The VMD and PRSA features suggested by Goovaerts et al. exclusively inform about the presence of fQRS and are extracted for the entire QRS complex, with no information with regards to its location. This can be inferred from the results of the model trained and evaluated on EU-CERT-ICD data. This EU-CERT-oriented model detected several false positive classifications, as indicated by the lower specificity, PPV and the results of the PR-curves, when compared to the UZL-oriented model. The difference in fQRS criteria explains the opposing behavior in terms of sensitivity and specificity for both datasets. While the models in which UZL data was included in the training detect most of the fragmented signals of EU-CERT-ICD, this also resulted in a higher number of false positive classifications in the EU-CERT-ICD data. On the contrary, an exclusively EU-CERT-ICD-oriented approach achieved low values of sensitivity in the UZL data, but the combined strategy did not compromise the performance on the UZL data while offering a trade-off when testing on the EU-CERT-ICD data. Hence, we propose that the combined model would be the most generally applicable approach to clinical practice, combining the different views on fragmentation.

The breakdown of results in terms of width of the QRS complex was not reported before for other automated approaches, despite its relevance in the fQRS analysis. Our results show a difference in performance depending on the QRS duration of the signals, especially in sensitivity and PPV in UZL data. The reduced sensitivity in narrow QRS complexes may indicate that the segmentation is more accurate in broad QRS complexes, but also that fragmentation is more evident in these signals. After observation of narrow QRS cases annotated as fQRS but receiving a low score from the algorithm, it was observed that many of these correspond to minimal fragmentation. The example in Figure 7 shows a minimal notch after the R-peak which is scored as fQRS when applying the criteria by Das et al., however the automated algorithm calculated a low score and considers this complex non-fragmented. These cases may be missed by the algorithm, particularly in narrow QRS complexes as there may not be enough samples capturing these abnormalities. These cases highlight the relevance of quantifying fQRS. Whether this minimal notching is clinically relevant is unknown and requires further investigation.

Lastly, previous work never evaluated automated fQRS detection and quantification in AF recordings and algorithms were applied exclusively to regular sinus rhythm signals. In our study, the results obtained for these signals are in line with those of narrow sinus signals in UZL data. For EU-CERT-ICD, these results vary more. It should be noted that these results are affected by the low number of AF signals and the low prevalence of fQRS in AF signals.

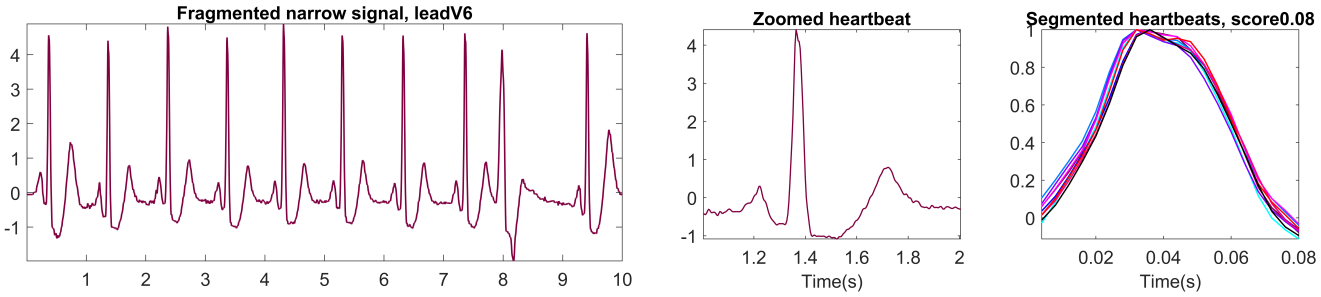

Figure 7. Example of a signal with slight fQRS in the R-peaks, visible in the zoomed heartbeat and in the segmented heartbeats. This is one of the cases in which the labels and the automatic score disagree for narrow QRS signals.

Limitations and future directions

In view of the known limitations of visual inspection of fQRS, the main limitation of our work is the training of our machine learning classifier using an imperfect standard. There is also no clinical consensus on the preferred fQRS definition. We tried to avoid this by using only the signals where all 5 observers agreed in the UZL-dataset thus avoiding inter-observer variability and by combining both criteria in the final model.

Since our goal was to technically validate the algorithm for signal analysis, we processed signals of separate ECG leads individually. In future work the obtained score per signal/lead needs to be summarized in a global score and a score per region/combination of leads (anterior (V1-5), lateral (I, aVL, V6) and inferior (II, III, aVF). We compared and validated the model to clinically used fQRS scoring. In the future, this continuous output of the fQRS algorithm could be used in clinical risk stratification by deriving a combined/regional score and correlating it to the clinical endpoints, e.g., severe ventricular arrhythmia or mortality. Machine learning applications could be useful to overcome the challenge of inter-observer variability and deriving an optimal combination of scores to optimize risk stratification using fQRS. To this end, in the future representation and unsupervised learning techniques could be considered to unveil hidden patterns relevant for this task.

5 Conclusion

This work proposed an automated machine learning method for fQRS detection and quantification. A novel, publicly available multi-lead QRS segmentation method was presented, combining the information of the multi-lead ECG and the different heartbeats in the signal. After segmentation QRS features were extracted, based on VMD and PRSA, they were used to train a machine learning classifier (Support Vector Machines) to discriminate between fragmented and non-fragmented ECG-traces. Data from 5 centers in which 2 different fQRS criteria were applied, were used in the training and evaluation of this new model. The combined model, which was trained using data from multiple centers and annotated with both fQRS criteria, showed to achieve comparable results to the specific models while making the model more flexible and more generally applicable to multi-center data, bringing it closer to clinical practice. The algorithm was shown to be applicable to narrow and wide QRS complexes and regular and irregular rhythms, such as AF. These findings are crucial for the next steps to apply these machine learning models in clinical practice and assist clinicians in their diagnosis.

Supplementary material

Table S1. Results for Experiment 1 for each of the 3 kernels considered in the SVM classifier. The results are expressed as mean ± standard deviation, summarizing the 10 repetitions performed per kernel.

Approach	Kernel	Sens	Spec	PPV	ROC AUC	PR AUC	κ
Proposed	Linear	0.74 ± 0.02	0.92 ± 0.01	0.84 ± 0.02	0.92 ± 0.01	0.86 ± 0.02	0.69 ± 0.02
	Polynomial	0.70 ± 0.09	0.92 ± 0.01	0.84 ± 0.03	0.91 ± 0.03	0.85 ± 0.04	0.67 ± 0.06
	RBF	0.73 ± 0.02	0.92 ± 0.01	0.84 ± 0.02	0.92 ± 0.01	0.87 ± 0.02	0.69 ± 0.02
Reference	Linear	0.67 ± 0.03	0.89 ± 0.02	0.78 ± 0.02	0.88 ± 0.01	0.82 ± 0.02	0.62 ± 0.02
	Polynomial	0.67 ± 0.07	0.90 ± 0.02	0.80 ± 0.02	0.89 ± 0.02	0.83 ± 0.02	0.62 ± 0.04
	RBF	0.69 ± 0.04	0.90 ± 0.02	0.81 ± 0.02	0.88 ± 0.01	0.83 ± 0.02	0.63 ± 0.02

EU-CERT-ICD Study Investigators

Department of Cardiology, Semmelweis University Heart Center, Budapest/Hungary: Béla Merkely MD, Peter Perge MD, Zoltan Sallo MD, Gabor Szeplaki MD, Nandor Szegedi MD, and Klaudia Vivien Nagy MD; Department of Cardiology and Pneumology, Heart Center, University Medical Center, Göttingen, Germany: Markus Zabel MD, Lars Lüthje MD, Simon Schlögl MD, Rajeeva Sritharan MSc, Helge Haarmann MD, Leonard Bergau MD, Joachim Seegers MD, Gerd Hasenfuß MD, Pascal Munoz-Exposito MD, Tobias Tichelbäcker MD, and Aleksandra Kirova MD; DZHK (German Center for Cardiovascular Research), partner site Göttingen, Göttingen, Germany: Gerd Hasenfuß MD, Tim Friede MD, Markus Zabel MD, and Simon Schlögl MD; Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany: Tim Friede PhD and Markus Harden, PhD; National Heart and Lung Institute, Imperial College, London, United Kingdom: Marek Malik MD, and Katerina Hnatkova PhD; Department of Medical Physiology, University Medical Center Utrecht: Marc A. Vos PhD; Institute for Social Medicine, Epidemiology and Health Economics, Charité Universitätsmedizin Berlin, Berlin, Germany: Stefan N. Willich MD and Thomas Reinhold PhD; University Hospitals of Leuven, Leuven, Belgium: Rik Willems MD and Bert Vandenberk MD; Magdalena Klinika, Department of Cardiology, Krapinske Toplice, Croatia: Janko Szavits-Nossan MD and L. Rotkvić MD; Attikon University Hospital, 2nd Department of Cardiology, Athens, Greece: Panayota Flevari MD, Andreas Katsimardos MD, and Dimitrios Katsaras MD; Slovak Medical University NUSCH, Bratislava, Slovakia: Robert Hatala MD and Martin Svetlosak MD; Medical University of Lodz (MUL) WAM Hospital, Department of Cardiology, Lodz, Poland: Andrzej Lubinski MD and Tomasz Kuczejko MD; Gentofte Hospital, Copenhagen, Denmark: Jim Hansen MD; University Hospital, Department of Cardiology, Basel, Switzerland: Christian Sticherling MD and David Conen MD; KBC Sestre Milosrdnice, Department of Cardiology, Zagreb, Croatia: Nikola Pavlović MD, Šime Manola MD, Ozren Vinter MD, and Ivica Benko RN; University Medical Center Utrecht, Department of Cardiology/Physiology: Anton Tuinenburg, MD, David Sprenkeler, MD, Agnieszka Smoczynska, MD, and Marc A. Vos PhD; University Hospital Tübingen, Department of Cardiology: Axel Bauer MD, Christine Meyer-Zürn MD, and Christian Eick MD; Rigshospitalet, The Heart Centre, Department of Cardiology, Copenhagen University Hospital, and Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark: Jesper Hastrup Svendsen MD; IDIBAPS, Department of Cardiology, Hospital Clinic Barcelona, Spain: Josep Brugada MD and Elena Arbelo MD; SUSSCH, Department of Cardiology, Banska Bystrica, Slovakia: Gabriela Kaliska MD and Jozef Martinek PhD; Technische Universität München, Med. Klinik und Poliklinik I, Klinikum rechts der Isar, Munich, Germany: Georg Schmidt MD, Michael Dommasch MD, and Alexander Steger MD; Klinikum Großhadern, Department of Cardiology, Ludwig-Maximilians-Universität Munich, Germany: Stefan Kääb MD, Axel Bauer MD, Moritz F. Sinner MD, Konstantinos D Rizas MD, and Wolfgang Hamm MD; Acibadem City Clinic Tokuda Hospital, Department of Cardiology: Vassil Traykov MD; Medical University of Lodz (MUL), CKD Hospital, Department of Cardiology, Lodz, Poland: Iwona Cygankiewicz MD, Pawel Ptaszyński MD, K. Kaczmarek MD, and I. Poddebska MD; St. Ekaterina University Hospital, Department of Cardiology, Sofia, Bulgaria: Svetoslav Iovev MD; University Hospital Brno, Department of Internal Medicine and Cardiology, Brno, Czech Republic: Tomáš Novotný MD and Milan Kozak MD; Oulu University Hospital and University of Oulu, Medical Research Center, Finland: Heikki Huikuri MD, Tuomas Kenttä MD, and Ari Pelli MSc; Bieganski Hospital, Chair and Department of Cardiology, Medical University of Lodz (MUL), Lodz, Poland: Jaroslaw D. Kasprzak MD and Dariusz Qavoq MD; KBC Rijeka, Department of cardiovascular disease, Rijeka, Croatia: Sandro Brusich MD, Ervin Avdovic MD, and Marina Klasan RN; University Hospital, Department of Cardiology, Olomouc, Czech Republic: Jan Galuszka MD, and Milos Taborsky MD; St. Anna Hospital, Department of Cardiology, Sofia, Bulgaria: Vasil Velchev MD; Klinikum Reinkenheide Bremerhaven, Department of Cardiology, Germany: Rüdiger Dissmann MD; National Heart Hospital, Department of Cardiology, Sofia, Bulgaria:Tchavdar Shalganov MD; Poznan Medical University, HSUH Hospital Department of Cardiology, Poznan, Poland: P. Guzik, T. Krauze; Marienkrankenhaus Bonn, Department of Cardiology, Germany: Dieter Bimmel MD and Christiane Lieberz RN; Klinikum Ludwigsburg, Department of Cardiology, Germany: Stefan Stefanow MD, Norman Rüb MD, and Christian Wolpert MD; University Hospital Regensburg, Department of Cardiology, Germany: Joachim Seegers MD and Lars S. Meier MD; Vivantes Humboldt Klinikum Berlin, Department of Cardiology: Steffen Behrens MD; KBC Split, Department of Cardiology, Croatia: Zrinka Jurisic; Karolinska Institutet, Department of Cardiology, Stockholm, Sweden: Frieder Braunschweig MD; Charité Campus Virchow Klinikum, Department of Cardiology, Berlin, Germany: Florian Blaschke MD and Burkert Pieske MD; General Hospital Zadar, Croatia: Zoran Bakotic MD and Ante Anic MD; Klinikum Weiden, Department of Cardiology, Germany: Robert H.G. Schwinger MD; Lund University Hospital, Department of Cardiology, Lund, Sweden: Pyotr Platonov MD.

References

1. Das, M. K., Khan, B., Jacob, S., Kumar, A., Mahenthiran, J. (2006). Significance of a fragmented QRS complex versus a Q wave in patients with coronary artery disease. *Circulation*, **113(21),** 2495-2501.
2. Vandenberk, B., et al. (2017). 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*,

243, 223-228.

3. Das, M. K., et al. (2008). Fragmented wide QRS on a 12-lead ECG: a sign of myocardial scar and poor prognosis. *Circulation: Arrhythmia and Electrophysiology*, **1(4),** 258-268.
4. Vandenberk, B., et al. (2018). Inter-and intra-observer variability of visual fragmented QRS scoring in ischemic and non-ischemic cardiomyopathy. *Journal of electrocardiology*, **51(3),** 549-554.
5. Torigoe, K., et al. (2012). The number of leads with fragmented QRS is independently associated with cardiac death or hospitalization for heart failure in patients with prior myocardial infarction. *Journal of cardiology*, **59(1),** 36-41.
6. Maheshwari, S., et al. (2013). An automated algorithm for online detection of fragmented QRS and identification of its various morphologies. *Journal of the Royal Society Interface*, **10(89),** 20130761.
7. Haukilahti, M. A. E., Eranti, A., Kenttä, T., Huikuri, H. V. (2016). QRS fragmentation patterns representing myocardial scar need to be separated from benign normal variants: hypotheses and proposal for morphology based classification. *Frontiers in physiology*, **7,** 653.
8. Malik, M. (2013). Electrocardiographic smoke signals of fragmented QRS complex. *Journal of Cardiovascular Electrophysiology*, **24(11),** 1267-1270.
9. Bono, V., et al. (2013). Development of an automated updated selvester QRS scoring system using SWT-based QRS fractionation detection and classification. *IEEE journal of biomedical and health informatics*, **18(1),** 193-204.
10. Jin, F., Sugavaneswaran, L., Krishnan, S., Chauhan, V. S. (2017). Quantification of fragmented QRS complex using intrinsic time-scale decomposition. *Biomedical Signal Processing and Control*, **31,** 513-523.
11. Goovaerts, G., et al. (2018). A machine-learning approach for detection and quantification of QRS fragmentation. *IEEE journal of biomedical and health informatics*, **23(5),** 1980-1989.
12. Dragomiretskiy, K., Zosso, D. (2013). Variational mode decomposition. *IEEE transactions on signal processing*, **62(3),** 531-544.
13. Bauer, A., et al. (2006). Phase-rectified signal averaging detects quasi-periodicities in non-stationary data. *Physica A: Statistical Mechanics and its Applications*, **364,** 423-434.
14. Goovaerts, G., et al. (2018, September). Risk Assessment of All-Cause Mortality in ICD Patients Using a Novel QRS Fragmentation Score. In *2018 Computing in Cardiology Conference (CinC)* (Vol. **45,** pp. 1-4). IEEE.
15. Zabel, M., et al. (2020). Clinical effectiveness of primary prevention implantable cardioverter-defibrillators: results of the EU-CERT-ICD controlled multicentre cohort study. *European heart journal*, **41(36),** 3437-3447.
16. Sticherling, C., et al. (2018). Sex differences in outcomes of primary prevention implantable cardioverter-defibrillator therapy: combined registry data from eleven European countries. *Ep Europace*, **20(6),** 963-970.
17. Pelli, A., et al. (2020). Electrocardiogram as a predictor of survival without appropriate shocks in primary prophylactic ICD patients: A retrospective multi-center study. *International journal of cardiology*, **309,** 78-83.
18. Moeyersons, J., Amoni, M., Van Huffel, S., Willems, R., Varon, C. (2019). R-DECO: An open-source Matlab based graphical user interface for the detection and correction of R-peaks. *Peerj computer science*, **5,** e226.
19. Beraza, I., Romero, I. (2017). Comparative study of algorithms for ECG segmentation. *Biomedical Signal Processing and Control*, **34,** 166-173.
20. Martínez, J. P., Almeida, R., Olmos, S., Rocha, A. P., Laguna, P. (2004). A wavelet-based ECG delineator: evaluation on standard databases. *IEEE Transactions on biomedical engineering*, **51(4),** 570-581.
21. Demski, A., Soria, M. L. (2016). ecg-kit: a Matlab toolbox for cardiovascular signal processing. *Journal of open research software*, **4(1)**.
22. Villa, A., Padhy, S., Willems, R., Van Huffel, S., Varon, C. (2018, September). Variational mode decomposition features for heartbeat classification. In *2018 Computing in Cardiology Conference (CinC)* (Vol. **45,** pp. 1-4). IEEE.
23. Goovaerts, G., Vandenberk, B., Varon, C., Willems, R., Van Huffel, S. (2016, September). Phase-rectified signal averaging for automatic detection of QRS fragmentation. In *2016 Computing in Cardiology Conference (CinC)* (pp. 637-640). IEEE.
24. Cortes, C., Vapnik, V. (1995). Support vector machine. *Machine learning*, **20(3),** 273-297.
25. Platt, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, **10(3),** 61-74.
26. Ozenne, B., Subtil, F., Maucort-Boulch, D. (2015). The precision–recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. *Journal of clinical epidemiology*, **68(8),** 855-859.

Acknowledgements

BV is supported by a research grant of the Frans Van de Werf Fund for Clinical Cardiovascular Research. RW is supported as postdoctoral clinical researcher by the Fund for Scientific Research Flanders (FWO Vlaanderen). SVH, AV and CV received funding from the Flemish Government (AI Research Program) and are affiliated to Leuven.AI - KU Leuven institute for AI, B-3000, Leuven, Belgium. EU-CERT-ICD is funded by the European Commission within the 7th Framework Programme under Grant Agreement n°602299.

Author contributions statement

All authors contributed to the study meaningfully, providing feedback on the interpretation of the results and on the writing of the paper. AV conducted the study, analysis, implementation of the methodologies and the algorithms. BV, TK, SI, HVH, MZ, TF, CS, AT, MM and RW provided the data and the medical feedback during the analysis. AV, BV, SVH, RW and CV designed the experiments. CV supervised the study. AV wrote the original draft manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.V. The **codes** for multi-lead QRS segmentation are available at https://github.com/avillago/multiLeadSegmentation.