Towards the Prediction of Atrial Fibrillation Using Interpretable ECG Features

Alexander Hammer, Hagen Malberg, Martin Schmidt

Institute of Biomedical Engineering, TU Dresden, Dresden, Germany

Abstract

Atrial fibrillation (AF) is our society's most common cardiac arrhythmic disease, leading to increased morbidity and mortality. Predicting AF episodes during sinus rhythm based on electrocardiograms (ECGs) allows timely interventions. It is known, that changes in selected ECG morphology features are a predictor for the onset of AF, but no systematic investigation of different ECG features' temporal changes has been performed so far. We split sinus rhythm episodes of 60 minutes preceding AF from the MIT-BIH AF database into segments of 5 minutes with 50% overlap ($n = 644$) and calculated 155 features of different domains per segment. Logistic regression analyses between the segments preceding AF and others revealed the most significant effects for segments ending 5 minutes before AF onset, with PQ interval slope ($p < 0.01$), PQ interval correlation ($p < 0.05$), and median RR time ($p < 0.05$) being the most relevant features. A decision tree ensemble, trained with all features, achieved an accuracy of 0.87 when distinguishing 8 segment clusters. Our results confirm expected changes in ECG features (e.g., PQ interval) before AF episodes, indicating impaired atrial excitation, and show that the combination of interpretable features is sufficient to discriminate at different points in time before AF onset. For advanced analyses, more extensive databases should be included.

1. Introduction

Atrial fibrillation (AF) is the leading cardiac arrhythmic disease globally, leading to increased morbidity and mortality [1,2]. Paroxysmal AF occurs in episodes of less than seven days but often becomes persistent. With early detection of AF episodes, interventions can avoid chronification and decrease morbidity and mortality, which is why predicting AF episodes during sinus rhythm based on electrocardiograms (ECGs) is highly important.

The transition from normal sinus rhythm to AF is not fully understood but is often associated with changes in atrial electrophysiological properties and thus with discontinuous atrial pacing [3]. Therefore, we hypothesize that there is increased variability in beat morphology and heart rate before the onset of AF. It is already known that the variability in P wave duration increases before AF onset [4]. However, to our knowledge, the temporal course of different features describing other characteristics has not yet been investigated. Therefore, we examine the temporal change before AF onset and the predictive value of various interpretable ECG features from different domains.

2. Methods

To examine the predictability of AF using interpretable features, we split normal sinus rhythm episodes that preceded AF and calculated features of different domains for each segment. Using logistic regression, we studied the relevance of each feature to discriminate between the segment before the AF onset (preAF) and those further ahead. Furthermore, we trained a decision tree ensemble to investigate the predictive value of combined features. Figure 1 shows the processing and validation pipeline. The implementation was entirely done with Matlab R2021b (MathWorks Inc., Natick, MA, USA).

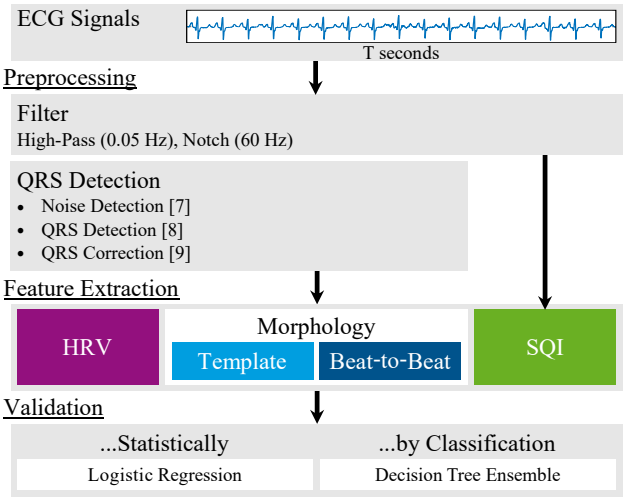

Figure 1. Pipeline for feature extraction and validation.

2.1. Data

We used the first lead of the MIT-BIH AF database [5,6] containing 23 long-term two-leads ECGs (10 hours each) with AF episodes of primarily paroxysmal AF and rhythm annotations. The database included 274 sinus rhythm episodes of 21 subjects preceding AF. As a trade-off between period size and data set size, we only used sinus rhythm episodes of at least one hour in length. From each of the remaining 28 episodes, we extracted 23 segments with a duration of 5 minutes and 50% overlap, concluding with the onset of AF.

2.2. Preprocessing

All ECGs were high-pass and notch filtered at 0.05 Hz and 60 Hz, respectively, with previous zero-padding to avoid boundary effects. In filtered ECGs, noisy parts were detected [7] and ignored for QRS detection [8] with subsequent QRS correction, based on amplitude heights and signs as well as peak-to-peak and peak-to-signal edge distances [9].

2.3. Feature Extraction

We extracted 155 features of different domains, including 27 heart rate variability (HRV) features, 113 morphology features and 15 signal quality indices.

2.3.1. Heart Rate Variability Features

We extracted RR intervals from consecutive QRS complexes, filtered physiological and non-physiological RR intervals, and added the filtration ratio (filt rate) to the feature set. The physiological RR intervals were used to calculate standard heart rate metrics, including mean, median, minimum, and maximum, and statistical, geometric, non-linear, and frequency-based HRV features [10].

2.3.2. Morphology Features

To extract morphological features, we extracted beats b_j according to [11] between onset

$$t_{b_j^{on}} = \begin{cases} t_{R_j} - 370\,\text{ms}, & \overline{RR} \geq 500\,\text{ms} \\ t_{R_j} - 1/2\overline{RR}, & \overline{RR} < 500\,\text{ms} \end{cases}, \quad (1)$$

which was fitted to be dependent on $\overline{RR}$, and offset

$$t_{b_j^{off}} = \begin{cases} t_{R_{j+1}} - 240\,\text{ms}, & \overline{RR} \geq 720\,\text{ms} \\ t_{R_j} + 2/3\overline{RR}, & \overline{RR} < 720\,\text{ms} \end{cases}, \quad (2)$$

where $\overline{RR}$ is the length and t_{R_j} is the location of R peak j. The beats were averaged to create a template beat, and fiducial points were detected using the *ECGdeli* toolbox [12]. Subsequently, we extracted the height, length, baseline slope, area, and skew of the P and T waves, as well as the QRS complex. Length and slope were also determined for the PQ and QT intervals.

To consider morphological beat-to-beat changes, we calculated the mean, median, and standard deviation (SD) between beats of the same feature, using two-dimensional signal warping [13]. Furthermore, we extracted standard QT interval variability (QTV) features, including SD of QT intervals and the QTV index (QTVi) as well as the T wave amplitude corrected measures cQTV, cSDQT, and cQTVi [14]. Additionally, we used the mean, median, minimum, and maximum QT length. In total, we received 22 template features and 91 beat-to-beat features.

2.3.3. Signal Quality Indices

We considered the signal quality entity by applying different SQIs from the *fecgsyn* toolbox [15] to the complete but filtered signals.

2.4. Validation

We performed logistic regression analyses for each segment versus preAF for each feature separately after removing the baseline and outliers. The baseline was defined feature dependent as the mean of the three segments closest to AF, and outliers were defined as values that were more than three scaled median absolute deviations away from the median, using the Matlab function `isoutlier`.

Furthermore, we examined the predictive value of feature combinations using decision tree ensembles based on the GentleBoost algorithm. To reduce the number of classes, we clustered segments as shown in Figure 2, with the cluster size increasing with distance from the AF episode. We set the learning rate to 0.1 and the number of bins to 256. To determine the optimal model design, we applied a grid search, varying the number of learners between 128 and 768 and the number of splits between 2 and 4. The models were validated using 10-fold cross-validation and the following measures: accuracy, the area under the precision-recall curve (AUPRC), and F1.

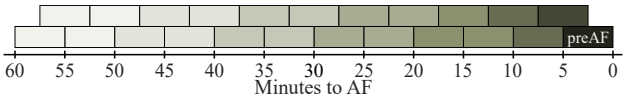

Figure 2. Clusters of segments, between 60 minutes before atrial fibrillation (AF) to the segments preceding AF (preAF), for training the decision tree ensemble.

3. Results

Logistic regression analyses showed the greatest number of significant ($p < 0.05$) effects for segments that ended 5 minutes before AF onset ($n = 21$), followed by segments that ended 7.5 ($n = 18$), 40 ($n = 18$), and 42.5 ($n = 19$) minutes before AF onset. Other segments showed significant effects but less frequently.

Maximum RR ($n = 15$) and the HRV feature of approximate entropy ($n = 13$) were the most relevant features, according to the number of significant segments.

For segments that ended 5 minutes before AF onset, PQ interval slope ($p < 0.01$), median PQ correlation ($p < 0.05$), and median RR time ($p < 0.05$) were indicated as most relevant (see Table 1). However, we found significant effects for all domains' features, dominated by P/QRS/T wave features, with the frequency bands' relative power of 5-20 Hz to 5-45 Hz (p-SQI) being the only SQI.

Table 1. Features for which logistic regression found significant differences between segments ending five minutes before versus just before the AF onset, with feature domains highlighted by color and the sign of slope between both segments. Standard deviation is abbreviated by SD.

Rank	Feature	Metric	p value	Sign of slope
1	PQ slope		0.004	+
2	PQ correlation	median	0.011	+
3	RR	median	0.012	+
4	P area-to-length	mean	0.014	+
5	QRS area slope	median	0.017	-
6	QRS area slope	mean	0.017	-
7	PQ correlation	mean	0.018	+
8	T height	mean	0.019	+
9	T height		0.020	+
10	QRS area slope		0.027	-
11	T baseline slope	SD	0.029	+
12	QRS area slope	SD	0.030	-
13	RR	maximum	0.031	+
14	PQ slope	mean	0.031	+
15	RR	mean	0.035	+
16	P area		0.036	+
17	T height	median	0.040	+
18	QRS area-to-length		0.041	+
19	T area		0.045	+
20	P area-to-length	median	0.046	+
21	p-SQI		0.048	-

Legend: Heart rate variability (HRV); Morphology: Template, Beat-to-beat; Signal quality indices (SQIs)

According to the grid search (see Figure 3), the classification performance for each lead set increases with model complexity up to a certain point. The optimum was achieved with 640 learners and 4 splits ($accuracy = 0.87$; $F1 = 0.45$; $AUPRC = 0.35$).

For both settings, classifying all clusters or only preAF versus the rest, HRV, beat-to-beat, and SQI features were among the ten most relevant features (see Figure 4) with an overlap of 6 out of 10. Template-based features were represented only once with the P wave baseline's slope for all clusters and not for preAF versus the rest. Besides beat-to-beat P wave and PQ interval features, the most important features were dominated by HRV features, including the HRV index, which is the ratio of the SD to the mean of the heart rate and the median RR interval [10]. In addition to the SD along the identity line (SD2) and its perpendicular (SD1) from the Poincaré plot, the dominant HRV features furthermore include the density of low frequencies (LF), very low frequencies (VLF), and high frequencies (HF) as well the proportion of HF (pHF) from the interpolated RR tachogram [10].

4. Discussion and Conclusion

The dominance of P wave and PQ interval features in both analyzes confirms the assumption of altered electrophysiological properties in the atrium. However, particularly for the classification, HRV and beat-to-beat features of T wave and QRS complex were relevant, indicating that self-terminating circling excitations occur shortly before the onset of AF. Circling excitations lead to a superimposition of all beat characteristics with F waves and an irregular transmission of the excitation to the ventricles, which is reflected in an increased heart rate (variability). This is supported by the increase in median, mean, and maximum RR to preAF as indicated by the sign of slope in Table 1.

Using the decision tree ensemble, we achieved decent accuracies of up to 0.87 in distinguishing between different clusters of segments. F1 score and AUPRC, on the other hand, were lower, which is because our classifier has high specificity but low sensitivity. That means the rate of correctly as *others* classified segments is high, but the rate of correctly classified *correct* segments is low.

Reasons for this can be the small data set, from which we could only use 28 episodes, and the sampling frequency below the recommendation for extracting morphological features. Furthermore, a visual examination of the signals

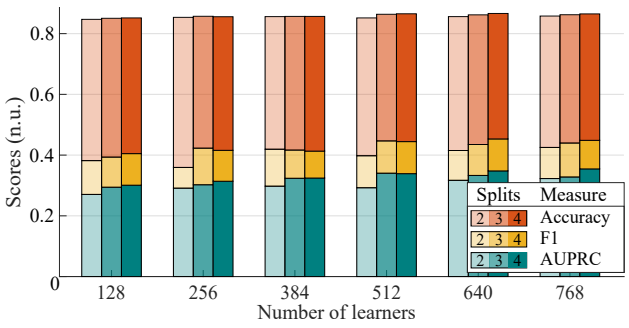

Figure 3. Results of the grid search for different numbers of learners and splits, including accuracy, F1 score, and area under precision recall curve (AUPRC).

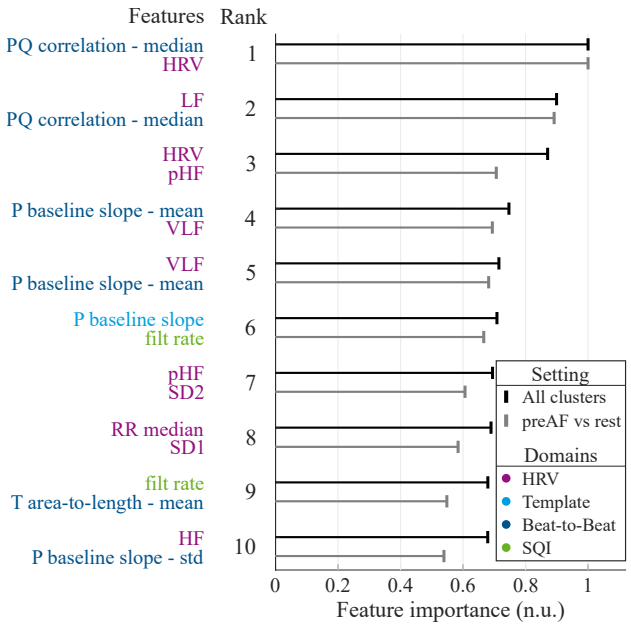

Figure 4. Relative importance of the 10 most important features in classifying all clusters of segments or the segment preceding AF (preAF) versus the rest.

revealed that episodes annotated as sinus rhythm sometimes contain AF.

Our study demonstrates the potential of feature-based approaches using machine learning methods for the interpretable prediction of AF from ECG. For this purpose, however, future studies should include more extensive databases to verify the value of the respective ECG features for AF prediction on an extended time horizon.

Acknowledgments

This study was supported by grants from the European Union's Horizon 2020 research and innovation program (TIMELY, No. 101017424).

References

[1] Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ, Gillum RF, Kim YH, McAnulty JH, Zheng ZJ, Forouzanfar MH, Naghavi M, Mensah GA, Ezzati M, Murray CJ. Worldwide Epidemiology of Atrial Fibrillation. Circulation February 2014;129(8):837–847.

[2] Tanaka Y, Shah NS, Passman R, Greenland P, Lloyd-Jones DM, Khan SS. Trends in Cardiovascular Mortality Related to Atrial Fibrillation in the United States, 2011 to 2018. Journal of the American Heart Association August 2021; 10(15):e020163.

[3] Dilaveris PE, Gialafos JE. P-Wave Dispersion: A Novel Predictor of Paroxysmal Atrial Fibrillation. Annals of Noninvasive Electrocardiology The Official Journal of the International Society for Holter and Noninvasive Electrocardiology Inc April 2001;6(2):159–165.

[4] Martínez A, Alcaraz R, Rieta JJ. Study on the P-Wave Feature Time Course as Early Predictors of Paroxysmal Atrial Fibrillation. Physiological Measurement November 2012; 33(12):1959–1974.

[5] Moody GB, Mark RG. A New Method for Detecting Atrial Fibrillation Using R-R Intervals. Computers in Cardiology 1983;10:227–230.

[6] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. Physiobank, Physiotoolkit, and Physionet: Components of a New Research Resource for Complex Physiologic Signals. Circulation June 2000;101(23):e215–e220.

[7] Datta S, Puri C, Mukherjee A, Banerjee R, Choudhury AD, Singh R, Ukil A, Bandyopadhyay S, Pal A, Khandelwal S. Identifying Normal, AF and Other Abnormal ECG Rhythms Using a Cascaded Binary Classifier. In 2017 Computing in Cardiology. September 2017; .

[8] Johnson AE, Behar J, Andreotti F, Clifford GD, Oster J. R-Peak Estimation Using Multimodal Lead Switching. In Computing in Cardiology 2014. September 2014; 281–284.

[9] Hammer A, Scherpf M, Ernst H, Weiß J, Schwensow D, Schmidt M. Automatic Classification of Full- And Reduced-Lead Electrocardiograms Using Morphological Feature Extraction. In 2021 Computing in Cardiology, volume 48. September 2021; .

[10] Vollmer M. A Robust, Simple and Reliable Measure of Heart Rate Variability Using Relative RR Intervals. In 2015 Computing in Cardiology. 2015; 609–612.

[11] Laguna P, Moody GB, García J, Goldberger AL, Mark RG. Analysis of the ST-T Complex of the Electrocardiogram Using the Karhunen—Loeve Transform: Adaptive Monitoring and Alternans Detection. Medical Biological Engineering Computing 1999;37(2):175–189.

[12] Pilia N, Nagel C, Lenis G, Becker S, Dössel O, Loewe A. ECGdeli - An Open Source ECG Delineation Toolbox for Matlab. SoftwareX January 2021;13:100639.

[13] Schmidt M, Baumert M, Porta A, Malberg H, Zaunseder S. Two-Dimensional Warping for One-Dimensional Signals—Conceptual Framework and Application to ECG Processing. IEEE Transactions on Signal Processing 2014; 62(21):5577–5588.

[14] Schmidt M, Baumert M, Malberg H, Zaunseder S. T Wave Amplitude Correction of QT Interval Variability for Improved Repolarization Lability Measurement. Frontiers in Physiology 2016;7:216.

[15] Andreotti F, Behar J, Zaunseder S, Oster J, Clifford GD. An Open-Source Framework for Stress-Testing non-Invasive Foetal ECG Extraction Algorithms. Physiological Measurement 2016;37(5):627–648.

Address for correspondence:

Alexander Hammer
Institute of Biomedical Engineering, TU Dresden
Fetscherstr. 29, 01307 Dresden, Germany
alexander.hammer@tu-dresden.de