Atrial Fibrillation Recurrence Risk Prediction from 12-lead ECG Recorded Pre- and Post-Ablation Procedure

Eran Zvuloni[1], Sheina Gendelman[1], Sanghamitra Mohanty[2], Jason Lewen[3], Andrea Natale[2], Joachim A. Behar[1]

[1]Faculty of Biomedical Engineering, Technion-IIT, Haifa, Israel
[2]Texas Cardiac Arrhythmia Institute, St David's Medical Center, Austin, TX, USA
[3]BioSig Technologies Inc.

Abstract

Introduction: 12-lead electrocardiogram (ECG) is recorded during atrial fibrillation (AF) catheter ablation procedure (CAP). It is not easy to determine if CAP was successful without a long follow-up assessing for AF recurrence (AFR). Therefore, an AFR risk prediction algorithm could enable a better management of CAP patients. In this research, we extracted features from 12-lead ECG recorded before and after CAP and train an AFR risk prediction machine learning model. Methods: Pre- and post-CAP segments were extracted from 112 patients. The analysis included a signal quality criterion, heart rate variability and morphological biomarkers engineered from the 12-lead ECG (804 features overall). 43 out of the 112 patients (n) had AFR clinical endpoint available. These were utilized to assess the feasibility of AFR risk prediction, using either pre or post CAP features. A random forest classifier was trained within a nested cross validation framework. Results: 36 features were found statistically significant for distinguishing between the pre and post surgery states (n=112). For the classification, an area under the receiver operating characteristic (AUROC) curve was reported with $AUROC_{pre} = 0.64$ and $AUROC_{post} = 0.74$ (n=43). Discussion and conclusions: This preliminary analysis showed the feasibility of AFR risk prediction. Such a model could be used to improve CAP management.

1. Introduction

Analysis of 12-lead electrocardiogram (ECG) signals is essential for the diagnosis of heart conditions and patient monitoring following intervention. Atrial fibrillation (AF) is the most common arrhythmia with about 3% prevalence in adults and is associated with 5-fold increase in strokes [1]. Catheter ablation procedure (CAP) is a treatment for AF [2]. CAP is considered a long-term success if the patient does not experience AF recurrence (AFR) within a 3-year follow-up [2]. It is estimated that around 70% [1] of AF CAPs fail within this timeline. Therefore, there is a need to better understand who the patients are that may best benefit from CAP intervention. Moreover, it is important to closely monitor those who are at high risk of AFR following their CAP. This will support a better management of AF patients.

Engineered features extracted from ECG have been investigated as input for machine learning (ML) algorithms [3], [4], supporting complex analysis tasks. These have been providing necessary insights on patients' conditions for risk prediction or diagnosis [5], [6]. A number of research studies used echocardiography for AFR prediction. Yet, there was no agreement on a single echocardiography feature enabling AFR prediction post CAP [7]. One work by Fornengo et al [8] attempted to harness ML techniques and predicted AFR in cardioversion patients. Their results showed an area under the receiver operating characteristic (AUROC) curve of 0.66. In an ECG analysis work by Cheng et al [9], the group extracted the f-wave amplitude from three 10-second ECG leads prior to the CAP and analyzed them separately. Two leads were found significant as AFR predictors using the f-wave feature with best result of $Se = 0.75$ and $Sp = 0.73$.

This research aims to develop a risk prediction model for AFR using features engineered from 12-lead ECG sections taken either before (pre) or after (post) CAP. Accordingly, we attempt to address two fundamental questions: (i) can pre segments classification predict the CAP success rate for a given patient? (ii) can post segments classification answer if a patient treated with CAP is likely to develop AFR? We extracted a large and diverse set of features ($n_x = 804$) from the 12-lead ECG recorded. Then we analyzed the results statistically and within a supervised learning framework. Our results demonstrate the feasibility of predicting AFR on both conditions based on a small dataset of $n = 43$ patients.

2. Methods

Database and segment extraction: Patients treated with CAP for paroxysmal AF using the PURE EP system (BioSig Technologies Inc.) totaling 137 patients were included in this research. A continuous 12-lead ECG (standard lead system) was recorded throughout the surgery, i.e., starting when the patient received anesthetization before the CAP and until its fading after the treatment. Data was recorded at $f_s = 2,000 Hz$ with $0.03 \mu V$ amplitude quantization and had a median and interquartile duration of 1.6 and 0.98 hours, respectively. Fig. 1 describes each recording processing linked to the dataset experimental settings. Seven recordings were corrupted and thus excluded. Then, the first and last 5 minutes from each recording were analyzed. This was intended to reflect the patient state pre and post surgery. Next, representative segments were extracted from each of these 5 minutes. To select the segments with the highest quality, the signals were scanned with a moving signal quality index (bSQI) [5]. Since we later applied different statistical and ML tasks, the window had a different size in each task (treated as an hyperparameter), extracting segments with adequate durations: 10 and 60 seconds for the statistical analysis and for the classification tasks, respectively. The bSQI window had 5 seconds of overlapping between the scanned segments. bSQI was computed using two peak detectors, epltd from the WFDB library and jQRS as a reference. Computation was done with the custom PEBM toolbox [4]. Each ECG lead was computed separately and the mean bSQI over the 12 leads was then stored. For each recording, a single segment was extracted for the statistical analysis and the 5 segments with the highest bSQI were extracted for the classification tasks. A total of 18 recordings were discarded because of low quality (bSQI< 0.8). Among the remaining 112 recordings, 43 had AFR labels available (age= 63.5 ± 10.9; 61.4% males). The labels were based on a follow-up of 320.5 ± 120.34 days with a minimum of 154 days after the CAP.

Feature engineering: Two types of ECG features were engineered: heart rate variability (HRV) and morphological (MOR) biomarkers. For the HRV, features 1-20 from Chocron et-al [3] with an additional three features denoted "extended parabolic phase space mapping" [10] features were computed. MOR features were computed using the PEBM toolbox [4] for each ECG cycle and the median and standard deviation statistic were computed for each segment. Overall, we obtained 23 HRV and 2×22 MOR features for each lead, thus totaling 804 for the 12 leads. These were used in the statistical analysis. Additional demographics features (META) of age and sex were added to the classification tasks.

Statistical analysis: Statistical significance analysis between the ECG features extracted from the 10-second pre
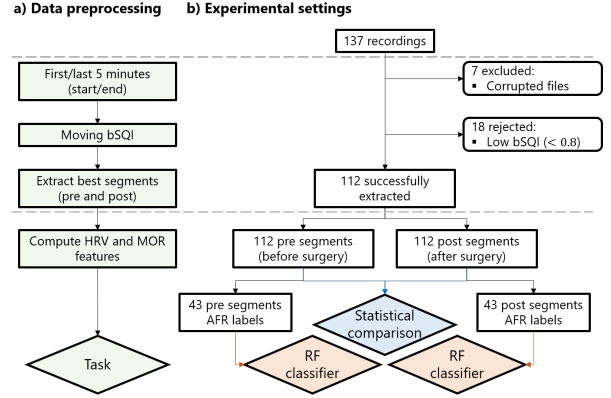

Figure 1. Dataset elaboration and experimental setting. a) Data preprocessing and feature engineering: the first and last 5 minutes were scanned to select a pre and post segment(s). Best segments were selected by computing signal quality index (bSQI) in a moving window. Heart rate variability (HRV) and morphological (MOR) features were extracted from the best quality segments. These were used as input for the final tasks (statistical analysis or classification). b) Experimental settings. 43 out of the 112 had atrial fibrillation recurrence (AFR) labels and were classified for AFR prediction using a random forest (RF) model.

and post highest bSQI segments was performed. A p-value was computed using a paired samples t-test applied on pre-CAP vs post-CAP segment features. Accordingly, each feature obtained a p-value to indicate its significance. Moreover, a mean fold-change (FC) was computed for each feature as $\mu_{FC}^{feature} = \frac{1}{n} \sum_{i=1}^{n} \frac{f_i^{post}}{f_i^{pre}}$, where f_i^{pre} and f_i^{post} are a feature computed from pre and post segments of patient i, respectively, and n is the number of patients. A volcano plot (Fig. 2) was used for display (bioinfokit [11]).

Machine learning: A random forest (RF) classifier was trained using the scikit-learn library [12] for the binary classification tasks. Given the low number of recordings having a clinical endpoint available ($n = 43$), a nested cross-validation approach was taken. The data was split in a K-fold manner into train, validation and test sets (with no overlap of the recordings), where $K_{test} = 8$ for the train-test (outer loop – train includes the validation set) and $K_{val} = 8$ for the train-validation (inner loop). A median imputer and a standardizing scaler fitted to the train folds in both loops were used. In addition, the outer loop included minimum redundancy maximum relevance (mRMR) algorithm for feature selection, implemented in MATLAB (Mathworks). In this way the same features were used for all the inner K validations, yet possibly varied between the K test sets. The number of selected features was optimized according to AUROC scores taken from the inner loop. Other hyperparameter optimizations

Table 1. Pre- and post-CAP segment statistical analysis.

Type	Significant feature found (as in Fig. 2)
HRV	AVNN, IALS, medHR, minRR, PACEv, PAS, PIP, PNN20, PSS, sq_map_linear, sq_map_intercept
MOR-median	$Jpoint$, PR_{int}, $PR2_{int}$, PR_{seg}, $Pwave_{int}$, QRS_{int}, QT_{int}, QT_{cB}, R_{dep}, RR_{int}, R_{wave}, TP_{seg}, $Twave_{int}$, $Twave$
MOR-std	PR_{int}, $PR2_{int}$, $Pwave_{int}$, QRS_{int}, QT_{int}, QT_{cF}, QT_{cH}, $Rwave$, ST_{seg}, TP_{seg}, $Twave_{int}$

were performed inside the inner loop using a Bayesian search (scikit-optimize) to tune the RF hyperparameters. The search was set to maximize the AUROC. Since we applied data augmentation, the AUROC was computed based on a majority vote over the different segments from a given patient. Moreover, the final AUROC was taken as the mean of the different outer loop 8-fold scores (Fig. 3). With this configuration three different models were trained: META, ECG (HRV+MOR), and META+ECG.

3. Results

We performed a statistical analysis between the features extracted from the pre and post segments and included the 804 features tested in a volcano plot (Fig. 2). The thresholds (gray lines) determine both statistical ($p^{feature}_{value} < \alpha = 0.05$) and FC ($|\mu^{feature}_{FC}| > 1$) significance. Accordingly, features crossing both thresholds (i.e., closer to the graph top corners) can be considered significant for distinguishing between the patients' pre- and post-CAP states. With this analysis, we found 36 features to be significant both statistically and FC-wise (Table 1).
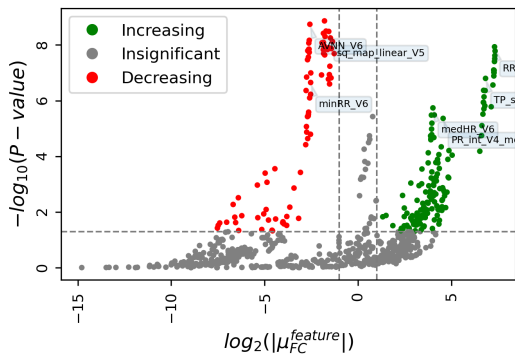

Figure 2. Volcano plot showing the engineered features. Features that were found significant (above gray line thresholds) are colored (green and red for an increasing and decreasing fold-change (FC), respectively). The labels show some of these by name and ECG lead.

The classification results are shown in Fig. 3. mRMR feature selection led to 5 features being selected. The

three models we trained (META, ECG and META+ECG) allowed us to observe the separated and joint effect of the different feature types. Using the META features alone, the RF classifier achieved an $AUROC_{META} = 0.5$. In the cases involving the extracted ECG features, the results were $AUROC_{pre-ECG} = 0.6$, $AUROC_{pre-META+ECG} = 0.64$, $AUROC_{post-ECG} = 0.67$ and $AUROC_{post-META+ECG} = 0.74$.
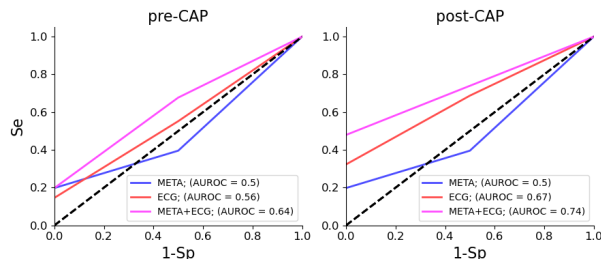

Figure 3. Test set receiver operating characteristic curve (ROC) for both RF experiments. Left: experiments obtained by analyzing the ECG acquired from the pre-catheter ablation procedure (CAP) segments. Right: the ECG was analyzed post-CAP.

4. Discussion

The statistical analysis implies the feasibility of distinguishing between the pre- and post-CAP patients. Importantly, we found significant features constructed from both HRV and MOR analysis (Table 1). This feature range demonstrates how conduction is affected by the surgery and may be quantified. Moreover, it emphasizes the importance of combining different feature engineering approaches (e.g., HRV and MOR) and utilizing multiple channels, as all contributed to the separation. Interestingly, the post segments were apparently affected by an increase in heart rate (e.g. using isoproterenol), applied to assess the heart activity post-treatment. This effect may be recognized in the statistical analysis, for example, the median heart rate (medHR) feature was found as a significant discriminator with higher value after surgery (Fig. 2).

The ROC curves in Fig. 3 show a moderate classifications performance for both AFR risk prediction experiments. The extracted ECG features benefit the classification task when compared to using META features alone. Specifically, META had an AUROC of 0.5 versus 0.64 and 0.74 for pre-CAP and post-CAP when using META and ECG features combined. These results match performance reported by others for the task of AFR risk prediction [8], [9] using echocardiography or a single ECG lead, although, these research experiments were only used on the pre-treatment 12-lead ECG measurements.

The main limitation of our study is the need to consider the anesthetization effect on patients before and after the CAP. This might have caused bias in the extracted features, which would not correctly reflect the patient state. Thus, ideally, it might be important to acquire ECG data before anesthetization and until enough time has passed post surgery to assume that the drug was washed out. The second main limitation is the low number of patients for which we had an AFR clinical endpoint (only 43), which intrinsically restricts the performance we were able to reach using ML approach.

5.   Conclusions

HRV and MOR features were extracted from 12-lead ECG recording segments of pre- and post-CAP for AFR treatment. With these we obtained statistically significant separation between patients' pre and post states, implying a heart electrical activity modification caused by the treatment. Moreover, these features were also used to classify between patients that did or did not experience AFR post-treatment (clinical endpoint). The classification showed a moderate AUROC performance for both pre-CAP and post-CAP analysis. Our results serve as a proof of concept and demonstrate how data taken from a 12-lead ECG can be used as a predictor for both treatment success (pre) and arrhythmia likelihood of recurrence (post). In future work, we intend to grow the dataset and to investigate the pre and post differences as features, as well as evaluate deep learning approaches [6] to reach higher performance; thus, allowing the clinical deployment of our model.

Acknowledgments

References

[1]   P. Kirchhof, S. Benussi, D. Kotecha, *et al.*, "2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS," *European Heart Journal*, vol. 37, no. 38, pp. 2893–2962, Oct. 2016.

[2]   I. Garvanski, I. Simova, L. Angelkov, and M. Matveev, "Predictors of Recurrence of AF in Patients After Radiofrequency Ablation," *European Cardiology Review*, vol. 14, no. 3, pp. 165–168, Dec. 2019.

[3]   A. Chocron, J. Oster, S. Biton, *et al.*, "Remote Atrial Fibrillation Burden Estimation Using Deep Recurrent Neural Network," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 8, pp. 2447–2455, Aug. 2021.

[4]   S. Gendelman, S. Biton, R. Derman, *et al.*, "PhysioZoo ECG: Digital electrocardiography biomarkers to assess cardiac conduction," in *2021 Computing in Cardiology (CinC)*, vol. 48, 2021, pp. 1–4.

[5]   J. Behar, J. Oster, Q. Li, and G. D. Clifford, "ECG signal quality during arrhythmia and its application to false alarm reduction," *IEEE transactions on biomedical engineering*, vol. 60, no. 6, pp. 1660–1666, 2013.

[6]   S. Biton, S. Gendelman, A. H. Ribeiro, *et al.*, "Atrial fibrillation risk prediction from the 12-lead electrocardiogram using digital biomarkers and deep representation learning," *European Heart Journal - Digital Health*, Aug. 2021.

[7]   A. Liżewska-Springer, A. Dabrowska-Kugacka, E. Lewicka, Ł. Drelich, T. Królak, and G. Raczak, "Echocardiographic predictors of atrial fibrillation recurrence after catheter ablation: A literature review," *Cardiology Journal*, vol. 27, no. 6, pp. 848–856, Dec. 2020.

[8]   C. Fornengo, M. Antolini, S. Frea, *et al.*, "Prediction of atrial fibrillation recurrence after cardioversion in patients with left-atrial dilation," *European Heart Journal – Cardiovascular Imaging*, vol. 16, no. 3, pp. 335–341, Mar. 2015.

[9]   Z. Cheng, H. Deng, K. Cheng, *et al.*, "The Amplitude of Fibrillatory Waves on Leads aVF and V 1 Predicting the Recurrence of Persistent Atrial Fibrillation Patients Who Underwent Catheter Ablation," *Annals of Noninvasive Electrocardiology*, vol. 18, no. 4, pp. 352–358, Jul. 2013.

[10]   S. Moharreri, S. Rezaei, N. J. Dabanloo, and S. Parvaneh, "Extended Parabolic Phase Space Mapping (EPPSM): Novel quadratic function for representation of Heart Rate Variability signal," *Computing in Cardiology*, pp. 417–420, 2014.

[11]   R. Bedre, *reneshbedre/bioinfokit: Bioinformatics data analysis and visualization toolkit*, Jan. 2021.

[12]   F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine learning in Python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

Address for correspondence:

Dr. Joachim Behar
jbehar@technion.ac.il