



# Heart Rate Variability and Clinical Features as Predictors of Atrial Fibrillation Recurrence After Catheter Ablation: A Pilot Study

Javier Saiz-Vivo<sup>1,2\*</sup>, Valentina D. A. Corino<sup>2</sup>, Robert Hatala<sup>3</sup>, Mirko de Melis<sup>1</sup> and Luca T. Mainardi<sup>2</sup>

<sup>1</sup> Medtronic Bakken Research Center B.V., Maastricht, Netherlands, <sup>2</sup> Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy, <sup>3</sup> Department of Arrhythmias and Cardiac Pacing, National Institute of Cardiovascular Diseases, Bratislava, Slovakia

## OPEN ACCESS

### Edited by:

Bastiaan J. Boukens,  
University of Amsterdam, Netherlands

### Reviewed by:

Merle Friederike Fenner,  
University of Copenhagen, Denmark  
Hidde Bleijendaal,  
Heart Center, Amsterdam UMC,  
Netherlands

### \*Correspondence:

Javier Saiz-Vivo  
javier.saizvivo@medtronic.com

### Specialty section:

This article was submitted to  
Cardiac Electrophysiology,  
a section of the journal  
Frontiers in Physiology

Received: 26 February 2021

Accepted: 12 April 2021

Published: 25 May 2021

### Citation:

Saiz-Vivo J, Corino VDA, Hatala R,  
de Melis M and Mainardi LT (2021)  
Heart Rate Variability and Clinical  
Features as Predictors of Atrial  
Fibrillation Recurrence After Catheter  
Ablation: A Pilot Study.  
Front. Physiol. 12:672896.  
doi: 10.3389/fphys.2021.672896

Single-procedure catheter ablation success rate is as low as 52% in atrial fibrillation (AF) patients. This study evaluated the feasibility of using clinical data and heart rate variability (HRV) features extracted from an implantable cardiac monitor (ICM) to predict recurrences in patients prior to undergoing catheter ablation for AF. HRV-derived features were extracted from the 500 beats preceding the AF onset and from the first 2 min of the last AF episode recorded by an ICM of 74 patients (67% male; 57 ± 12 years; 26% non-paroxysmal AF; 57% AF recurrence) before undergoing their first AF catheter ablation. Two types of classification algorithm were studied to predict AF recurrence: single classifiers including support vector machines, classification and regression trees, and K-nearest neighbor classifiers as well as ensemble classifiers. The sequential forward floating search algorithm was used to select the optimum feature set for each classification method. The optimum weighted voting method, which used an optimum combination of the single classifiers, was the best overall classifier (accuracy = 0.82, sensitivity = 0.76, and specificity = 0.87). Clinical and HRV features can be used to predict rhythm outcome using an ensemble classifier which would enable a more effective pre-ablation patient triage that could reduce the economic and personal burden of the procedure by increasing the success rate of first catheter ablation.

**Keywords:** atrial fibrillation, classification, ablation, implantable cardiac monitor (ICM), heart rate variability, recurrences, machine learning

## INTRODUCTION

Catheter ablation, specifically pulmonary vein isolation (PVI), has become a common treatment over the decades for atrial fibrillation (AF) patients, especially those highly symptomatic (Arbelo et al., 2014) or those where antiarrhythmic drug therapy has not been sufficient (or tolerated) for rhythm stabilization (Kirchhof et al., 2017). However, the long-term outcomes of catheter ablation in AF reported single-procedure success rates as low as 66.6% in paroxysmal AF (PAF) patients and 51.9% in non-paroxysmal AF (NPAF) patients (Ganesan et al., 2013). Several well-established scoring systems aimed at predicting rhythm outcome after catheter ablation, including

thromboembolic risk predictors like CHADS<sub>2</sub> or CHA<sub>2</sub>DS<sub>2</sub>-VASc, have shown modest prediction capabilities (Letsas et al., 2014). Other specific rhythm outcome predictors such as APPLE (Kornej et al., 2015), SUCCESS (Jud et al., 2019), and MB-LATER (Mujović et al., 2017) have achieved better results. However, most studies done so far have the drawback of relying on 24-h Holter monitoring to detect AF recurrences, which was shown to have a rather poor detection rate for subclinical AF of 5.5% (Ramkumar et al., 2018).

Implantable cardiac monitors (ICMs) offer the advantage of long-term monitoring and use highly sensitive AF detection algorithms, with detection rates of up to 96% (Hindricks et al., 2010). These devices continuously classify the heart rhythm of a patient by analyzing its cardiac cycle and, in addition, store the R-R intervals of the beats preceding the AF episode and the first beats of the AF episode. Hence, heart rate variability (HRV)-derived features can be extracted. HRV has been proven to be a predictor of chronic AF recurrence after electrical cardioversion (Lombardi et al., 2001; Akyürek et al., 2003), and extensive work has been done in describing the changes in HRV features before and after ablation (Kocovic et al., 1993; Hsieh et al., 1999; Kang et al., 2014; Seaborn et al., 2014).

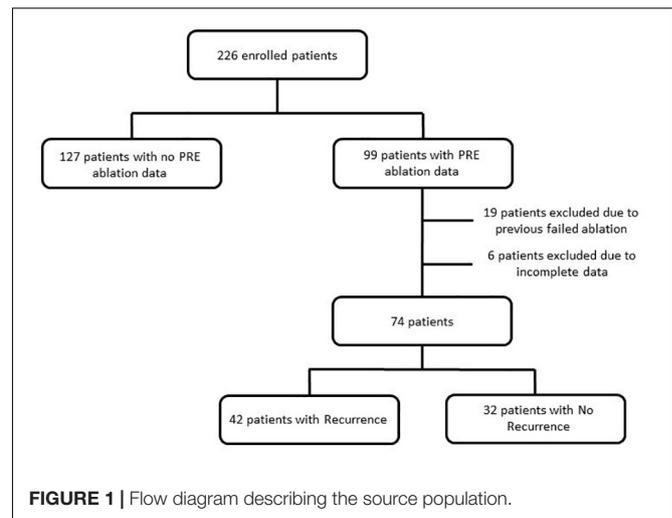
This article proposes the use of common HRV-derived features in conjunction with clinical data to predict recurrences within the first 12 months after catheter ablation in a continuously monitored patient population. To accomplish this, the article will evaluate several single classification methods including support vector machines (SVM), with linear (Christianini and Shawe-Taylor, 2013), polynomial (SVMp), and Gaussian (SVMg) kernels (Djerioui et al., 2019), classification and regression trees (CART), and K-nearest neighbor (KNN) algorithms. In addition, the capabilities of ensemble learning methods (Hosseini et al., 2020) in which a weighted combination of the single classifiers is used as the predictor of AF recurrence will be explored.

## MATERIALS AND METHODS

### Patient Population

This retrospective study included patients from two cohorts: the Reveal LINQ usability study (ClinicalTrials.gov identifier: NCT01965899), a multicenter single-arm clinical study, and clinical data derived from a single center with extensive experience in evaluating the long-term outcome of AF ablation by means of ICM (Bou Ezzeddine et al., 2015). The patients of both cohorts provided written informed consent, and the study protocols were reviewed and approved by the Human Research Ethics Committee of each participating institution.

Out of the 226 enrolled patients, only 99 had pre-ablation data, out of which 19 were excluded due to previously failed ablation and another six due to incomplete data such as no medical and/or ablation records. A schematic presentation describing the patient cohort can be found in **Figure 1**. The selected 74 patients (67% male;  $57 \pm 12$  years; 26% NPAF) were divided into two classes: those with AF recurrence ( $R = 42$  patients, 57% of the total) and those with no AF recurrence ( $NR = 32$ , 43%). AF recurrence



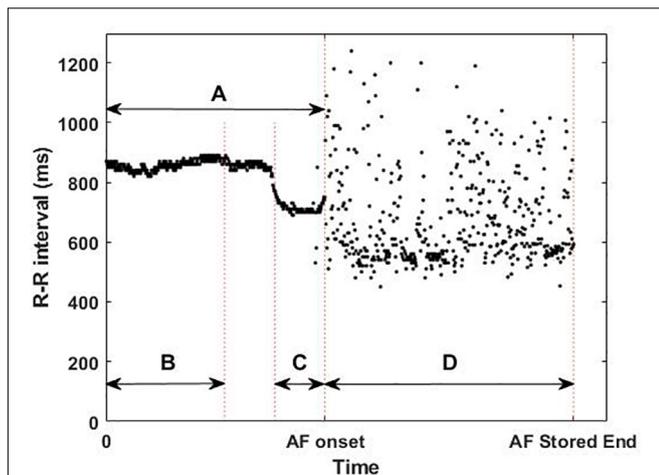
was defined as presence of an AF episode longer than 2 min as recorded by the ICM after a 3-month blanking period following catheter ablation. The blanking period of 3 months is used as suggested by the 2012 Consensus Statement of Catheter and Surgical Ablation of Atrial Fibrillation to report the efficacy of the ablation as early recurrences could be caused by post-ablation inflammation or short-term autonomic imbalance (Calkins et al., 2012). Both cohorts had more than 12 months follow-up for AF recurrence after catheter ablation.

The ICM was implanted  $5.9 \pm 3.8$  months before the ablation procedures, which were classified as PVI or PVI plus Extra Lesions. These long monitoring periods ensure the detection of the AF onset of both paroxysmal and non-paroxysmal patients. The devices used in the usability and the single-center studies were Reveal LINQ and Reveal XT (Medtronic Inc, Minneapolis, MN, United States), respectively, and were implanted within the fourth intercostal space (V2–V3 electrode orientation). Both devices sense and detect the rhythm with a sampling frequency of 256 Hz and then, to optimize memory slots, store the R-peak timestamps and the ECG of the first 2 min of the AF episodes. For the last AF episode recorded, the device also stores the timestamps of the beats preceding the AF onset (Flashback). For the analysis, the first beats of the last recorded AF episode ( $477 \pm 71$  beats) before the catheter ablation and its Flashback ( $483 \pm 33$  beats) were extracted (D and A in **Figure 2**). The last AF episodes occurred between 1 and 183 days before the ablation.

### Data Collection and Feature Extraction

From these two types of signals (Flashback and last AF episode) 4 different areas of interest (AoI) were defined: the whole Flashback (A), the first 300 beats of the Flashback (B), the last 100 beats of the Flashback (C), and the first beats of the AF episode (D) as shown in **Figure 2**.

Classical HRV-derived features describing the variability and irregularity of the R–R intervals (see **Table 1**) were computed from the different AoIs and then categorized into four groups: whole Flashback (FB, AoI: A), last 100 beats of Flashback (L100, AoI: C), and delta and first beats of AF episode (AF, AoI: D).



**FIGURE 2** | Example of available implantable cardiac monitor data: **(A)** whole flashback, **(B)** first 300 beats of the flashback, **(C)** last 100 beats of the flashback, and **(D)** first 2 min of the last atrial fibrillation (AF) episode recorded before catheter ablation, illustrating the short-term changes in the R–R intervals (ms) prior to the AF onset and the differences in R–R intervals between the flashback **(A)** and the AF episode **(D)**.

Delta was defined as the percentage difference between the first 300 beats of the Flashback (AoI: B) and the last 100 beats of the Flashback (AoI: C) and was computed as:

$$\text{Delta (feature)} = \frac{\text{First 300 (feature)} - \text{Last 100 (feature)}}{\text{First 300 (feature)}} * 100$$

with the aim of studying the changes occurring within the Flashback itself.

In addition to variability and irregularity features, clinical information such as the age, AF type (paroxysmal or non-paroxysmal), hypertension presence, and ablation type (PVI or PVI plus extra lesions) were included in the analysis. A total of 14 R–R interval variability and irregularity features were computed for the four areas of interest, except pNN50 and pNN20 for Delta as the values for the first 300 beats of these features were 0 for some patients, and the relative change could not be computed. A total of 58 features were considered per patient, including the four clinical features and the variability and irregularity features.

## Classification

Based on the extracted features, eight classification algorithms were evaluated to predict AF recurrence after a 3-month blanking period. As a first step, a test set was randomly selected, containing 22% of the data (eight patients with recurrence and eight with no recurrence), which was used to evaluate the performance of the classifiers on never-before-seen data. The remaining 78% of the observations were processed by the classification algorithm, using 2/3 of the data to train the classification model (the training set) and 1/3 to validate the trained model (validation set) as illustrated in **Figure 3**. The classification algorithm included the feature selection and the model training, out of which the validation performance metrics were computed. As a feature selection tool, the sequential forward floating search (SFFS) algorithm was used. This algorithm considers and analyzes subsets with different number of features by iteratively selecting the features that increase the overall accuracy of the model (Somol et al., 2006). Further information on the algorithm can be found in the **Supplementary Material**.

The trained model was evaluated on the test set, on which the performance metrics were computed.

The classification was performed using two types of classifiers: single classifiers and ensemble classifiers. The single classifiers were SVM, CART, and KNN; the ensemble classifiers were mean voting (MV), accuracy weighted voting (AWV), and optimum weighted voting (OWV).

For the ensemble classifiers, the five classifications of the single classifiers were weighed and combined to classify a given observation  $x$ :  $C(x)$ , which is the result of the scalar product between the weight vector ( $\mathbf{W}$ ) and the voting vector ( $\mathbf{V}(x)$ ):

$$C(x) = \begin{cases} 1 & \text{if } \mathbf{W} \cdot \mathbf{V}(x)^T > 0.5 \\ 0 & \text{if } \mathbf{W} \cdot \mathbf{V}(x)^T \leq 0.5 \end{cases}$$

where, the weight vector:

$$\mathbf{W} = [W_{\text{SVM}}, W_{\text{SVMp}}, W_{\text{SVMg}}, W_{\text{CART}}, W_{\text{KNN}}]$$

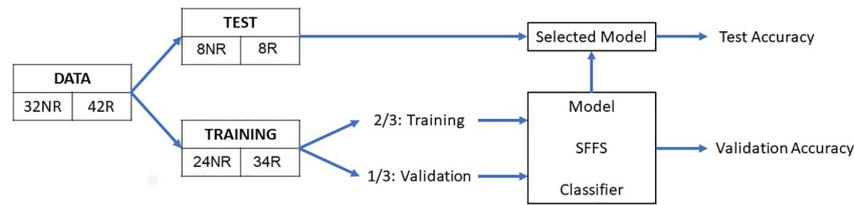
contains the weights for the single classifiers ( $W_i \in [0, 1]$ ,  $\sum W_i = 1$ ), and the voting vector:

$$\mathbf{V}(x) = [V_{\text{SVM}}(x), V_{\text{SVMp}}(x), V_{\text{SVMg}}(x), V_{\text{CART}}(x), V_{\text{KNN}}(x)]$$

is the binary classification of the observation  $x$  made by the single classifiers.

**TABLE 1** | Description of the heart rate variability features used in the analysis.

Features	Short descriptions	References
Mean	Mean of the R–R intervals (ms)	Camm et al., 1996
pNN50, pNN20	Percentage of the interval differences of successive R–R intervals greater than 50 and 20 ms (%)	Mietus et al., 2002
RMSSD	Mean square differences of successive R–R intervals (ms)	Camm et al., 1996
SDNN	Standard deviation of the R–R intervals (ms)	Camm et al., 1996
TINN	Triangular interpolation of interval histogram (ms)	Camm et al., 1996
TRI	Triangular index: integral of the density distribution	Camm et al., 1996
ApEn, SampEn	Approximate and sample entropy: estimators of the complexity of the R–R series	Delgado-Bonal and Marshak, 2019
SD1, SD2, SD1SD2ratio	Geometric descriptors of the Poincare plot	Tayel and AlSaba, 2015
DFA $\alpha_1$ , $\alpha_2$	Scaling exponents of short- and long-term fluctuations of the R–R intervals	Peng et al., 1995



**FIGURE 3** | Schematic representation of the overall method. NR, no recurrence; R, recurrence; SFFS, sequential forward floating search.

Each ensemble algorithm had a different weight configuration. The MV algorithm was defined as the average of all the classifications; thus, the single classifiers have the same weight:

$$W_{SVM} = W_{SVMp} = W_{SVMg} = W_{CART} = W_{KNN} = 0.2.$$

In the AWV method, the weights are proportional to the accuracy on the validation set of the single classifiers, thus being:

$$W_i = \frac{\text{Accuracy}_i}{\sum \text{Accuracy}_i} \text{ with } i \\ = \text{'SVM', 'SVMp', 'SVMg', 'CART', 'KNN'}$$

Finally, for the OWV method, all possible weight combinations (considering a step of 0.1 for each weight) were iteratively evaluated, and the set of weights that maximizes the overall accuracy of the validation set was selected.

Leave- $p$ -out cross-validation (where  $p$  is 1/3 of the data) was performed with 100 bootstrap repetitions, i.e., all the above-mentioned steps were repeated 100 times, randomizing the patient selection and allowing both patients with AF recurrence and those without to be part of the training and validation phases. Performance metrics such as accuracy, sensitivity, and specificity were then averaged over the repetitions, while F1-score was computed from the averaged confusion matrix, and were computed as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}; \text{ Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$\text{F1-score} = 2 \cdot \frac{\text{PPV} \cdot \text{Sensitivity}}{\text{PPV} + \text{Sensitivity}} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$$

where TP is true positive, TN is true negative, FP is false positive, FN is false negative, and PPV is positive predictive value. The feature extraction, selection, and classification were conducted using Matlab R2019b (The Mathworks Inc., Natick, MA, United States).

## Statistical Analysis

The statistical properties of the optimum set of features used by the classification method with the highest accuracy were

analyzed. Continuous data are presented as mean  $\pm$  SD if the null hypothesis  $H_0$  of the Shapiro–Wilk test ( $H_0$ : data is normally distributed) was not rejected, and these were compared with the unpaired Student's  $t$ -test. Otherwise, continuous data are presented as median (IQR), with IQR being the interquartile range, and compared using the Mann–Whitney  $U$  test. Conversely, categorical data is presented as absolute frequency (relative frequency in percentage) and compared with the Pearson chi-square method. A  $p$ -value  $< 0.05$  was considered for the rejection of the null hypothesis and set as the level of significance. All statistical analyses were conducted using SPSS, version 23 (SPSS Inc., Chicago, IL, Illinois).

## RESULTS

The available clinical baseline characteristics of the 74 patients are shown in **Table 2**. Even though there are no statistically significant differences between the clinical baseline characteristics used in the analysis of the patients with and without recurrences (age, paroxysmal AF, hypertension, and extra lesions), patients with AF recurrence were, on average, 3.6 years older than those without. There was also a higher proportion of paroxysmal AF patients and a lower proportion of arterial hypertension patients among those who did not have AF recurrences. Diabetes, coronary artery disease, and stroke were excluded from the analysis as these features were heavily underrepresented. The HRV-derived features for each group of interest are shown

**TABLE 2** | Clinical baseline characteristics of the patients that had no recurrence and those who did.

Feature	No recurrence (NR = 32, 43%)	Recurrence (R = 42, 57%)	$p$ -value
Age, years	55.47 $\pm$ 12.79	59.12 $\pm$ 11.55	0.20
Paroxysmal atrial fibrillation	26 (81.3%)	29 (69.1%)	0.18
Hypertension	19 (59.4%)	26 (61.9%)	0.83
Diabetes	1 (3.1%)	5 (11.9%)	<0.001
Coronary artery disease	0	5 (11.9%)	<0.001
Stroke	3 (9.4%)	2 (4.8%)	<0.001
Monitoring time pre-ablation, months	8.1 $\pm$ 3.7	4.3 $\pm$ 2.9	<0.001
Extra lesions	5 (15.6%)	10 (23.8%)	0.39

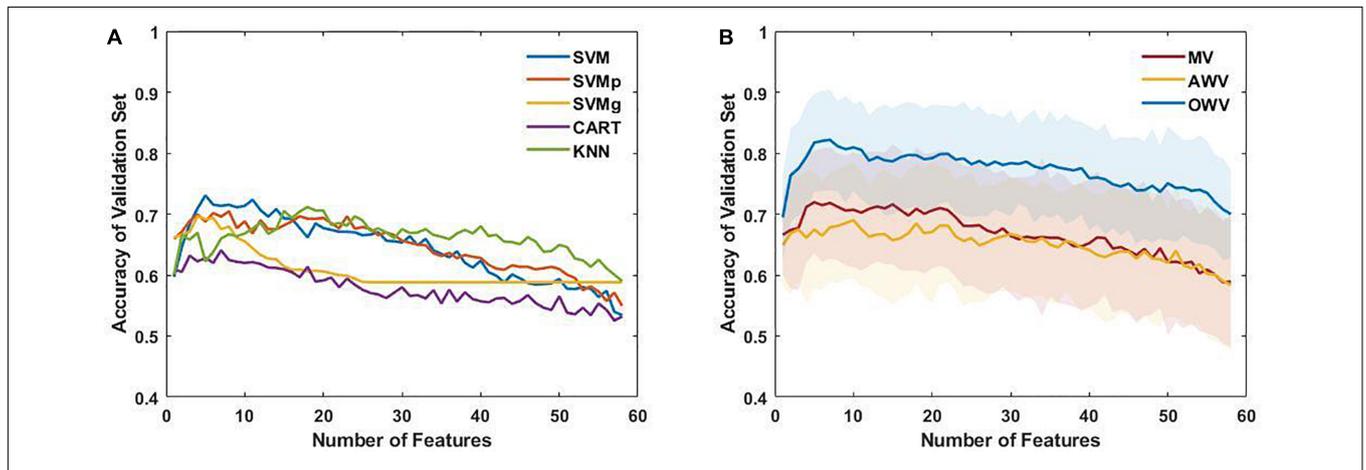
in **Supplementary Table 1** as mean  $\pm$  standard deviation for normally distributed data and as median (IQR) for non-normally distributed data. Only pNN20 for the whole Flashback, delta triangular index, and sample entropy in the AF episode have a statistically significant difference between patients with and without AF recurrences.

Firstly, the single classifiers were evaluated by computing the accuracy on the validation set for subsets with an increasing number of features as selected by the SFFS. **Figure 4A** shows the mean accuracies on the validation set of the single classifiers as a function of the number of selected features. It can be observed that all the classifiers reach a maximum of accuracy for the validation set with a subset containing less than 10 features except KNN (number of features = 18). For each single classifier, the feature subset that maximizes the validation accuracy is selected as the optimum feature set, which is used to evaluate the test

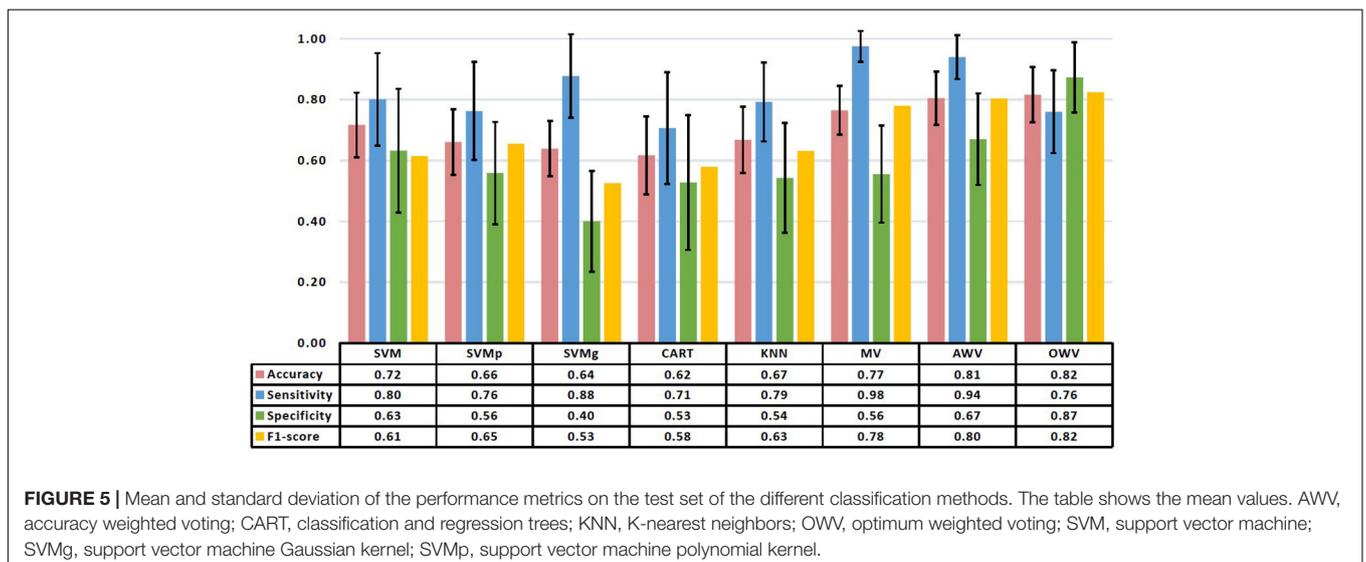
using the trained model. The performance evaluators for the test set were then computed in every iteration, and **Figure 5** depicts the mean and standard deviation. The F1-score, however, was computed from the averaged confusion matrix (also shown in **Figure 5**), as only the average score was of interest to compare different classifiers.

When working with never-seen data, SVM had the highest accuracy ( $0.72 \pm 0.11$ ) and specificity ( $0.63 \pm 0.20$ ), while SVMg had the highest sensitivity ( $0.88 \pm 0.14$ ). The highest F1-score ( $0.65$ ) was obtained by SVMp.

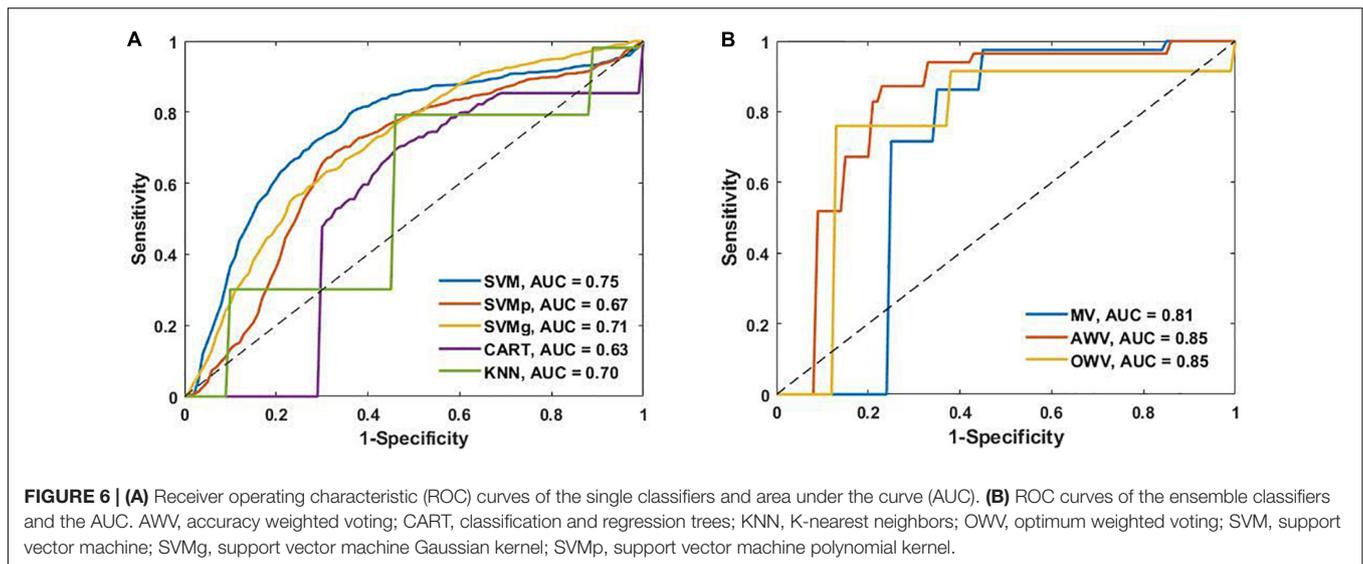
The single classifiers were then combined in an ensemble classifier in which a weighted combination of the single classification is used to compute the final classification. **Figure 4B** shows the mean (bold line) and the standard deviation (shaded area) of the accuracy of the validation set for the different ensemble classifiers. Similar to the single classifiers, the maximum



**FIGURE 4 | (A)** Mean accuracy of the single classifiers on the validation set, plotted as a function of the number of selected features. **(B)** Mean and standard deviation of the accuracy of the multiple classifiers (line and colored area, respectively) for each subset of features on the validation set. AWV, accuracy weighted voting; CART, classification and regression trees; KNN, K-nearest neighbors; OWV, optimum weighted voting; SVM, support vector machine; SVMg, support vector machine Gaussian kernel; SVMp, support vector machine polynomial kernel.



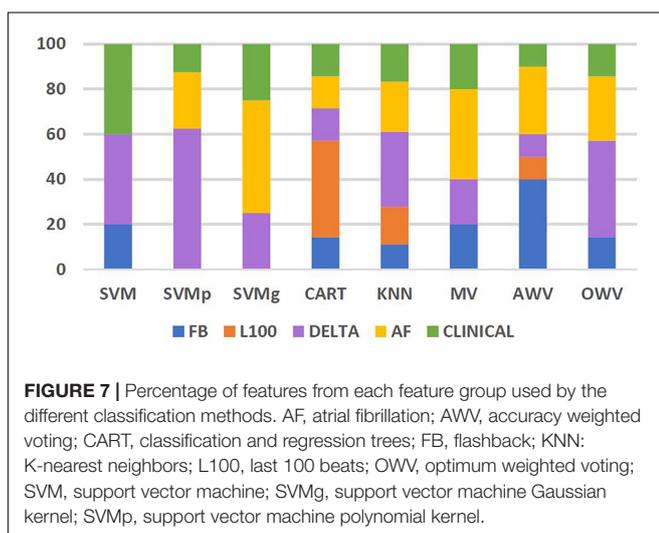
**FIGURE 5 |** Mean and standard deviation of the performance metrics on the test set of the different classification methods. The table shows the mean values. AWV, accuracy weighted voting; CART, classification and regression trees; KNN, K-nearest neighbors; OWV, optimum weighted voting; SVM, support vector machine; SVMg, support vector machine Gaussian kernel; SVMp, support vector machine polynomial kernel.



accuracy was reached with less than 10 features. The optimum feature set was determined as the subset that maximized the accuracy on the validation set, and the results for the ensemble classifiers' comparison are shown in **Figure 5**.

The best overall classifier with the highest F1-score (0.82) is the OWV method, in which the weights used to combine the different single classifiers are evaluated in each iteration. This method also has the highest accuracy ( $0.82 \pm 0.09$ ) and specificity ( $0.87 \pm 0.12$ ) on the test set, while MV has the highest sensitivity ( $0.98 \pm 0.05$ ).

This OWV method of classification used a set of seven features combining geometric delta features ("delta SD1SD2ratio") with complexity delta features ("delta ApEn"), statistical delta features ("delta RMSSD"), geometric AF features ("SD2 AF"), statistical AF features ("pNN20 AF"), statistical Flashback features ("pNN20"), and clinical features ("extra lesions").



To provide added information and insight on the performance of the different classification models, the receiver operating characteristic curves from the single and the ensemble classifiers are shown in **Figure 6** alongside their area under the curve (AUC) values. SVM had the highest AUC value (0.75) of the single classifiers, while AWW and OWV both obtained the highest overall AUC value (0.85).

The frequency of use of each of the feature groups, i.e., FB: Flashback, L100: Last 100, Delta, AF, and clinical by the different classification methods, was also analyzed. **Figure 7** shows the percentage of each feature group used by the different methods. Features from feature group Delta were the most frequently used by the different classifiers. On average, the classifiers used features from Delta group for 31% of their selected features, reaching 63% of the selected features for the SVMp method. However, it is worth noting that every classifier took at least "extra lesions" as one of the optimum features in their feature list, and classifiers such as SVM and SVMg had features from the clinical group comprising more than 25% of their features, with SVMg being the classifier with the highest percentage of use (40%).

## DISCUSSION

The main finding of this study is that a reduced set of HRV and clinical features extracted from an ICM can be used by an ensemble classifier to predict AF recurrence with a mean accuracy higher than 0.8 in patients that underwent single-procedure catheter ablation. If confirmed by future studies, these findings are potentially of significant clinical relevance for several reasons: first, catheter ablation of the AF substrate is a procedure with high economic and personal burden; secondly, due to the epidemic character of AF prevalence, these interventions cannot be offered (even in countries with developed healthcare systems) to all patients; and third, the selection of patients with higher probability of long-term elimination of AF has high priority. Until now, the various

clinical scoring systems based on phenotypic biomarkers used to this aim generate a rather poor prediction of limited clinical usefulness.

HRV has been extensively studied with respect to procedural outcome by analyzing the changes in HRV features before and after ablation (Kocovic et al., 1993; Hsieh et al., 1999; Kang et al., 2014; Seaborn et al., 2014). However, these studies mainly describe the effect of ablation on HRV and use non-continuous Holter monitoring of the patients. To the best of our knowledge, this is the first study that reports the combination of classical HRV and clinical features to predict AF recurrence in a continuously monitored ablation cohort using a variety of classification methods.

Although all of the different HRV and clinical features were initially introduced in the algorithm, SFFS iteratively selected the optimum set, and only one of the eight different methods used had a peak performance with more than 10 features and OWV only used seven: delta SD1SD2 ratio, delta ApEn, delta RMSSD, pNN20 AF, SD2 AF, and extra lesions. Out of the seven features found as optimum, only one was a clinical feature, while the others were classical HRV features extracted from three different feature groups: delta, Flashback, and AF. RMSSD and Poincare descriptor SD1SD2 ratio both describe the variability of the R-R intervals, and so when analyzing the correlation between the features, it was not surprising to observe that they had an  $r = 0.506$ ,  $p < 0.01$ , in the two-tailed Pearson correlation test. The rest of the features had a weak correlation between them.

The most used feature group appeared to be delta, with an average use of 31% in the different classifiers and a maximum of 63% in SVMp, which shows the importance of studying the onset of the AF contained within the Flashback.

Several studies have investigated different methods of AF recurrence prediction by analyzing the impact of different clinical scores. While thromboembolic risk predictors like CHADS<sub>2</sub> or CHA<sub>2</sub>DS<sub>2</sub>-VASc showed a relatively modest prediction (Letsas et al., 2014), other specific rhythm outcome predictors such as APPLE (Kornej et al., 2015), SUCCESS (Jud et al., 2019), and MB-LATER (Mujović et al., 2017) have been introduced as yielding better results. The APPLE score [one point for age > 65 years, persistent AF, impaired eGFR < 60 ml/min/1.73 m<sup>2</sup>, left atrial (LA) diameter ≥ 43 mm, and ejection fraction (EF) < 50%] was originally developed to predict AF recurrences after first ablation, with area under the receiver operating characteristic curve (AUC) of 0.634 (Kornej et al., 2015), but it has also been tested to predict recurrence in repeated ablations, with AUC of 0.617 (Kornej et al., 2017). Based on this score, the SUCCESS score was created by adding one point for each preciously performed ablation, and although it did demonstrate an improvement over APPLE (AUC 0.657 vs 0.620), the findings were not significant in the study (Jud et al., 2019). The MB-LATER score [male, bundle branch block, left atrium ≥ 47 mm, type of AF (paroxysmal, persistent, or long-standing persistent), and ER - AF = early recurrent AF] was associated with patients who will develop very late AF recurrence, i.e. recurrence documented more than 12 months after the ablation procedure (AUC 0.782) (Mujović et al., 2017),

and was also proven to predict late AF recurrence (AUC 0.62) (Potpara et al., 2019). However, this score has a drawback in that it uses early recurrence of AF as a feature, and it cannot be used as a baseline predictor. All these scoring systems have the disadvantage of relying on conventional Holter devices to detect AF recurrence in patients and the need of image-based parameters such as LA diameter or EF, while the proposed method uses easily obtainable clinical information and classical HRV features extracted from an ICM which continuously monitors the patient.

In the review of AF recurrence predictors developed by Balk et al. (2010) the relationship between success of ablation for AF and clinical features was systematically evaluated. The multivariable analyses showed that neither age, AF type, nor hypertension showed a significant association to ablation success. However, in the case of age, Balk et al. (2010) suggested that this result was due to the limitations of the existing literature rather than a true lack of association, as only relatively young patients were included in the analyses (40–70 years). The patient population analyzed in this study was relatively young ( $57.5 \pm 12.15$  years old), and this could be the reason why age was not included in the final feature set. Even though the multivariate analyses performed by Balk et al. (2010) failed to show a significant association between AF type and ablation success, the univariate analysis found that patients with NPAF had a 60% increased risk of AF recurrence compared to those with PAF, and it was hypothesized that it would be a good clinical indicator of the likelihood of AF recurrence. In this study, AF type was chosen by only two of the eight classification methods, and while one was AWV, which had the second highest F1 score, the other one was SVMp, which had one of the lowest scores. This could be explained by the underrepresentation in the patient population of non-paroxysmal patients (25.7%) which translated into 18.8% of those which did not have AF recurrence and 31.0% of those who did. Ablation type was the only feature used by every method: from these preliminary results, patients that had extra lesions also had a higher chance of having AF recurrence. Although there are some confounding factors to consider, i.e., the need of extra lesions may be due to a more advanced AF with a higher presence of fibrotic tissue in the atria; the extra scar tissue could be the foci of new re-entry circuits that could develop and sustain AF. Nonetheless, this feature was also heavily biased, as most of the patients underwent PVI ablations and only 15 patients (20.3%) also had extra lesions, so further work would have to confirm this.

This retrospective study was made using a limited patient population from two different cohorts with different clinical information, which limited the clinical features that could be used to those which were collected in both studies. Furthermore, the features that were included, such as AF type and ablation type, were heavily biased, as persistent patients (26% of the total number of patients) and ablation strategies with extra lesions (20%) were underrepresented. However, the main part of this study was focused on HRV features, and even though the use of the clinical features increased the accuracy and future work should be done to understand their impact, the presented

classification method and the results are still clinically relevant. The data were extracted from the Reveal LINQ ICM which automatically detected AF episodes longer than 2 min and, due to memory restrictions, stored the R–R intervals of the episodes detected and their Flashbacks. Therefore, episodes longer than 30 s, which are defined as AF episodes by the guidelines, but shorter than 2 min were undetected by the ICM. The lack of stored ECG signals also limited the number of features that could be extracted and restricted the analysis to HRV-derived features that could be extracted from the R–R intervals. Despite these drawbacks, the advantage of having continuous monitoring of the patients before and after the ablation greatly outweighs the disadvantages of possible information loss due to device resolution or memory restrictions. Although the study shows promising results and serves as proof of the feasibility of the method described, being a pilot study, the results should be validated on an external database.

Recurrence of AF after ablation can be predicted with varying degrees of accuracy using simple classification methods and an iteratively selected feature set of easily obtainable HRV and clinical features. The best approach is an optimally weighted combination of single classifiers which uses HRV (Poincare descriptors SD1SD2 ratio, pNN20 approximate entropy, RMSSD, and triangular index) and clinical (extra PVI lesions) features. This could be a first step into a more effective pre-ablation patient triage that could reduce the economic and personal burden of the procedure by increasing the success rate of the first catheter ablation to achieve long-term AF termination.

## DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions. Dataset owned by Medtronic and by the

Department of Arrhythmias and Cardiac Pacing, National Institute of Cardiovascular Diseases of Slovakia. Requests to access these datasets should be directed to JS-V and RH.

## AUTHOR CONTRIBUTIONS

JS-V contributed to the design of the study, wrote and implemented the code, analyzed the results, created the figures and tables, and wrote the first draft of the manuscript. RH designed the clinical study and obtained and provided clinical data. MM provided clinical data, contributed to the design of the study, analyzed the results, and supervised the project. VC wrote code, contributed to the design of the study, analyzed the results, and supervised the project. LM contributed to the design of the study, analyzed the results, and supervised the project. All authors contributed to manuscript revision and read and approved the submitted version.

## FUNDING

This work was supported by the MY-ATRIA consortium. The MY-ATRIA project has received funding from the European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant agreement no. 766082.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphys.2021.672896/full#supplementary-material>

## REFERENCES

- Akyürek, Ö, Diker, E., Güldal, M., and Oral, D. (2003). Predictive value of heart rate variability for the recurrence of chronic atrial fibrillation after electrical cardioversion. *Clin. Cardiol.* 26, 196–200. doi: 10.1002/clc.4960260411
- Arbelo, E., Brugada, J., Hindricks, G., Maggioni, A. P., Tavazzi, L., Vardas, P., et al. (2014). The atrial fibrillation ablation pilot study: an European survey on methodology and results of catheter ablation for atrial fibrillation conducted by the European heart rhythm association. *Eur. Heart J.* 35, 1466–1478. doi: 10.1093/eurheartj/ehu001
- Balk, E. M., Garlitski, A. C., Alsheikh-Ali, A. A., Terasawa, T., Chung, M., and Ip, S. (2010). Predictors of atrial fibrillation recurrence after radiofrequency catheter ablation: a systematic review. *J. Cardiovasc. Electrophysiol.* 21, 1208–1216. doi: 10.1111/j.1540-8167.2010.01798.x
- Bou Ezzeddine, H., Vachulova, A., Svetlosak, M., Urban, L., Hlivak, P., Margitfalvi, P., et al. (2015). Occurrence of symptoms after catheter ablation of atrial fibrillation. *Bratisl. Med. J.* 116, 461–464. doi: 10.4149/BLL\_2015\_086
- Calkins, H., Kuck, K. H., Cappato, R., Brugada, J., Camm, A. J., Chen, S. A., et al. (2012). 2012 HRS/EHRA/ECAS expert consensus statement on catheter and surgical ablation of atrial fibrillation: recommendations for patient selection, procedural techniques, patient management and follow-up, definitions, endpoints, and research trial design. *Heart Rhythm* 9, 632–696. doi: 10.1016/j.hrthm.2011.12.016
- Camm, A. J., Malik, M., Bigger, J. T., Breithardt, G., Cerutti, S., Cohen, R. J., et al. (1996). Heart rate variability: standards of measurement, physiological. *Eur. Heart J.* 17, 354–381. doi: 10.1161/01.CIR.93.5.1043
- Christianini, N., and Shawe-Taylor, J. (2013). *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge: Cambridge University Press.
- Delgado-Bonal, A., and Marshak, A. (2019). Approximate entropy and sample entropy: a comprehensive tutorial. *Entropy* 21:541. doi: 10.3390/e21060541
- Djeriou, M., Brik, Y., Ladjal, M., and Bilal, A. (2019). Neighborhood component analysis and support vector machines for heart disease prediction. *Ingenierie Des Syst. d'Inform.* 24, 591–595. doi: 10.18280/isi.240605
- Ganesan, A. N., Shipp, N. J., Brooks, A. G., Kuklik, P., Lau, D. H., and Lim, H. S. (2013). Long-term outcomes of catheter ablation of atrial fibrillation: a systematic review and meta-analysis. *J. Am. Heart Assoc.* 2:603. doi: 10.1161/JAHA.112.004549
- Hindricks, G., Pokushalov, E., Urban, L., Taborsky, M., Kuck, K., and Lebedev, D. (2010). Performance of a new leadless implantable cardiac monitor in detecting and quantifying atrial fibrillation results of the XPECT Trial. *Circ. Arrhythm. Electrophysiol.* 3, 141–147. doi: 10.1161/CIRCEP.109.877852
- Hossini, M.-P., Hosseini, A., and Ahi, K. (2020). A Review on machine learning for EEG Signal processing in bioengineering. *IEEE Rev. Biomed. Eng.* 14, 204–218. doi: 10.1109/rbme.2020.2969915
- Hsieh, M. H., Chiou, C. W., Wen, Z. C., Wu, C. H., Tai, C. T., Tsai, C. F., et al. (1999). Alterations of heart rate variability after radiofrequency catheter

- ablation of focal atrial fibrillation originating from pulmonary veins. *Circulation* 100, 2237–2243. doi: 10.1161/01.CIR.100.22.2237
- Jud, F. N., Obeid, S., Duru, F., and Haegeli, L. M. (2019). A novel score in the prediction of rhythm outcome after ablation of atrial fibrillation: the SUCCESS score. *Anatol. J. Cardiol.* 21, 142–149. doi: 10.14744/AnatolJCardiol.2018.76570
- Kang, K. W., Kim, T. H., Park, J., Uhm, J. S., Joung, B., Hwang, C., et al. (2014). Long-term changes in heart rate variability after radiofrequency catheter ablation for atrial fibrillation: 1-year follow-up study with irrigation tip catheter. *J. Cardiovasc. Electrophysiol.* 25, 693–700. doi: 10.1111/jce.12398
- Kirchhof, P., Benussi, S., Kotecha, D., Ahlsson, A., Atar, D., Casadei, B., et al. (2017). 2016 ESC guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *Eur. Heart J.* 37, 2893–2962. doi: 10.1093/eurheartj/ehw210
- Kocovic, D. Z., Harada, T., Shea, J. B., Soroff, D., and Friedman, P. L. (1993). Alterations of heart rate and of heart rate variability after radiofrequency catheter ablation of supraventricular tachycardia: delineation of parasympathetic pathways in the human heart. *Circulation* 88, 1671–1681. doi: 10.1161/01.CIR.88.4.1671
- Kornej, J., Hindricks, G., Arya, A., Sommer, P., Husser, D., and Bollmann, A. (2017). The APPLE score—a novel score for the prediction of rhythm outcomes after repeat catheter ablation of atrial fibrillation. *PLoS One* 12:e0169933. doi: 10.1371/journal.pone.0169933
- Kornej, J., Hindricks, G., Shoemaker, M. B., Husser, D., Arya, A., Sommer, P., et al. (2015). The APPLE score: a novel and simple score for the prediction of rhythm outcomes after catheter ablation of atrial fibrillation. *Clin. Res. Cardiol.* 104, 871–876. doi: 10.1007/s00392-015-0856-x
- Letsas, K. P., Efremidis, M., Giannopoulos, G., Deftereos, S., Lioni, L., Korantzopoulos, P., et al. (2014). CHADS2 and CHA2DS2-VASc scores as predictors of left atrial ablation outcomes for paroxysmal atrial fibrillation. *Europace* 16, 202–207. doi: 10.1093/europace/eut210
- Lombardi, F., Colombo, A., Basilio, B., Ravaglia, R., Garbin, M., Vergani, D., et al. (2001). Heart rate variability and early recurrence of atrial fibrillation after electrical cardioversion. *J. Am. Coll. Cardiol.* 37, 157–162. doi: 10.1016/S0735-1097(00)01039-1
- Mietus, J. E., Peng, C. K., Henry, I., Goldsmith, R. L., and Goldberger, A. L. (2002). The PNNx files: re-examining a widely used heart rate variability measure. *Heart* 88, 378–380. doi: 10.1136/heart.88.4.378
- Mujović, N., Marinković, M., Marković, N., Shantsila, A., Lip, G. Y. H., and Potpara, T. S. (2017). Prediction of very late arrhythmia recurrence after radiofrequency catheter ablation of atrial fibrillation: the MB-LATER clinical score. *Sci. Rep.* 7:40828. doi: 10.1038/srep40828
- Peng, C. K., Havlin, S., Stanley, H. E., and Goldberger, A. L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos* 5, 82–87. doi: 10.1063/1.166141
- Potpara, T. S., Mujovic, N., Sivasambu, B., Shantsila, A., Marinkovic, M., Calkins, H., et al. (2019). Validation of the MB-LATER score for prediction of late recurrence after catheter-ablation of atrial fibrillation. *Int. J. Cardiol.* 276, 130–135. doi: 10.1016/j.ijcard.2018.08.018
- Ramkumar, S., Nerlekar, N., Souza, D. D., Pol, D. J., Kalman, J. M., and Marwick, T. H. (2018). Atrial fibrillation detection using single lead portable electrocardiographic monitoring?: a systematic review and meta-analysis. *BMJ Open* 8:e024178. doi: 10.1136/bmjopen-2018-024178
- Seaborn, G. E. J., Todd, K., Michael, K. A., Baranchuk, A., Abdollah, H., Simpson, C. S., et al. (2014). Heart rate variability and procedural outcome in catheter ablation for atrial fibrillation. *Ann. Noninvasive Electrocardiol.* 19, 23–33. doi: 10.1111/anec.12098
- Somol, P., Novovičová, J., and Pudil, P. (2006). “Flexible-Hybrid Sequential Floating Search in Statistical Feature Selection,” in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) LNCS*, Vol. 4109, eds D. Y. Yeung, J. T. Kwok, A. Fred, F. Roli, and D. de Ridder, (Berlin: Springer), 632–639. doi: 10.1007/11815921\_69
- Tayel, M., and AlSaba, E. (2015). Poincaré plot for heart rate variability. *Int. J. Med. Health Biomed. Bioeng. Pharm. Eng.* 9, 708–711.

**Conflict of Interest:** MM is a Medtronic employee and JS-V is affiliated to Medtronic and supported by an EU grant.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Saiz-Vivo, Corino, Hatala, de Melis and Mainardi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.