Comparing Predictive Models for AF Recurrence Post-Catheter Ablation: CHA2DS2-VASc Score vs. HRV-Derived Features from Implantable Cardiac Monitors

Javier Saiz-Vivo[1], Mirko de Melis[1], Yong K Cho[2]

[1] Medtronic: Bakken Research Center, Maastricht, The Netherlands
[2] Medtronic: Mounds View, Minneapolis, MN, USA

Abstract

This study aims to evaluate the performance in predicting rhythm outcome post-catheter ablation of the CHA_2DS_2-VASc score against a refined set of clinical and Heart Rate Variability (HRV) features obtained from cardiovascular implantable electronic devices (CIEDs) in continuously monitored patients. 239 AF patients (age 67 ± 9 years, 40% female) from the Optum® de-identified Electronic Health Record data set (Optum® EHR) (2007-2020) were selected and linked with the Medtronic CareLink database of CIEDs. We compared the AF recurrence predictive performance of the CHA_2DS_2-VASc score with a Support Vector Machine (SVM) classifier. Feature selection among the clinical and HRV-derived features was performed using the sequential forward floating search algorithm. Among the patients, 182 (76%) experienced AF recurrence, defined as detected of an AF episode outside the 3-month blanking period. We utilized a CHA_2DS_2-VASc cutoff of 2 to predict AF recurrence. The SVM algorithm (F1-score = 0.58, AUC = 0.75), employing nine clinical and HRV-derived variables, outperformed the CHA_2DS_2-VASc score (F1-score = 0.47, AUC = 0.64). The optimized selection of clinical and HRV-derived features demonstrated superior predictive capability for rhythm outcome compared to thromboembolic risk predictors.

1. Introduction

As the population ages, atrial fibrillation (AF) is increasingly prevalent, affecting millions worldwide. Its incidence rises sharply from 0.7% in people aged 55-59 to nearly 18% in those 85 and older [1], and its management and the improvement of long-term outcomes present significant challenges [2]. Catheter ablation, particularly pulmonary vein isolation (PVI), has become a key treatment strategy for AF. It is now a class I recommendation for patients with symptomatic AF who have not responded to antiarrhythmic drugs (AADs), or in cases where AADs are contraindicated, poorly tolerated, or not preferred. PVI is also a first-line therapy for selected patients, typically younger individuals with few comorbidities [3].

Despite its advantages over AADs, AF recurrence recurs in about one-third of patients after catheter ablation, leading to higher morbidity, more hospitalizations, and increased healthcare costs [4]. Identifying patients at high risk of recurrence is crucial for personalizing treatment and improving outcomes.

Previous efforts to predict AF recurrence after ablation have focused on widely used thromboembolic risk predictors like the CHA_2DS_2-VASc score [5], or rhythm outcome predictors such as APPLE [6, 7], SUCCESS [8], and MB-LATER [9], which combine clinical parameters but offer only modest predictive power. Additionally, these studies often relied on suboptimal monitoring, like 24-hour Holter monitoring, to detect AF recurrences.

To address this, implantable cardiac monitors (ICMs) have become valuable tools for continuous cardiac rhythm monitoring, offering detailed insights into arrhythmia burden and recurrence patterns. ICMs continuously analyze heart rhythm, storing R–R intervals of beats before (hereinafter the Flashback) and during the first 2 minutes of an AF episode, allowing the extraction of heart rate variability (HRV)-derived features. The rich datasets from ICMs present a unique opportunity to develop predictive models for AF recurrence after ablation.

The aim of this study is to explore the AF recurrence prediction capabilities of the CHA_2DS_2-VASc score in a continuously monitored patient cohort and compare its performance to a support vector machine (SVM) classifier fed with an optimally selected set of heart rate variability (HRV)-derived and clinical features. By doing so, this research seeks to offer a more effective approach to patient triage and personalized treatment strategies.

2. Materials and Methods

2.1. Patient Population

The cohort used in this study was retrieved from the Optum® de-identified Electronic Health Record data set (Optum® EHR) (2007-2020), which contained the clinical interactions such as ablation procedures of patients across

the US, and were linked with Medtronic's CareLink database of CIEDs, providing the continuous monitoring data of the patients. Out of the 641 patients initially identified in Optum® EHR, 239 (age 67 ± 9 years, 40% female) met the inclusion criteria of the study as shown in Figure 1.

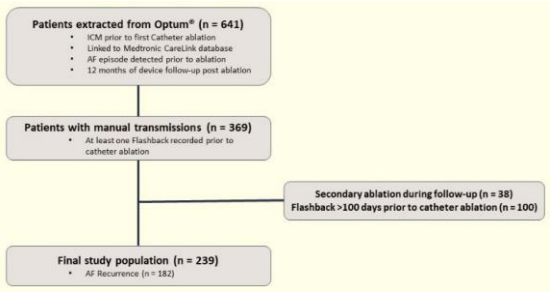

Figure 1 Flow diagram describing source population.

The AF recurrence incidence for this cohort was 76% and was defined as an ICM-detected AF episode longer than 2 minutes after a 3-month blanking period following catheter ablation. Early recurrences (those within the 3-months window) could be caused by post-ablation inflammation or short-term autonomic imbalance and were not considered AF recurrences [10].

The patients were monitored with Reveal LINQ and LINQ-II (Medtronic Inc, Minneapolis, MN, United States). These devices evaluate the rhythm every 2 minutes to detect arrhythmias, and once an AF episode is detected, the device stores the R-peak timestamps (marked as 'AF' in red in Figure 2) and the ECG of the first 2 minutes of the episode. In addition, the device stores the Flashback preceding the onset of the episode (marked as 'Pre-AF' in blue in Figure 2).

2.2. Feature extraction

The CHA_2DS_2-VASc score was determined by assigning one point for each factor such as congestive heart failure, hypertension, age 65 to 74, diabetes mellitus, vascular disease and female gender, and 2 points for age 75 years or older and history of transient ischemic attach and/or stroke [5].

Classical HRV-derived features describing the variability and irregularity of the R-R intervals including the mean R-R interval, pNN50 and pNN20, the root means square of successive differences (RMSSD), the standard deviation of R-R intervals (SDNN), the approximate (ApEn) and sample (SampEn) entropy [11], and the geometric descriptors of the Poincare plot (SD1, SD2, and SD1SD2ratio) [12], were computed for both the Flashback and the first 2 minutes of the AF episode.

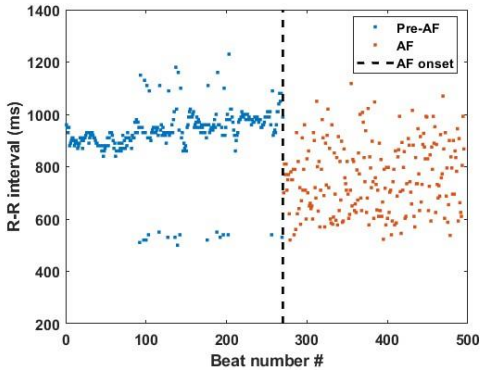

Figure 2 Example of implantable cardiac monitor data. In blue, pre-AF beats (Flashback), in red, first 2 minutes of atrial fibrillation (AF) episode, and in black, line marking AF onset.

In addition, clinical features such as age, gender, AF type (paroxysmal or non-paroxysmal), hypertension, diabetes mellitus, previous stroke, presence of vascular disease, coronary artery disease (CAD), previous myocardial infarction (MI) including ST-segment elevation and non-ST-segment elevation, heart failure, medication treatment (oral anticoagulation, beta blockers), and ablation type (PVI or PVI plus extra lesions) were also included in the analysis.

2.3. Classification

As a first classification attempt, the study population was categorized into two risk groups based on their CHA_2DS_2-VASc scores: low risk, score 0-1, and high risk, score ≥ 2 [3].

Then, using the extracted features, an SVM algorithm was evaluated to predict AF recurrence. The aim of SVMs are to devise a computationally efficient way of finding separating hyperplanes in an N-dimensional space that will minimize the generalization error, those which have the largest distance to the nearest training-data point of any class (so-called functional margin) [13]. Support vectors are defined as data points that are closer to a given hyperplane.

The data was split into training and test set (80-20). The test set in not used during the training phase and it's only introduced to evaluate the generalizability of the algorithm with never-before-seen data. The F1-score, the harmonic mean between precision and recall, in the metric the algorithm aimed to maximize during its training. The algorithm was trained and tested using a 100 bootstrap repetition scheme.

During the training phase, the data was further divided intro training and validation set using leave-p-out cross-validation (where p is 1/3 in this application). In addition, as the number of features was relatively high, the

sequential forward floating search (SFFS) algorithm was included as a feature selection tool. This feature selection algorithm considers and analyses subsets with different number of features by iteratively including and excluding features to increase the performance of the model [14]. As mentioned, once the optimum feature set was selected and the model trained, the performance metrics were computed from the test set.

3. Results

The demographics and baseline characteristics of the 239 patients are shown in Table *1*.

Table 1 Clinical baseline characteristics of the patients

	Total (n = 239)	No AF rec (n = 57)	AF rec (n = 182)	p-value
Female, n (%)	95 (40%)	15 (26%)	80 (44%)	0.020
Age, years, mean ± SD	67.4 ± 9.5	63.1 ± 9.8	68.7 ± 9.0	< 0.001
Paroxysmal AF, n (%)	215 (90%)	57 (100%)	158 (87%)	0.002
Coronary Artery Disease, n (%)	77 (32%)	12 (21%)	65 (36%)	0.050
Diabetes, n (%)	57 (24%)	10 (18%)	47 (26%)	0.218
Heart Failure, n (%)	41 (17%)	N<5 (<9%)	37 (20%)	0.025
Hyperlipidaemia, n (%)	145 (61%)	28 (49%)	117 (64%)	0.044
Hypertension, n (%)	182 (76%)	36 (63%)	146 (80%)	0.012
Myocardial Infarction, n (%)	23 (9%)	N<5 (<9%)	20 (11%)	0.302
Stroke, n (%)	35 (15%)	8 (14%)	27 (15%)	1.000
Vascular Disease, n (%)	30 13%)	5 (9%)	25 (14%)	0.370
CHA_2DS_2-VASc ≥ 2, n (%)	176 (74%)	31 (54%)	145 (80%)	< 0.001

The population was 40% female, relatively young with 79% under 75 years old and predominantly paroxysmal AF patients (90%). 176 patients had a CHA_2DS_2-VASc ≥ 2 and there was a statistically significant difference in the proportion of patients with a high score between recurrence and non-recurrence groups (Recurrence 80% vs No Recurrence 54%; p-value < 0.001).

The performance evaluators for both CHA_2DS_2-VASc and SVM on the test set were computed for every iteration and summarized in Figure *3* (left) where the bars show the average and standard deviation of the metrics while the table underneath shows only the average for clarity reasons.

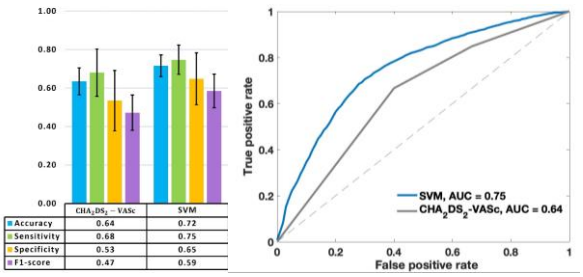

Figure 3 Mean and standard deviation of the performance metrics on the test set of the different classification methods (left), and the Receiver operator characteristic (ROC) curves with their respective area under the curve (AUC) (right).

In addition, the Receiver operator characteristic (ROC) curves of both predictor models with their respective AUC values are represented in Figure *3* (right). The SVM algorithm (F1-score = 0.58, AUC = 0.75), employing nine clinical and HRV-derived variables, outperformed the CHA_2DS_2-VASc score (F1-score = 0.47, AUC = 0.64). The features used by the SVM were age, heart failure, myocardial infarction, AF type, pNN20 extracted from the beats preceding the AF episode, and pNN50, approximate entropy and Poincare's SD1SD2 ratio extracted from the AF episode.

4. Discussion

This study explores the performance of support vector machines using a reduced set of HRV-derived and clinical features extracted from ICMs in predicting AF recurrence and shows its superiority to using thromboembolic risk predictors like CHA_2DS_2-VASc. Although statistically significant differences were observed between the proportion of patients with high CHA_2DS_2-VASc scores between recurrence and non-recurrence groups (80% vs 54%; p-value < 0.001), using CHA_2DS_2-VASc as a dichotomous variable (with a cut-off of 2) to predict AF recurrence yielded an AUC of 0.64, confirming its modest predictive value found in literature [5]. Other rhythm outcome predictors such as APPLE [6, 7], SUCCESS [8], and MB-LATER [9], offer better performances while having the disadvantage of needing image-based parameters such as LA diameter or EF, and in the case of MB-LATER, early recurrence of AF, preventing their use as a baseline predictor.

The limited predictive ability of scoring systems has motivated the development of machine-learning models for recurrence prediction. A recent review of 40 studies proposing 54 models, using clinical and/or radiomic features, showed moderate to high performance [15]. However, many models had high bias, lacked independent validation, and had challenges in explaining their predictions. Additionally, most studies (93%) used 12-lead

ECGs or 24-hour Holter monitoring for AF recurrence detection, which are less sensitive than implantable cardiac monitors [16].

This algorithm, a simplified light-weight version of the algorithm previously described in Saiz-Vivo et al. [17], improved the performance capabilities of scoring systems and by using ICMs, ensured a more sensitive AF recurrence detection. The selected features used by the algorithm could be classified into clinical features, and features extracted from the beat trend and the AF episode. The clinical features selected included age, heart failure, myocardial infarction, and AF type. These features were combined with statistical features (pNN20) extracted from the beat trend, and pNN50, complexity (approximate entropy) and geometric (Poincare's SD1 and SD1SD2ratio) features extracted from the AF episode.

Some limitations of the study include a relatively small number of patients, the fact that some of the variables used such as type of AF were heavily biased (only 10% were non-paroxysmal AF patients), and the 2-minute AF detection threshold the ICM has, by which episodes longer than 30s (as defined by the guidelines), but shorter than 2 minutes, were undetected by the ICM. Despite these drawbacks, the advantage of having continuously monitored patients before and after the ablation greatly outweighs the disadvantages of device resolution.

5. Conclusion

Predicting AF recurrence after catheter ablation using simple machine learning models and easily obtainable HRV and clinical features marks a significant advancement. By utilizing accessible data and predictive models, clinicians may better identify high-risk patients, enabling personalized treatment and optimized resource use. This study also paves the way for future research to refine these models and integrate them into clinical practice, improving outcomes and reducing the burden of AF recurrence.

Conflict of interest: Mirko De Melis, Yong K Cho and Javier Saiz-Vivo are Medtronic employees.

Bibliography

1. Heeringa, J., et al., *Prevalence, incidence and lifetime risk of atrial fibrillation: The Rotterdam study.* European Heart Journal, 2006. **27**(8): p. 949-953.
2. Chugh, S.S., et al., *Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study.* Circulation, 2014. **129**(8): p. 837-47.
3. Joglar, J.A., et al., *2023 ACC/AHA/ACCP/HRS Guideline for the Diagnosis and Management of Atrial Fibrillation: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines.* Circulation, 2024. **149**(1): p. e1-e156.
4. Wilber, D.J., et al., *Comparison of antiarrhythmic drug therapy and radiofrequency catheter ablation in patients with paroxysmal atrial fibrillation: a randomized controlled trial.* Jama, 2010. **303**(4): p. 333-40.
5. Letsas, K.P., et al., *CHADS2 and CHA2DS2-VASc scores as predictors of left atrial ablation outcomes for paroxysmal atrial fibrillation.* Europace, 2014. **16**(2): p. 202-207.
6. Kornej, J., et al., *The APPLE score - A novel score for the prediction of rhythm outcomes after repeat catheter ablation of atrial fibrillation.* PLoS ONE, 2017. **12**(1): p. 1-9.
7. Kornej, J., et al., *The APPLE score: a novel and simple score for the prediction of rhythm outcomes after catheter ablation of atrial fibrillation.* Clinical Research in Cardiology, 2015. **104**(10): p. 871-876.
8. Jud, F.N., et al., *A novel score in the prediction of rhythm outcome after ablation of atrial fibrillation: The SUCCESS score.* Anatolian Journal of Cardiology, 2019. **21**(3): p. 142-149.
9. Mujović, N., et al., *Prediction of very late arrhythmia recurrence after radiofrequency catheter ablation of atrial fibrillation: The MB-LATER clinical score.* Scientific Reports, 2017. **7**(December 2016): p. 1-12.
10. Calkins, H., et al., *2017 HRS / EHRA / ECAS / APHRS / SOLAECE expert consensus statement on catheter and surgical ablation of atrial fibrillation.* 2018: p. 1-160.
11. Delgado-Bonal, A. and A. Marshak, *Approximate entropy and sample entropy: A comprehensive tutorial*. Vol. 21. 2019.
12. Tayel, M. and E. AlSaba, *Poincaré Plot for Heart Rate Variability.* International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering, 2015. **9**(9): p. 708-711.
13. Cristianini, N. and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. 2000, Cambridge: Cambridge University Press.
14. Somol, P., J. Novovičová, and P. Pudil, *Flexible-Hybrid Sequential Floating Search in Statistical Feature Selection*, in *Structural, Syntactic, and Statistical Pattern Recognition*. 2006. p. 632-639.
15. Zhang, H., et al., *Predictive Value of Machine Learning for Recurrence of Atrial Fibrillation after Catheter Ablation: A Systematic Review and Meta-Analysis.* Reviews in Cardiovascular Medicine, 2023. **24**(11).
16. Bernstein, R.A., et al., *Effect of Long-term Continuous Cardiac Monitoring vs Usual Care on Detection of Atrial Fibrillation in Patients With Stroke Attributed to Large- or Small-Vessel Disease: The STROKE-AF Randomized Clinical Trial.* JAMA, 2021. **325**(21): p. 2169-2177.
17. Saiz-Vivo, J., et al., *Heart Rate Variability and Clinical Features as Predictors of Atrial Fibrillation Recurrence After Catheter Ablation: A Pilot Study.* Front Physiol, 2021. **12**: p. 672896.

Address for correspondence:
Javier Saiz-Vivo
javier.saizvivo@medtronic.com