Machine learning for predicting AF ablation outcomes using daily heart rhythm data at baseline

Authors:
L Tang¹ , K Ho¹ , R Tam¹ , N Hawkins¹ , M Lim¹ , J Andrade¹ , ¹University of British Columbia - Vancouver - Canada

Topic(s):
Catheter Ablation of Arrhythmias

Background: While numerous studies have shown that catheter ablation is superior to antiarrhythmic drug (AAD) in treating atrial fibrillation (AF), the long term outcomes have been limited by arrhythmia recurrence. Reliable data and methods to predict ablation outcomes will thus be valuable for treatment planning.

Objective: To evaluate the utility of machine learning and various types of input variables, viz. patient characteristics at baseline, and daily heart rhythm data recorded prior to ablation for outcome prediction.

Methods: We acquired permission to analyze data collected from a randomized clinical trial that recorded daily biomeasures from >345 patients who were referred for first catheter ablation due to AF refractory to at least one AAD. After standardizing the dataset, each patient sample is characterized by a set of daily measures, viz. heart rate variability (HRV) and AF burden (AFB), which is the total minutes in AF per day. We next performed comparative analyses on 19 candidate model variants to evaluate each model’s ability in identifying patients who were to experience at least one episode of AF recurrence during post-ablation period starting from day 91 up to day 365 post-ablation, per standard guidelines. We examined: i) use of a set of daily biomeasures jointly with baseline sex and age; and ii) observation lengths of the pre-ablation period. We also examined the use of baseline CHA2DS2-VASc scores, left-atrial volume (LAV), atrial diameter, medical history. We conducted multiple sets of 3-fold cross validation (CV) experiments, each fold independently trained a candidate model with 236 samples (two thirds of the dataset) and performed evaluation on the left-out samples. About 50% of cohort belongs to one class. Each fold scored a model and its input variables in terms of sensitivity (SEN), specificity (SPEC), area under receiver operating characteristic curve (AUC), etc. To circumvent risks of overfitting highly parameterized models to our training subset, we shortlisted 19 models that have few hyper-parameters, e.g. stepwise regression, random forest (RF), linear discriminant analysis (LDA).

Results: CV results demonstrated that LDA and RF gave comparable performances, with RF achieving highest AUC of 0.68±0.06 using 30 days of rhythm data prior to ablation (SEN of 65.9±7.82; SPEC of 66.3±0.57). When observation period extended to 90 days prior, AUC improved to 0.691±0.02. In contrast, use of LAV alone was not adequate to predict outcome (AUC~0.5), and when combined with all aforementioned baseline variables, the best model achieved AUC of 0.58±0.05. Feature analyses from the trained models suggest that AFB had highest relevance in predicting outcome. Using only daily AFB, RF and LDA respectively achieved AUC of 0.608±0.04 and 0.652±0.04.

Conclusions: Our results suggest the value of pre-ablation rhythm data for improving outcome-prediction. Future work will validate these findings using large public datasets.