# SUPPLEMENTAL MATERIAL

## Supplemental Methods

### Extraction of Inductive Features from SimAF

For the purpose of extracting inductive features of SimAF, the left atrium (LA) was divided into 11 anatomical regions: base, inferior posterior LA, left of the posterior LA, left atrial appendage, left inferior pulmonary vein, left superior pulmonary vein, mitral valve, mid anterior LA, mid posterior LA, right inferior pulmonary vein, and right superior pulmonary vein. In each patient-specific LA model, we recorded the pacing sites from which sustained AF was induced and the regions where RD or MAT were localized.[1]

The inductive SimAF features were defined via analysis of the training data, then calculated for the training and validation or test data for the inner and outer loops of cross-validation, respectively. They were learned in an unsupervised manner as follows: for each LA model in the training data set and each of the 11 anatomical regions in that model, 6 characteristics of SimAF results were computed: the number of RDs ($n_{RD}$), MATs ($n_{MAT}$), and reentries ($n_{RD+MAT}$) located in the given region, and the proportions of pacing sites in the region (out of the total number of pacing sites) from which RD ($p_{RD}$), MAT ($p_{MAT}$), or a reentry ($p_{RD+MAT}$) was induced. For each characteristic of SimAF results, the differences between patients who did and did not experience AF for each of the 11 anatomical regions were ranked in significance with the Wilcoxon rank sum test. Several anatomical regions with the highest significance were chosen in the calculation of each inductive feature for the training and validation/test sets (Table 1). The numbers of anatomical regions selected from the ranked list for feature calculation were treated as hyperparameters and were optimized as described in the following section.

**Training, Optimization, and Evaluation of QDA Classifier**

Ten-fold nested cross-validation was used to train, validate, and test the AF recurrence risk classifier. In each fold of the "outer loop" of cross-validation, 10% of the data was set aside to use as a test set. The remaining 90% were used for validation and training in the "inner loop" of cross validation. In each fold of the "inner loop", 10% of the data was set aside for validation. The remaining 90% were used to train QDA classifiers with weighted loss to account for any class imbalance in the training data. Hyperparameters of the classifier included the number of features selected and the numbers of anatomical regions used in the calculation of the various inductive simulation features.

In each fold of cross-validation, feature selection was performed using a random forest developed with the training data set, which consisted of out-of-bag permuted predictor importance estimates using a bagged ensemble of 300 regression trees. For each tree in the forest, feature selection was performed with the interaction-curvature test, which chose the split predictor that minimized the p-value of chi-square tests of independence between each feature and the outcome, and that minimized the p-value of a chi-square test of independence between each pair of features and the outcome.[2] The outcome was a binary value indicating whether a given patient experienced AF recurrence. Feature importance for each predictor was determined by randomly permuting predictor values for each tree and observing the effect on the classification error. There was no manual interaction by the researchers in the feature selection process and features from various sources (imaging, SimAF) were not treated differently in any way.

In the "inner loop" of cross-validation, a grid search was used to find the optimal hyperparameters by training QDA classifiers with all possible combinations of the following: K features of highest importance ($1 \leq K \leq 5$), N anatomical regions for inductive SimAF results features that involved

specific reentry locations ($n_{RD}$, $n_{MAT}$, and $n_{RD+MAT}$; $1 \leq N \leq 4$), and P anatomical regions for inductive SimAF results features that involve specific pacing locations ($p_{RD}$, $p_{MAT}$, and $p_{RD+MAT}$; $1 \leq P \leq 4$). Each trained QDA classifier was then used to predict the probability of AF recurrence for the validation set. Following the "inner loop", the hyperparameters which maximized the area under the validation receiver operating characteristic (ROC) curve (AUC), were selected and used to train a classifier with all the patients from the inner loop.

The trained QDA classifier with optimal hyperparameters was used to predict the probability of AF recurrence for the left-out test set. Training, validation, and testing results were aggregated over all loops of cross-validation to create training and validation ROC curves.

**References**


1. Zahid S, Cochet H, Boyle PM, Schwarz EL, Whyte KN, Vigmond EJ, Dubois R, Hocini M, Haïssaguerre M, Jaïs P, et al. Patient-derived models link re-entrant driver localization in atrial fibrillation to fibrosis spatial pattern. *Cardiovasc Res*. 2016;110:443–454.

2. Yoh WY. Regression Trees with. Unbiased Variable Selection. *Korean J Appl Stat*. 2004;17:459–473.